Regression analysis for current status data using the EM algorithm

Christopher S. McMahan, a Lianming Wang b and Joshua M. Tebbs b*, †

We propose new expectation–maximization algorithms to analyze current status data under two popular semi-parametric regression models: the proportional hazards (PH) model and the proportional odds (PO) model. Monotone splines are used to model the baseline cumulative hazard function in the PH model and the baseline odds function in the PO model. The proposed algorithms are derived by exploiting a data augmentation based on Poisson latent variables. Unlike previous regression work with current status data, our PH and PO model fitting methods are fast, flexible, easy to implement, and provide variance estimates in closed form. These techniques are evaluated using simulation and are illustrated using uterine fibroid data from a prospective cohort study on early pregnancy. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords: data augmentation; maximum likelihood; monotone splines; proportional hazards model; proportional odds model; semiparametric regression; survival analysis

1. Introduction

Current status data commonly arise in demographic, epidemiological, and medical studies. The salient feature of current status data is that the failure time is not observed exactly but, instead, is known to be smaller than or larger than the time of examination. One common characteristic of studies that produce current status data is that subjects are observed only once, possibly because of the nature of the failure event or because of resource limitations. For example, in tumorigenicity studies conducted by the National Toxicology Program, rats are exposed to a test agent at different levels to assess the toxicity of the agent. Tumor onset time is not measured directly, especially when the test agent is nonlethal, and tumor status is determined only at the death time of the animal. Therefore, the tumor onset time is either smaller than or larger than the death time depending on whether or not a tumor is detected.

In the analysis of current status data, also known as ‘case 1’ interval-censored data in the statistics literature [1, 2], it is often of interest to estimate survival functions for different treatment groups and to assess the significance of covariate effects on the survival time. Many parametric and semiparametric regression methods have been developed for these purposes. Although parametric models such as the gamma and Weibull are appealing because of their simplicity, the corresponding regression methods may give misleading conclusions because of their overly restrictive model assumptions. In contrast, semiparametric regression models typically enjoy great flexibility and are popular choices when analyzing survival data. In this paper, we focus on two popular semiparametric models, the proportional hazards (PH) model [3] and the proportional odds (PO) model, and we propose new expectation–maximization (EM) algorithms to fit each model with current status data.

Many approaches have been investigated and/or proposed to analyze current status data under the PH and PO models. Under the PH model, among others, Huang [4] explored efficiency issues and established asymptotic results for the maximum likelihood (ML) estimator of both the regression parameter and the baseline cumulative hazard function; Pan [5] extended the iterative convex minorant algorithm

a Department of Mathematical Sciences, Clemson University, Clemson, SC 29634, U.S.A.

b Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

* Correspondence to: Joshua M. Tebbs, 209G LeConte, Department of Statistics, University of South Carolina, Columbia, SC 29208, U.S.A.

† E-mail: tebbs@stat.sc.edu

Although many regression techniques have been developed for current status data, and more generally for interval-censored data, most require substantial effort to implement – not only for practitioners but also for researchers in this area. Furthermore, computational techniques for regression analysis with current status and interval-censored data remain notably underdeveloped [14]. Available R packages for non-Bayesian inference under the semiparametric PH model include intcox [15] based on the modeling approach due to Pan [5], and C1.coxph [14], which implements a generalized Gauss–Seidel algorithm. Unfortunately, both intcox and C1.coxph can produce biased parameter estimates, intcox does not provide variance estimates directly, and variance estimates from C1.coxph are sometimes unsatisfactory [7]. We know of two Bayesian packages available to fit the PH model with interval-censored data: BITE [16] and survBayes [17]. We know of no available software packages to analyze current status or interval-censored data under the PO model.

In this paper, we propose new EM algorithms to analyze current status data under both the PH and PO models. We use monotone splines [18] to model the baseline cumulative hazard function in the PH model and the baseline odds function in the PO model. Our algorithms then make judicious use of Poisson latent variables to fit each model. Our new methods are easy to implement, remarkably efficient, and provide variance estimates in closed form.

2. Models, the observed data likelihood, and monotone splines

2.1. Proportional hazards and proportional odds models

The PH model [3] is the most commonly used regression model for survival data. It specifies that the covariates have a multiplicative effect on the hazard function \( \lambda(t|\mathbf{x}) \), that is,

\[
\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\mathbf{x}' \mathbf{\beta}),
\]

where \( \lambda_0(t) \) is the baseline hazard function and \( \mathbf{x} \) is a vector of covariates. The regression parameter \( \mathbf{\beta} \) is interpreted as the multiplicative effect that covariates have on the hazard function of the failure time. Equivalently, one can rewrite the PH model in terms of the cumulative hazard function,

\[
\Lambda(t|\mathbf{x}) = \Lambda_0(t) \exp(\mathbf{x}' \mathbf{\beta}),
\]

where \( \Lambda(t|\mathbf{x}) = \int_0^t \lambda(s|\mathbf{x})ds \) and \( \Lambda_0(t) = \int_0^t \lambda_0(s)ds \) denote the cumulative hazard functions for the treatment group with covariates \( \mathbf{x} \) and the baseline group, respectively.

The PO model specifies that the covariates have a proportional effect on the odds function, that is,

\[
\frac{F(t|\mathbf{x})}{1 - F(t|\mathbf{x})} = \left\{ \frac{F_0(t)}{1 - F_0(t)} \right\} \exp(\mathbf{x}' \mathbf{\beta}),
\]

where \( F(t|\mathbf{x}) \) and \( F_0(t) \) are the cumulative distribution functions of the failure time for the treatment group with covariates \( \mathbf{x} \) and the baseline group, respectively. Defining \( \Lambda(t|\mathbf{x}) = F(t|\mathbf{x})/(1 - F(t|\mathbf{x})) \) and \( \Lambda_0(t) = F_0(t)/(1 - F_0(t)) \), the odds functions for the treatment and baseline groups, respectively, one obtains an unified expression in Equation (1) for both the PH and PO models but with a different interpretation for \( \Lambda_0(t) \). The PH and PO models are semiparametric in nature as the form of \( \Lambda_0(t) \) is typically left unspecified. Theoretically, for \( F(t|\mathbf{x}) \) to be a valid cumulative distribution function, \( \Lambda_0(t) \) is required to be a nondecreasing function of \( T \) with \( \Lambda_0(0) = 0 \) and \( \Lambda_0(\infty) = \infty \). For practical data analysis over a finite time domain, the only requirement is that \( \Lambda_0(t) \) be nondecreasing with \( \Lambda_0(0) = 0 \).
2.2. Observed data likelihood

Consider a study that consists of \( n \) independent subjects. Let \( T_i \) denote the random failure time for subject \( i \), which is unobserved for current status data. Let \( C_i \) denote the random examination time or censoring time for subject \( i \) and let \( \delta_i = I(T_i \leq C_i) \) denote the censoring indicator; that is, \( \delta_i = 1 \) for left-censoring and \( \delta_i = 0 \) for right-censoring. The observed data can be expressed as a sample of triplets \( \{ (\delta_i, C_i, x_i)' \}, i = 1, \ldots, n \} \), where \( x_i \) is a \( p \times 1 \) covariate vector associated with subject \( i \). We assume throughout that the covariates \( x_i \) are not time dependent. Under the assumption that the failure time and censoring time are independent conditional on the covariates, the observed data likelihood is

\[
\mathcal{L}_{\text{obs}} = \prod_{i=1}^{n} \{ F(C_i | x_i) \}^{\delta_i} \{ 1 - F(C_i | x_i) \}^{1-\delta_i},
\]

where \( F(t | x) \) denotes the cumulative distribution function of the failure time given the covariate \( x \) and takes the form \( F(t | x) = 1 - \exp\{-\Lambda_0(t) \exp(x' \beta)\} \) under the PH model and \( F(t | x) = \Lambda_0(t) \exp(x' \beta) / \{1 + \Lambda_0(t) \exp(x' \beta)\} \) under the PO model.

2.3. Monotone splines

The observed data likelihood in Equation (2) is a function of the regression parameter \( \beta \) and the cumulative baseline hazard function or the baseline odds function \( \Lambda_0(t) \). Unlike with right censored data, partial likelihood methods can not be used with current status data; instead, one needs to estimate both \( \beta \) and \( \Lambda_0(t) \) simultaneously [2]. Estimating \( \Lambda_0(t) \) is challenging because of its infinite dimension and the sparse amount of information available in current status data. It has been shown that the ML estimator of \( \Lambda_0(t) \) converges to \( \Lambda_0(t) \) at an \( n^{1/3} \) rate for current status data [4, 6], much slower than the \( n^{1/2} \) convergence rate for right-censored data. The number of parameters in \( \Lambda_0(t) \) to estimate is on the order of the sample size \( n \) when the censoring time distribution is continuous. In this case, standard maximization algorithms such as Newton–Raphson often experience convergence problems and/or exhibit poor performance.

In the survival literature, approximating unknown functions with splines is common [19–21]. The use of splines bridges the gap between nonparametric and parametric modeling approaches because, while providing adequate flexibility, only a finite number of parameters need to be estimated. In this paper, we adopt the monotone splines of Ramsay [18] to model the baseline cumulative hazard function and the baseline odds function in the PH and PO models, respectively, as

\[
\Lambda_0(t) = \sum_{l=1}^{L} \gamma_l b_l(t),
\]

where the \( b_l \)'s are integrated spline basis functions, each of which is nondecreasing from 0 to 1, and the \( \gamma_l \)'s are nonnegative spline coefficients that ensure monotonicity. This approximation is valid within a finite interval formed by the minimum and maximum of the censoring times in the observed data. The spline basis functions are essentially piecewise polynomial functions. To construct these functions, one needs only to specify a sequence of \( m \) increasing points as knots and to choose a degree for the splines. The degree controls the smoothness of the splines and takes on values 1, 2, and 3 for linear, quadratic, and cubic functions, respectively. The \( L = m + \text{degree} - 2 \) basis functions are completely determined once the degree and the knots have been specified.

The placement of knots plays an important role in specifying monotone splines. In his seminal paper, Ramsay [18] recommended to use only a few knots, say, at the median or at the three quartiles. Cai et al. [7], Wang and Dunson [9], and Lin and Wang [22] similarly adopted the use of monotone splines for different nonparametric functions under various regression models for interval-censored data and recommended using 10–30 equally-spaced knots. Unlike our focus, these latter three references take a Bayesian approach and adopt a shrinkage prior for the spline coefficients to prevent over-fitting problems that may result from using too many knots. In this paper, we use a fixed number of knots to illustrate our proposed non-Bayesian model fitting methods. However, because the choice of \( m \) should ultimately depend on the (unknown) baseline function \( \Lambda_0(t) \), we propose fitting the PH or PO model across different values of \( m \) and then using suitable model selection criteria to select the ‘best’ value. This proposal was also made by Rosenberg [23], and we illustrate this strategy in Section 5.
3. Expectation-maximization algorithms for the proportional hazards and proportional odds models

3.1. Proportional hazards model

Under the PH model, the observed data likelihood in Equation (2) becomes

$$L_{\text{obs}}(\theta) = \prod_{i=1}^{n} \left[ 1 - \exp \left\{ -\Lambda_0(C_i) \exp(x'_i\beta) \right\} \right] \delta_i \exp \left\{ -\Lambda_0(C_i) \exp(x'_i\beta)(1 - \delta_i) \right\}, \quad (4)$$

where $\Lambda_0(\cdot)$ is written using the monotone spline representation in Equation (3). Denote by $\gamma = (\gamma_1, \ldots, \gamma_L)'$ and $\theta = (\beta', \gamma')'$. Our primary goal is to make inference about $\theta$.

Although the first and second derivatives of the logarithm of the observed data likelihood $\log L_{\text{obs}}(\theta)$ with respect to $\theta$ are available, it has been our experience that Newton-type algorithms are not reliable. A traditional Newton–Raphson algorithm can encounter numerical instability – even when analytic expressions of the gradient and Hessian are self-specified – resulting in a lack of convergence or the algorithm getting ‘stuck’ at local extrema. Similarly, we have found that ‘off-the-shelf’ quasi-Newton optimization routines (i.e., those routines that evaluate the gradient numerically) can also experience numerical problems even when $m$ is small. Unfortunately, even when an analytic form of the gradient is specified, these routines can still be extremely sensitive to initial starting values.

To overcome the computational challenges outlined in the last paragraph and to provide a more reliable maximization procedure, we propose an EM algorithm that uses two layers of Poisson latent variables. In the first layer, we relate the censoring indicator $\delta_i$ to the latent variable $Z_i$ according to $\delta_i = I(Z_i > 0)$, where $I(\cdot)$ denotes the usual indicator function and where $Z_i \sim \text{Poisson} \{ \Lambda_0(C_i) \exp(x'_i\beta) \}$. The likelihood function involving the latent $Z_i$’s and the observed data is given by

$$L_*(\theta) = \prod_{i=1}^{n} \delta_i^{I(Z_i>0)}(1 - \delta_i)^{I(Z_i=0)} \left\{ \Lambda_0(C_i) \exp(x'_i\beta) \right\}^{Z_i} \exp \left\{ -\Lambda_0(C_i) \exp(x'_i\beta) \right\} / Z_i!. \quad (5)$$

Note that one can retrieve the observed data likelihood in Equation (4) by integrating Equation (5) over all of the $Z_i$’s. In the second layer, we again exploit the Poisson distribution and also the monotone spline representation of $\Lambda_0(t)$. Specifically, we introduce the independent latent random variables $Z_{il}$, $l = 1, \ldots, L$, where $Z_{il} \sim \text{Poisson} \{ \gamma_l b_l(C_i) \exp(x'_i\beta) \}$ so that $Z_i = \sum_{l=1}^{L} Z_{il}$. At this layer, what we call the complete data likelihood (now involving the latent $Z_i$’s/Z_{il}$’s and the observed data) is given by

$$L_c(\theta) = \prod_{i=1}^{n} \delta_i^{I(Z_i>0)}(1 - \delta_i)^{I(Z_i=0)} I \left( Z_i = \sum_{l=1}^{L} Z_{il} \right) \times \prod_{l=1}^{L} \left\{ \gamma_l b_l(C_i) \exp(x'_i\beta) \right\}^{Z_{il}} \exp \left\{ -\gamma_l b_l(C_i) \exp(x'_i\beta) \right\} / Z_{il}!. \quad (6)$$

To see the connection between the complete data likelihood in Equation (6) and the observed data likelihood in Equation (4), note that one can retrieve Equation (5) by integrating over the $Z_{il}$’s in Equation (6). Our dual-layered choice of latent variables, which takes advantage of the PH model structure and the monotone spline representation for $\Lambda_0(t)$, was also used by Cai et al. [7] in their Bayesian approach.

We now describe our proposed EM algorithm for the PH model. The E-step of the algorithm involves taking the expectation of $\log L_c(\theta)$ with respect to all of the latent variables (the $Z_i$’s and the $Z_{il}$’s) conditional on the observed data and the current parameter $\hat{\theta}^{(d)} = (\beta^{(d)'}, \gamma^{(d)'}')$. This yields the $Q$ function

$$Q(\theta; \hat{\theta}^{(d)}) = \sum_{i=1}^{n} \sum_{l=1}^{L} \left\{ E(Z_{il}) \left[ \log \gamma_l b_l(C_i) + x'_i\beta \right] - \gamma_l b_l(C_i) \exp(x'_i\beta) - E \{ \log (Z_{il}) \} \right\}, \quad (7)$$
where

\[
E(Z_{it}) = \frac{E(Z_i)E(Y_i^{(d)}b_l(C_i)\Lambda_0^{(d)}(C_i))}{\Lambda_0^{(d)}(C_i)}
\]

and \(\Lambda_0^{(d)}(C_i) = \sum_{l=1}^{L} \gamma_i^{(d)} b_l(C_i)\). Note that although all expectations in Equation (7) are conditional on the observed data and the current parameter estimate \(\theta^{(d)}\), we suppress this in our notation henceforth for ease of exposition, that is, we write \(E(Z_i)\) to denote \(E(Z_i|\delta_i, C_i, x_i, \theta^{(d)})\) and \(E(Z_{it})\) to denote \(E(Z_{it}|\delta_i, C_i, x_i, \theta^{(d)})\). The expression for \(E(Z_i)\) is obtained easily by using Equation (5). The expression for \(E(Z_{it})\) is found by observing that the conditional distribution of \(Z_{it}\) given \(Z_i\) is binomial. The M-step finds \(\theta^{(d+1)} = \arg \max_{\theta} Q(\theta, \theta^{(d)})\); note that the partial derivatives of \(Q(\theta, \theta^{(d)})\) are available in closed form and are given by

\[
\frac{\partial Q(\theta, \theta^{(d)})}{\partial \beta} = \sum_{i=1}^{n} \{E(Z_i) - \Lambda_0(C_i) \exp(x_i^{l}\beta)\} x_i,
\]

\[
\frac{\partial Q(\theta, \theta^{(d)})}{\partial \gamma_l} = \sum_{i=1}^{n} \{\gamma_i^{l-1} E(Z_{il}) - b_l(C_i) \exp(x_i^{l}\beta)\}, \quad l = 1, \ldots, L.
\]

Setting these derivatives equal to zero and solving the resulting system of equations for \(\theta\) yields \(\theta^{(d+1)}\). Towards this end, note that solving \(\frac{\partial Q(\theta, \theta^{(d)})}{\partial \gamma_l} = 0\) leads to an explicit form of \(\gamma_i^{l(d+1)}\), which itself is a function of \(\beta^{(d+1)}\). Inserting \(\gamma_i^{l(d+1)}\) into the equations \(\frac{\partial Q(\theta, \theta^{(d)})}{\partial \beta} = 0\), one then solves for \(\beta^{(d+1)}\).

For implementation purposes, we now summarize our proposed EM algorithm succinctly for the PH model. First, initialize \(\theta^{(d)} = (\beta^{(d)'}, \gamma^{(d)'})'\) for \(d = 0\) and then simply repeat the following two steps until convergence:

1. Calculate \(\beta^{(d+1)}\) by solving the system of \(p\) equations

\[
\sum_{i=1}^{n} \left[ E(Z_i) - \left\{ \sum_{l=1}^{L} b_l(C_i) \sum_{i'=1}^{L} E(Z_{i'^{l}}) \right\} \exp(x_i^{l}\beta) \right] x_i = 0.
\]

2. Calculate \(\gamma_i^{l(d+1)}\) according to

\[
\gamma_i^{l(d+1)} = \frac{\sum_{i=1}^{n} E(Z_{it})}{\sum_{i=1}^{n} E(Z_{il})} b_l(C_i) \exp(x_i^{l}\beta^{(d+1)}), \quad l = 1, \ldots, L.
\]

and update \(d = d + 1\).

Denote the final value at convergence by \(\hat{\theta} = (\hat{\beta}', \hat{\gamma}')\). In the Web-based Supporting Materials\(^1\), we show that \(\hat{\theta}\) solves the score equations associated with the observed data likelihood at the point of convergence; that is, that \(\hat{\theta}\) is a ML estimator. One will note that in the our algorithm, spline coefficients are always updated in closed form; this results in increased stability when compared with Newton-type optimization routines. In addition, our algorithm explicitly acknowledges the natural constraints placed on the spline coefficients; that is, \(\gamma_l \geq 0\), for \(l = 1, \ldots, L\).

A direct appeal to the missing information principle and Louis’s method [24] gives a closed-form expression for the covariance matrix of \(\hat{\theta}\), that is,

\[
\text{var}(\hat{\theta}) = \left\{ -\frac{\partial^2 \log L_{\text{obs}}(\theta)}{\partial \theta \partial \theta'} \right\}^{-1},
\]

\(^1\)Supporting information may be found in the online version of this article
where

\[-\frac{\partial^2 \log \mathcal{L}_{\text{obs}}(\theta)}{\partial \theta \partial \theta'} = - \frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \theta \partial \theta'} - \text{var} \left\{ \frac{\partial \log \mathcal{L}_c(\theta)}{\partial \theta} \right\} \]

Another appealing characteristic of our approach is that all quantities involved in \( \text{var}(\hat{\theta}) \) are available in closed form and are easy to calculate; see Appendix A. An estimate of the covariance matrix is obtained by evaluating \( \text{var}(\hat{\theta}) \) at \( \theta \). Wald inference involving \( \beta \) and the construction of large-sample confidence intervals for the baseline CDF and survival functions can be carried out in the usual way.

### 3.2. Proportional odds model

The approach in Section 3.1 can be adapted to the PO model by writing the survival function \( S(t|x) \) under the PO model as the integrated conditional survival function under the frailty PH model with respect to an exponential frailty \( \phi \); that is,

\[ S(t|x) = \{1 + \Lambda_0(t) \exp(x' \beta)\}^{-1} = \int_0^\infty \exp \{ -\Lambda_0(t) \exp(x' \beta) \phi \} \exp(-\phi) d\phi. \]

By using this connection, one can expand the observed data likelihood in Equation (2) under the PO model by introducing a latent exponential \( \mathcal{E}(1) \) variable \( \phi_i \) for subject \( i \) as follows:

\[ \mathcal{L}_c(\theta) = \prod_{i=1}^n \left[ 1 - \exp \left\{ -\Lambda_0(C_i) \exp(x_i' \beta) \phi_i \right\} \right] \exp \left\{ -\Lambda_0(C_i) \exp(x_i' \beta) \phi_i (1 - \delta_i) \right\} \exp(-\phi_i). \]  

(8)

This likelihood takes the same form as the observed data likelihood in Equation (4) under the PH model except for the additional \( \phi_i \) terms. To develop an EM algorithm under the PO model, we introduce the following latent random variables:

\[ Z_i | \phi_i \sim \text{Poisson} \{ \Lambda_0(C_i) \exp(x_i' \beta) \phi_i \}, \]

\[ Z_{il} | \phi_i \sim \text{Poisson} \{ \gamma_l b_l(C_i) \exp(x_i' \beta) \phi_i \}, \quad l = 1, \ldots, L. \]

where the \( Z_{il} \)'s, conditional on \( \phi_i \), are independent. With these selections, the complete data likelihood can be written as

\[ \mathcal{L}_c(\theta) = \prod_{i=1}^n \delta_i^{I(Z_i > 0)} (1 - \delta_i)^{I(Z_i = 0)} \left( Z_i = \sum_{l=1}^L Z_{il} \right) \exp(-\phi_i) \]

\[ \times \prod_{l=1}^L \{ \gamma_l b_l(C_i) \exp(x_i' \beta) \phi_i \}^{Z_{il}} \exp\left\{ -\gamma_l b_l(C_i) \exp(x_i' \beta) \phi_i \right\} / Z_{il}!. \]

(9)

Integrating over the \( Z_{il} \)'s and \( Z_i \)'s in Equation (9) leads to the expanded likelihood in Equation (8); similarly, integrating over the \( \phi_i \)'s in Equation (8) leads to the observed data likelihood in Equation (2) under the PO model.

The E-step of the proposed algorithm for the PO model involves taking the expected value of \( \log \mathcal{L}_c(\theta) \) in Equation (9) with respect to all of the latent variables (here, the \( Z_i \)'s/\( Z_{il} \)'s and the \( \phi_i \)'s) conditional on the observed data and the current parameter estimate \( \theta^{(d)} \); this yields

\[ Q(\theta, \theta^{(d)}) = \sum_{i=1}^n \sum_{l=1}^L \left\{ E \left[ Z_{il} \log \gamma_l b_l(C_i) \exp(x_i' \beta) \phi_i \right] \right\} - \gamma_l b_l(C_i) \exp(x_i' \beta) E(\phi_i) \]

(10)

up to an additive term that is a function of \( \theta^{(d)} \) but is free of \( \theta \). In the M-step, \( Q(\theta, \theta^{(d)}) \) is then maximized with respect to \( \theta \) to obtain \( \theta^{(d+1)} \). The proposed EM algorithm for the PO model proceeds by initializing \( \theta^{(d)} \) for \( d = 0 \) and then repeating the following two steps until convergence:

1. Calculate \( \beta^{(d+1)} \) by solving the system of \( p \) equations

\[ \sum_{i=1}^n \left[ E(Z_i) - E(\phi_i) \right] \left\{ \sum_{l=1}^L \frac{b_l(C_i) \sum_{i' = 1}^n E(Z_{il})}{\sum_{i' = 1}^n b_l(C_{i'}) \exp(x_{i'}' \beta) E(\phi_{i'})} \right\} \exp(x_i' \beta) \]  

\[ x_i = 0. \]
4. Simulation evidence

4.1. Simulation models and description

We used simulation to assess the performance of our EM algorithms under the PH and PO models. The failure time $T_i$ was generated from

$$F(t|x_i) = 1 - \exp\{-\Lambda_0(t) \exp(x_{i1}\beta_1 + x_{i2}\beta_2)\}$$

under the PH model and from

$$F(t|x_i) = \frac{\Lambda_0(t) \exp(x_{i1}\beta_1 + x_{i2}\beta_2)}{1 + \Lambda_0(t) \exp(x_{i1}\beta_1 + x_{i2}\beta_2)}$$

under the PO model, where $\Lambda_0(t) = \log(1 + t + t^{3/2})$, $x_{i1} \sim N(0, 0.25)$, $x_{i2} \sim \text{Bernoulli}(0.5)$, $x_i = (x_{i1}, x_{i2})'$, and where both $\beta_1$ and $\beta_2$ can take on the values $-0.5, 0,$ or $0.5$, resulting in 9 parameter configurations under each model. The censoring time $C_i$ was generated from a truncated exponential distribution $\mathcal{E}(1)$ with support $(0, 2)$, and the censoring indicator was obtained by $\delta_i = I(T_i \leq C_i)$. Note that $\Pr(T_i > 2)$ is small for all covariate and parameter configurations under both models, which explains why we selected this specific truncated distribution.

We generated 500 independent data sets, each with a sample size of $n = 200$, at each parameter configuration under both models. For the monotone spline specifications, we used both $m = 5$ and $m = 10$ equally-spaced knots within the minimum and maximum censoring times and used degree 3 for adequate smoothness. We used the starting value $\theta^{(0)} = (\beta^{(0)})'(\beta^{(0)})'(Y_2, I_2)'$, $L = m + 1$, for the results presented herein; however, we observed in a separate investigation that the performance of our EM algorithms, in terms of accuracy and speed, was largely invariant to the choice of $\theta^{(0)}$. Convergence was declared when the maximum change in $\theta^{(d)}$ was less than 0.001. For the 500 simulated data sets under each model, parameter configuration, and choice of $m$ (9000 data sets in total), the average time it was less than 0.001. For the 500 simulated data sets (9000 data sets in total), the average time it was less than 0.001. For the 500 simulated data sets (9000 data sets in total), the average time it was less than 0.001.
4.2. Simulation results

We first examined the performance of our EM algorithm for the PH model and how it compared with the two competing packages mentioned in Section 1: \texttt{intcox} [15] and \texttt{C1.coxph} [14]. Table I summarizes the results. First, one notes the sample means of the 500 ML estimates of $\beta_1$ and $\beta_2$ are all close to the true parameter values for EM with $m = 5$ knots (also when $m = 10$, shown in the Web-based Supporting Materials). The same can not be said for the competing methods that can exhibit high bias, a finding also observed in [7]. Second, for EM, the sample standard deviation of the 500 estimates and the averaged standard error are in close agreement, indicating that the covariance matrix of \[ \hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)^t \] is being estimated correctly on average. Recall that the \texttt{intcox} package does not produce variance estimates automatically. Furthermore, SD and SE for \texttt{C1.coxph} often largely disagree; not surprisingly, the Wald coverage probability estimates associated with \texttt{C1.coxph} are incongruously low. On the other hand, the same coverage estimates calculated from EM are always within the margin of Monte Carlo error.

Because our EM algorithms produce the ML estimator $\hat{\theta}$, one might wonder how an EM solution compares to the solution one would obtain by attempting to maximize the observed data likelihood in Equation (2) directly. Towards investigating this, we examined different off-the-shelf quasi-Newton optimization routines to maximize the observed data log-likelihood function. However, as noted in

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>Method</th>
<th>Bias SD</th>
<th>SD SE</th>
<th>MSE Cov</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\hat{\beta}_1$</td>
<td>$\hat{\beta}_2$</td>
<td>$\hat{\beta}_1$</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.50</td>
<td>EM</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>-0.01</td>
<td>-0.06</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>-0.08</td>
<td>-0.07</td>
<td>0.28</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.00</td>
<td>EM</td>
<td>-0.03</td>
<td>-0.01</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>-0.08</td>
<td>-0.01</td>
<td>0.27</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.50</td>
<td>EM</td>
<td>-0.03</td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>-0.01</td>
<td>0.07</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>-0.08</td>
<td>0.07</td>
<td>0.28</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.50</td>
<td>EM</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.05</td>
<td>-0.07</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EM</td>
<td>0.01</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>0.00</td>
<td>0.00</td>
<td>EM</td>
<td>0.06</td>
<td>0.01</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.02</td>
<td>0.01</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>0.01</td>
<td>0.05</td>
<td>0.24</td>
</tr>
<tr>
<td>0.00</td>
<td>0.50</td>
<td>EM</td>
<td>0.04</td>
<td>0.10</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.00</td>
<td>0.12</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>0.00</td>
<td>0.02</td>
<td>0.23</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.50</td>
<td>EM</td>
<td>0.06</td>
<td>0.07</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.09</td>
<td>0.11</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>0.01</td>
<td>0.01</td>
<td>0.24</td>
</tr>
<tr>
<td>0.50</td>
<td>0.00</td>
<td>EM</td>
<td>0.07</td>
<td>0.02</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.10</td>
<td>0.02</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{C1.coxph}</td>
<td>0.02</td>
<td>0.02</td>
<td>0.23</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>EM</td>
<td>0.07</td>
<td>0.07</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>\texttt{intcox}</td>
<td>0.11</td>
<td>0.11</td>
<td>0.31</td>
</tr>
</tbody>
</table>

SD: standard deviation; SE: standard error.
Section 3.1, we found that most of these frequently failed when applied to either model, especially those that required numerical evaluation of the gradient. Further investigation on our part revealed that these routines are sensitive to the choice of initial starting values in $\theta^{(b)}$ and often converged to local maxima. Ultimately, we experienced success with one constrained quasi-Newton algorithm [25] but only
when we (a) provided the maximization routine an analytic expression for the gradient and (b) specified incredibly informative starting values. Of course, the latter may not be possible to do in practice. Table II displays the results from comparing our EM algorithms with a direct ML approach by using the algorithm described in [25] with \( m = 5 \) knots; the same comparison with \( m = 10 \) is given in the Web-based Supporting Materials. The results from Table II reveal the computational advantages that are possible with our EM algorithms. For the PH model, there is some evidence that the covariance matrix of \( \hat{\beta} \) is not being estimated correctly with the direct ML routine, potentially leading to Wald confidence intervals being slightly anti-conservative, and that EM provides estimates with smaller mean squared error (MSE). The \( m = 10 \) results, shown in the Web-based Supporting Materials, suggest that the ML routine can experience problems for the PO model; on the other hand, estimates from the EM algorithm are less variable and lead to Wald inference at nominal levels. We also compared EM and ML estimates of the baseline cumulative distribution function \( F_0(t) \). Table III gives the mean MSE and maximum MSE of the estimate \( \hat{F}_0(t) \) calculated over a pre-specified set of time points (i.e., 20 equally-spaced points across the support of the censoring time distribution) for both models with \( m = 5 \) and \( m = 10 \). These two measures provide a global assessment of how well \( \hat{F}_0(t) \) estimates the true baseline cumulative distribution function, and the results in Table III indicate that our EM approach provides smaller values for both. Overall, even with the special computational efforts given to make the direct ML routine viable for either the PH or PO model, our EM results are as good or better.

5. Uterine fibroid data analysis

Right from the Start (RFTS) is a prospective cohort study of early pregnancy conducted in North Carolina, Tennessee, and Texas. Participants in the study are required to be at least 18 years old and were to have become pregnant just before the time of enrollment. A detailed description of the study can be found in Laughlin et al. [26]. One objective of RFTS was to estimate the cumulative incidence of uterine fibroids for African American and white subjects, and to identify potential fibroid risk factors.

| Table III. Mean mean squared error (MSE) (×10^3) and maximum MSE (×10^3) of the estimated baseline cumulative distribution function \( \hat{F}_0(t) \) under the proportional hazards (PH) and proportional odds (PO) models with \( m = 5 \) and \( m = 10 \) knots. Mean and maximum values are calculated over a pre-specified set of time points as described in Section 4.2. Direct maximization (ML) was implemented using the constrained quasi-Newton optimization routine in [25]. |
|---------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| \( \beta_1 \)                 | \( \beta_2 \) |                |            |              |              |              |              |
|                                | \( m = 5 \)  | \( m = 10 \)  | \( m = 5 \) | \( m = 10 \) | \( m = 5 \)  | \( m = 10 \)  |              |
| Mean MSE                       |              |              |              |              |              |              |              |
| 0.50                           | 0.00         | 2.3          | 3.9          | 4.1          | 5.6          | 4.3          | 4.9          | 5.2          | 6.3          |
| 0.50                           | 0.00         | 2.5          | 4.1          | 4.4          | 8.0          | 4.8          | 5.3          | 5.6          | 6.9          |
| 0.00                           | 0.00         | 2.3          | 3.8          | 4.1          | 6.3          | 4.2          | 4.6          | 5.1          | 6.3          |
| 0.50                           | 0.00         | 2.7          | 3.6          | 3.7          | 5.0          | 4.3          | 4.8          | 5.1          | 6.0          |
| Mean MSE                       |              |              |              |              |              |              |              |
| Maximum MSE                    |              |              |              |              |              |              |              |
| 0.50                           | 0.00         | 2.8          | 3.4          | 3.7          | 4.9          | 5.0          | 5.4          | 5.9          | 7.3          |
| 0.50                           | 0.00         | 2.7          | 3.3          | 3.5          | 5.0          | 4.5          | 4.8          | 5.5          | 6.8          |
| 0.50                           | 0.00         | 2.6          | 3.2          | 3.4          | 4.7          | 4.4          | 4.7          | 5.4          | 6.6          |
Uterine fibroids, also known as uterine leiomyoma, have been found to be associated with adverse pregnancy outcomes, such as difficulty in conceiving, premature birth, and the need for cesarean delivery [27, 28].

In RFTS, each subject completed an ultrasound examination as early as possible in the pregnancy (the goal was to have this examination before the seventh gestational week). A subject was classified as ‘fibroid positive’ if any fibroid with diameter larger than five millimeters was detected during the ultrasound examination. Therefore, in the RFTS study, the onset time of fibroids was never exactly observed; instead, the onset time for a subject was smaller than or larger than the subject’s age at the time of examination, resulting in current status data. Because a participant’s ultrasound was scheduled as early as possible in the pregnancy and because the participants were not aware of their fibroid status before examination, we proceed under the assumption that the onset time of fibroids and the ultrasound examination time are independent so that the observed data likelihood in Equation (2) is valid.

For the purposes of illustration, we focus on the second RFTS substudy, which includes 227 African American and 1377 white subjects; the first substudy had serious problems with underreporting [9] and the third substudy is still ongoing. In addition to the examination time \( C_i \) and the fibroid status \( \delta_i \) for each subject, investigators collected the following covariates: parity status (whether a subject had given birth before (0 otherwise), age of menarche (when a subject had her first period), obesity status, and race (African American or white). Let \( x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})' \) denote the covariate vector for subject \( i \), where \( x_{i1} = 1 \) if subject \( i \) had given birth before (0 otherwise), \( x_{i2} \) is the standardized age of menarche, \( x_{i3} = 1 \) if subject \( i \) was categorized as obese (0 otherwise), and \( x_{i4} = 1 \) if subject \( i \) is African American (0 otherwise).

Table IV displays the regression estimates from fitting the PH and PO models by using our EM algorithms in Section 3. We fit both models by using \( m \) equally spaced knots, \( m \in \{3, 4, \ldots, 10\} \), across the minimum and maximum of the examination times; estimated standard errors were computed using the expressions provided in the appendices. Towards identifying a suitable choice of \( m \) for these data, under either model (PH or PO), we consider three standard model selection criteria: Akaike’s information criterion [29], the Bayesian information criterion, and the ‘leave-one-out’ cross validation statistic defined by

\[
\hat{CV} = \sum_{i=1}^{1604} \left( \hat{\delta}_i - \hat{F}_i(C_i|x_i) \right)^2.
\]

### Table IV. Uterine fibroid data. Regression estimates under the proportional hazards (PH) and proportional odds (PO) models by using \( m \) equally spaced knots. Estimated standard errors are in parentheses. The cross validation statistic \( \hat{CV} \) is calculated as described in Section 5. Values of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are also included.

<table>
<thead>
<tr>
<th>Parity status</th>
<th>Age of menarche</th>
<th>Obesity status</th>
<th>Race</th>
<th>Model selection criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m )</td>
<td>( \hat{\beta}_1 )</td>
<td>( \hat{\beta}_2 )</td>
<td>( \hat{\beta}_3 )</td>
<td>( \hat{\beta}_4 )</td>
</tr>
<tr>
<td>PH model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.299 (0.144)</td>
<td>-0.147 (0.066)</td>
<td>0.076 (0.175)</td>
<td>1.424 (0.167)</td>
</tr>
<tr>
<td>4</td>
<td>-0.302 (0.144)</td>
<td>-0.146 (0.066)</td>
<td>0.080 (0.176)</td>
<td>1.412 (0.167)</td>
</tr>
<tr>
<td>5</td>
<td>-0.298 (0.144)</td>
<td>-0.147 (0.065)</td>
<td>0.071 (0.175)</td>
<td>1.422 (0.167)</td>
</tr>
<tr>
<td>6</td>
<td>-0.293 (0.146)</td>
<td>-0.148 (0.066)</td>
<td>0.085 (0.176)</td>
<td>1.416 (0.168)</td>
</tr>
<tr>
<td>7</td>
<td>-0.295 (0.147)</td>
<td>-0.146 (0.066)</td>
<td>0.082 (0.177)</td>
<td>1.414 (0.170)</td>
</tr>
<tr>
<td>8</td>
<td>-0.299 (0.147)</td>
<td>-0.151 (0.066)</td>
<td>0.074 (0.176)</td>
<td>1.428 (0.171)</td>
</tr>
<tr>
<td>9</td>
<td>-0.293 (0.147)</td>
<td>-0.151 (0.066)</td>
<td>0.085 (0.177)</td>
<td>1.416 (0.177)</td>
</tr>
<tr>
<td>10</td>
<td>-0.293 (0.147)</td>
<td>-0.149 (0.066)</td>
<td>0.086 (0.177)</td>
<td>1.417 (0.183)</td>
</tr>
<tr>
<td>PO model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-0.340 (0.163)</td>
<td>-0.168 (0.078)</td>
<td>0.119 (0.201)</td>
<td>1.605 (0.201)</td>
</tr>
<tr>
<td>4</td>
<td>-0.341 (0.163)</td>
<td>-0.169 (0.078)</td>
<td>0.130 (0.202)</td>
<td>1.592 (0.201)</td>
</tr>
<tr>
<td>5</td>
<td>-0.337 (0.163)</td>
<td>-0.167 (0.078)</td>
<td>0.115 (0.201)</td>
<td>1.608 (0.200)</td>
</tr>
<tr>
<td>6</td>
<td>-0.333 (0.166)</td>
<td>-0.169 (0.078)</td>
<td>0.127 (0.202)</td>
<td>1.603 (0.202)</td>
</tr>
<tr>
<td>7</td>
<td>-0.334 (0.174)</td>
<td>-0.169 (0.079)</td>
<td>0.131 (0.203)</td>
<td>1.603 (0.205)</td>
</tr>
<tr>
<td>8</td>
<td>-0.341 (0.168)</td>
<td>-0.171 (0.079)</td>
<td>0.120 (0.203)</td>
<td>1.616 (0.213)</td>
</tr>
<tr>
<td>9</td>
<td>-0.339 (0.173)</td>
<td>-0.172 (0.079)</td>
<td>0.129 (0.204)</td>
<td>1.604 (0.249)</td>
</tr>
<tr>
<td>10</td>
<td>-0.338 (0.162)</td>
<td>-0.172 (0.079)</td>
<td>0.135 (0.202)</td>
<td>1.603 (0.146)</td>
</tr>
</tbody>
</table>
where \( \hat{F}_{(-i)}(\cdot) \) is the estimated cumulative distribution function based on all of the observations except the \( i \)th one. For the PH model, one notes that, like the estimates and their standard errors, the values of \( \hat{c}_V \) remain fairly constant across the values of \( m \). This suggests that, in terms of prediction, the estimated models are fairly robust to the number of knots used and ultimately a relatively small knot set provides adequate flexibility for estimating \( \Lambda_0(t) \). In this light, we would recommend selecting \( m = 3 \) as a final PH model (one interior knot), because this choice, as seen in Table IV, provides the smallest Akaike information criterion and Bayesian information criterion. Our recommendation for a final PO model would be identical.

Examining the PH and PO model results when \( m = 3 \), it appears that parity, age of menarche, and race are each significant risk factors for uterine fibroid development, whereas obesity status is not significant (this statement ignores multiple comparisons). In particular, the hazard rate of developing fibroids for African American subjects is estimated to be \( e^{1.424} \approx 4.15 \) times that for white subjects under the PH model, and the odds of developing fibroids for African American subjects is estimated to be approximately \( e^{1.605} \approx 4.98 \) times that for white subjects under the PO model.

Figure 1 presents the estimated cumulative incidence functions for African American and white subjects, under both models, when the standardized age of menarche \( x_2 = 0 \) for non-obese subjects (i.e., \( x_3 = 0 \)) with \( m = 3 \). Also included are large-sample 95% pointwise confidence bands, which we calculated using the Delta Method. Both plots show that African American subjects have a substantially higher risk of developing fibroids than white subjects over a large range of the ages. Note that we decided to truncate the horizontal axes in Figure 1 at 40, because only 14 subjects (out of 1604) had ages larger than this value.

6. Discussion

In this paper, we have proposed new EM algorithms for analyzing current status data under the PH and PO regression models. Our EM algorithms are easy to implement, converge quickly, and provide variance estimates of ML estimates in closed form. Our simulations show that these methods perform well when estimating both the regression parameters and the corresponding baseline functions, and we have applied these methods to current status data from a study on early pregnancy health. Our R programs for data analysis are available by request.

The use of monotone splines provides computational efficiency while maintaining adequate modeling flexibility, and it produces a smooth estimate of the baseline survival function. In the literature, many authors have chosen to model the logarithm of the baseline hazard function \( \log \Lambda_0(t) \) with B-splines or...
with power series polynomial splines under the PH model with survival data [17, 19–21]. This approach requires numerical approximation to evaluate \( \Lambda_0(t) \), because the corresponding integral does not exist in closed form. In contrast, the monotone spline representation in Equation (3) allows us to evaluate \( \Lambda_0(t) \) directly, and the additive form of \( \Lambda_0(t) \) as represented via monotone splines facilitates directly the use of data augmentation as shown in Section 3. The highly desirable properties of our EM algorithm approach cannot be realized when using B-splines.

The methods outlined in this paper could be generalized for use with more flexible semiparametric regression models with current status data, such as the class of linear transformation models with a \( G^p \) link [30]. We believe that our approach can also be generalized to analyze interval-censored data under both the PH and PO models and possibly further extended to handle clustered response and/or time-dependent covariates. Future efforts will be devoted to developing a comprehensive and user-friendly statistical package to analyze current status data and general interval-censored data under various semiparametric regression models based on the methods outlined in this paper.

**Appendix A. Calculating \( \text{var}(\hat{\theta}) \) under the proportional hazards model.**

Louis’s method [24] states that the covariance matrix of \( \hat{\theta} \) can be calculated via

\[
\text{var}(\hat{\theta}) = \left\{ -\frac{\partial^2 \log L_{\text{obs}}(\theta)}{\partial \theta \partial \theta'} \right\}^{-1}.
\]

where

\[
-\frac{\partial^2 \log L_{\text{obs}}(\theta)}{\partial \theta \partial \theta'} = -\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \theta \partial \theta'} - \text{var}\left\{ \frac{\partial \log L_c(\theta)}{\partial \theta} \right\}.
\]

The function \( Q(\theta, \hat{\theta}) \) is given by Equation (7) with \( \hat{\theta} \) replacing \( \theta^{(d)} \) and \( L_c(\theta) \) is given in Equation (6). We now present closed-form expressions for \( \frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \theta \partial \theta'} \) and \( \text{var}\{\partial \log L_c(\theta)/\partial \theta \} \) under the PH model in Section 3.1. Second derivatives of \( Q(\theta, \hat{\theta}) \) with respect to \( \theta \) are given by

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \beta \partial \beta'} = \sum_{i=1}^{n} \Lambda_0(C_i) \exp(x'_i \beta) x_i x'_i
\]

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \beta \partial \gamma_l} = \sum_{i=1}^{n} b_l(C_i) \exp(x'_i \beta) x_i
\]

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \gamma_l \partial \gamma_{l'}} = -\gamma_l^{-2} \sum_{i=1}^{n} E(Z_{il})
\]

and \( \frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \gamma_l \partial \gamma_{l'}} = 0 \), for \( l \neq l' \), where \( E(Z_{il}) = E(Z_i) \gamma_l b_l(C_i)(\Lambda_0(C_i))^{-1} \) and

\[
E(Z_i) = \Lambda_0(C_i) \exp(x'_i \beta) \delta_i \left[ 1 - \exp\left\{-\Lambda_0(C_i) \exp(x'_i \beta) \right\} \right]^{-1}.
\]

As noted in the paper, we write all expectations as unconditional expectations for brevity; see Section 3.1. The necessary block/scalar entries of \( \text{var}\{\partial \log L_c(\theta)/\partial \theta \} \) are given by

\[
\text{cov}\left( \frac{\partial \log L_c(\theta)}{\partial \beta}, \frac{\partial \log L_c(\theta)}{\partial \beta} \right) = \sum_{i=1}^{n} \text{var}(Z_i) x_i x'_i
\]

\[
\text{cov}\left( \frac{\partial \log L_c(\theta)}{\partial \beta}, \frac{\partial \log L_c(\theta)}{\partial \gamma_l} \right) = \gamma_l^{-1} \sum_{i=1}^{n} \text{cov}(Z_{il}, Z_i) x_i
\]

\[
\text{cov}\left( \frac{\partial \log L_c(\theta)}{\partial \gamma_l}, \frac{\partial \log L_c(\theta)}{\partial \gamma_{l'}} \right) = (\gamma_l \gamma_{l'})^{-1} \sum_{i=1}^{n} \text{cov}(Z_{il}, Z_{il'}). \]
where \( \text{var}(Z_i) = E(Z_i) - \{E(Z_i)^2 \exp(-\Lambda_0(C_i) \exp(x'_i \beta)) \}, \) \( \text{cov}(Z_{il}, Z_{i'l'}) = \gamma_{l'l'}b_l(C_i)b_{l'}(C_i)\{\text{var}(Z_i) - E(Z_i)\}{\Lambda_0(C_i)}^{-2} \) \( + \gamma_{l'l'}E(Z_i)I(l = l'){\Lambda_0(C_i)}^{-1} \).

The estimated covariance matrix is obtained by evaluating \( \text{var}(\hat{\theta}) \) at \( \hat{\theta} = (\hat{\beta}', \hat{\gamma}')' \).

**Appendix B. Calculating \( \text{var}(\hat{\theta}) \) under the proportional odds model.**

We again appeal to Louis’s method [24] with notation defined as in Appendix A. Consider the function \( Q(\theta, \hat{\theta}) \) given by the expression in Equation (10) with \( \hat{\theta} \) replacing \( \theta^{(d)} \). The second derivative matrix \( \partial^2 Q(\theta, \hat{\theta})/\partial \theta \partial \theta' \) contains the entries

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} \Lambda_0(C_i) \exp(x'_i \beta)E(x_i' \beta) x_i \]

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \beta \partial \gamma_t} = -\sum_{i=1}^{n} b_l(C_i) \exp(x'_i \beta)E(x_i' \beta) x_i \]

\[
\frac{\partial^2 Q(\theta, \hat{\theta})}{\partial \gamma_t \partial \gamma_{l'}} = -\gamma_t^{-2} \sum_{i=1}^{n} E(Z_{i'l'}) \]

and \( \partial^2 Q(\theta, \hat{\theta})/\partial \gamma_t \partial \gamma_{l'} = 0, \) for \( l \neq l' \), where \( E(Z_{i'l'}) = E(Z_{i'l'})\gamma_{l'1}b_{l'}(C_l)\{\Lambda_0(C_i)\}^{-1}, \) \( E(Z_i) = \{1 + \Lambda_0(C_i) \exp(x'_i \beta)\} \delta_i, \) and \( E(\phi_t) = \{1 + E(Z_i)\} \{1 + \Lambda_0(C_i) \exp(x'_i \beta)\}^{-1} \) are written as unconditional expectations. Consider the function \( L_c(\theta) \) given by Equation (9). The necessary block/scalar entries of \( \text{var}[\partial \log L_c(\theta)]/\partial \theta \) are given by

\[
\text{cov} \left( \frac{\partial \log L_c(\theta)}{\partial \beta}, \frac{\partial \log L_c(\theta)}{\partial \beta'} \right) = \sum_{i=1}^{n} \left[ \text{var}(Z_i) - 2\Lambda_0(C_i) \exp(x'_i \beta)\text{cov}(Z_i, \phi_t) \right. \nonumber \]

\[+\left. \Lambda_0(C_i) \exp(x'_i \beta)\right] x_i x'_i \]

\[
\text{cov} \left( \frac{\partial \log L_c(\theta)}{\partial \beta}, \frac{\partial \log L_c(\theta)}{\partial \gamma_t} \right) = \sum_{i=1}^{n} \left[ \gamma_t^{-1} \text{cov}(Z_{i'l'}, Z_i) - b_{l'}(C_i) \exp(x'_i \beta)\text{cov}(Z_i, \phi_t) \right] x_i \]

\[+ \sum_{i=1}^{n} \left[ b_l(C_i) \exp(x'_i \beta)\text{var}(\phi_t) - \gamma_t^{-1} \text{cov}(Z_i, \phi_t) \right] \]

\[\times \Lambda_0(C_i) \exp(x'_i \beta) x_i \]

\[
\text{cov} \left( \frac{\partial \log L_c(\theta)}{\partial \gamma_{l'}}, \frac{\partial \log L_c(\theta)}{\partial \gamma_{l'}} \right) = \sum_{i=1}^{n} \left[ \gamma_t^{-1} \gamma_{l'}^{-1} \text{cov}(Z_{i'l'}, Z_{i'l'}) - \gamma_t^{-1} b_{l'}(C_i) \exp(x'_i \beta)\text{cov}(Z_{i'l'}, \phi_t) \right. \nonumber \]

\[ - \gamma_t^{-1} b_l(C_i) \exp(x'_i \beta)\text{cov}(Z_{i'l'}, \phi_t) \nonumber \]

\[+ \left. b_l(C_i) b_{l'}(C_i) \exp(2x'_i \beta)\text{var}(\phi_t) \right], \]

where \( \text{var}(\phi_t) = \{1 + \Lambda_0(C_i) \exp(x'_i \beta)\}^{-2} + \delta_i \text{var}(Z_i) = \{1 + \Lambda_0(C_i) \exp(x'_i \beta)\} \Lambda_0(C_i) \exp(x'_i \beta) \delta_i, \)

\( \text{cov}(Z_{i'l'}, \phi_t) = \gamma_t b_l(C_i) \exp(x'_i \beta) \delta_i, \) and \( \text{cov}(Z_{i'l'}, Z_{i'l'}) = \) the same as in Equation (A.1). The estimated covariance matrix is obtained by evaluating \( \text{var}(\hat{\theta}) \) at \( \hat{\theta} = (\hat{\beta}', \hat{\gamma}')' \).

**Acknowledgements**

We would like to thank the Editor, the Associate Editor, and two anonymous referees for their helpful comments. Implementing their suggestions has improved our paper significantly.
References
