A Pseudo Partial Likelihood Method for Semi-Parametric Survival Regression with Covariate Errors

David M. Zucker

Department of Statistics, Hebrew University

Mt. Scopus, 91905 Jerusalem, Israel

E-mail: mszucker@mscc.huji.ac.il

April 4, 2005

AUTHOR'S FOOTNOTE

David M. Zucker is Associate Professor, Department of Statistics, Hebrew University, Mt. Scopus, 91905 Jerusalem, Israel (email: mszucker@mscc.huji.ac.il). The author thanks Donna Spiegelman for stimulating his interest in the area and for helpful input. The author also thanks the associate editor and referees for helpful comments leading to substantial improvements in the paper. Finally, the author thanks the U.S. National Heart, Lung, and Blood Institute (NHLBI) for providing data from the Framingham Heart Study (FHS), which was conducted and supported by the NHLBI in collaboration with the FHS investigators. The paper was not prepared in collaboration with the FHS investigators and does not necessarily reflect the opinions or views of the FHS or the NHLBI.

ABSTRACT

This paper presents an estimator for the regression coefficient vector in the Cox proportional hazards model with covariate error. The estimator is obtained by maximizing a likelihood-type function similar to the Cox partial likelihood. The likelihood function involves the cumulative baseline hazard function, for which a simple estimator is substituted. The method is capable of handling general covariate error structures: it is not restricted to the independent additive error model. It can be applied to studies with either an external or internal validation sample, and also to studies with replicate measurements of the surrogate covariate. The estimator is shown to be consistent and asymptotically normal, and an estimate of the asymptotic covariance matrix is derived. Some extensions to general transformation survival models are indicated. Simulation studies are presented for a setup with a single error-prone binary covariate and a setup with a single error-prone normally-distributed covariate. These simulation studies show that the method typically produces estimates with low bias and confidence intervals with accurate coverage rates. Efficiency results relative to fully parametric maximum likelihood are also presented. The method is applied to data from the Framingham Heart Study.

Key words: Cox model, proportional hazards, proportional odds, errors in variables

1. INTRODUCTION

Regression analysis of right-censored survival data commonly arises in many fields, especially medical science. The most popular survival data regression model is the Cox (1972) proportional hazards model, in which the hazard function $\lambda(t|\mathbf{x})$ for an individual with covariate vector $\mathbf{x} \in \mathbb{R}^p$ is modeled as

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{x}). \tag{1}$$

Here the baseline hazard function $\lambda_0(t)$ is of unspecified form, so that the model is semi-parametric. Cox developed a simple and elegant procedure for estimating the regression parameter vector $\boldsymbol{\beta}$ based on the notion of partial likelihood, and described likelihood-type asymptotic inference theory for the estimator. Subsequently the procedure was justified rigorously by other authors, including Tsiatis (1981), who used classical limit theory, and Andersen and Gill (1982), who used a martingale theory approach.

In many applications, the covariate $\mathbf{X}$ is not measured exactly, but rather is subject to some degree of measurement error. Thus, instead of observing $\mathbf{X}$, we observe a surrogate measure $\mathbf{Z}$. Starting from Prentice (1982), a considerable literature has developed on inference for the Cox proportional hazards model with covariate error (e.g., Zhou and Pepe, 1995; Huang and Wang, 2000; Xie, Wang, and Prentice, 2001). In the presence of covariate error, the simplicity of the Cox partial likelihood approach is compromised, and most of the methods proposed in the literature are substantially more complex than Cox's procedure. Moreover, the existing methods are subject to substantial limitations. Most work to date has focused on studies with an internal validation sample or situations with independent, additive (often normally distributed) covariate error, or have presented approximate methods involving some asymptotic bias. In particular, the method of Zhou and Pepe (1995) requires internal validation data, while the method of Huang and Wang (2000) requires replicate measurements and covers only the independent additive error model. Recently, Zucker and Spiegelman (2004) presented a consistent estimation procedure for the Cox model under covariate error of arbitrary structure, but their work focused on discrete covariates and relied on a stratification device which limits the applicability of the method.

This paper presents a simple general method for Cox model analysis with arbitrarily structured covariate error. The method is based on a pseudo partial likelihood approach described in Zucker and Yang (2005), which in turn grew out of the work of Yang and Prentice (1999) for the proportional odds model. This approach allows for semi-parametric modelling while avoiding high dimensional optimization. The approach is similar to the method presented recently by Chen, Jin, and Ying (2002) for transformation survival models, but is more closely patterned after Cox's original procedure for the proportional hazards model. In handling the covariate error aspect, some ideas are adapted from Zucker and Spiegelman (2004). There is also some connection with the method of Hu, Tsiatis, and Davidian (1998). We consider the case where the measurement error distribution is either known or estimated from external validation data, internal validation data, or replicate measurement data. In principle, the approach can be extended to cover transformation survival models, and we make some remarks on this extension at an appropriate point in the paper.

Our method overcomes certain limitations of previous methods for this problem proposed in the literature and goes beyond these methods in certain respects. A key advance is the capability of constructing an estimate of the regression coefficients using an estimate of the measurement error distribution that is based on data external to the main survival study. Thus, unlike the methods of Zhou and Pepe (1995), Zhou and Wang (2000), and Chen (2002), there is no need for a substantially-sized internal validation study with both the true covariate and the surrogate measured. Many studies involve an internal validation sample which includes only a very small fraction of the main study, and thus is essentially equivalent to an external validation study. Our method is applicable to this case. Similarly, unlike the methods of Huang and Wang (2000) and Xie et al. (2001), the method can be applied without replicate measurements of the surrogate in the survival study cohort. The regression calibration method (Wang, Hsu, Feng, and Prentice, 1997) and the methods of Hu et al. (1998) can be used with external validation data, but the regression calibration estimator is inconsistent, while the asymptotic behavior of the methods of Hu et al. has not been rigorously analyzed. Nakamura (1992), Kong and Gu (1999), and Hu and Lin (2002) present estimators based on the corrected score concept (Stefanski, 1989; Nakamura, 1990), but these papers deal only with the case of additive measurement error that is independent of the true covariate value. The method presented here provides a proven consistent estimator with external validation data in a very general setup; there does not appear to be any comparable method available elsewhere in the literature.

At the same time, internal validation studies and studies with replicate measurements are handled easily and naturally. In principle, the SIMEX approach (Carroll, Ruppert, and Stefanski, 1995, chap. 4) could be applied to this problem, but SIMEX relies on an extrapolation scheme which is uncertain and does not necessarily yield a consistent estimator. SIMEX also is a somewhat cumbersome multi-stage process, whereas our method can be implemented in a one-shot computer run. Our method allows seamlessly for possible dependence of the distribution of the errors on background covariates. As noted above, it admits an extension to the proportional odds model, for which, except for a paper of Cheng and Wang (2001), there appears to be no other work in the survival covariate error literature. Cheng and Wang's approach, like ours, makes use of the conditional distribution of $\mathbf{X}$ given $\mathbf{Z}$, but is otherwise quite different from our approach. Also, their work focuses on the classical independent additive error model, whereas our work is directed toward general covariate error structures. In short, relative to other methods proposed in the literature, our method has a distinctive flexibility and broadness of applicability.

The plan of the paper is as follows. Section 2 presents the setup and the proposed procedure for the case in which the covariate error distribution is known. Consistency and asymptotic distribution results are given. Section 3 presents simulation studies for the case of a single binary covariate with arbitrary misclassification and for the case of a single normally distributed covariate with normally distributed error. Section 4 discusses the case in which the covariate error distribution is estimated. Section 5 presents the details of this case in the setting of the normal error model with replicate measurements, including relevant theory and a simulation study. Section 6 presents an application to a real data set. Section 7 provides a general discussion. The Appendix gives the details of the theoretical development.

2. SETUP AND PROCEDURE

2.1. The Setup

We assume i.i.d. observations on n individuals. Associated with each individual i is a set of random variables $(T_i^0, T_i^\dagger, \mathbf{X}_i, \mathbf{Z}_i)$, with T_i^0 representing the time to event, $T_i^\dagger$ representing the time to censoring, $\mathbf{X}_i$ representing a p-vector of true covariate values, and $\mathbf{Z}_i$ representing a p-vector of observed covariate values. The observed data consist of $(T_i, \delta_i, \mathbf{Z}_i)$, where $T_i = \min(T_i^0, T_i^\dagger)$ is the follow-up time and $\delta_i = I(T_i^0 \leq T_i^\dagger)$, with $I(\cdot)$ being the indicator function, is the event indicator. In studies with an internal validation sample, some individuals will have both $\mathbf{X}_i$ and $\mathbf{Z}_i$ observed. We denote $S(t|\mathbf{x}) = \Pr(T^0 > t|\mathbf{X} = \mathbf{x})$. We posit the proportional hazards survival model

$$S(t|\mathbf{x}) = \exp(-\Lambda_0(t)\psi(\mathbf{x}; \boldsymbol{\beta})), \tag{2}$$

where Λ_0 is an unknown increasing, differentiable baseline cumulative hazard function of unspecified form and $\psi(\mathbf{x}; \boldsymbol{\beta})$, which involves a p-vector $\boldsymbol{\beta}$ of unknown parameters, expresses the covariate effects. The classical choice is $\psi(\mathbf{x}; \boldsymbol{\beta}) = e^{\boldsymbol{\beta}^T \mathbf{x}}$, but, following Thomas (1981), and Breslow and Day (1993, sec. 5.1(c)), we allow a general covariate effect function $\psi(\mathbf{x}; \boldsymbol{\beta})$. We assume that $\psi(\mathbf{x}; \boldsymbol{\beta})$ satisfies certain technical conditions stated in Appendix A.2 and that $\psi(\mathbf{x}; \mathbf{0}) = 1$ for all $\mathbf{x}$, which means simply that $\boldsymbol{\beta} = \mathbf{0}$ corresponds to no covariate effect. Often it will be desirable to take $\psi(\mathbf{x}; \boldsymbol{\beta})$ to be a function that is monotone in each component of $\mathbf{x}$ for all $\boldsymbol{\beta}$.

We express the conditional distribution of $\mathbf{X}$ given $\mathbf{Z}$ in terms of a conditional density function $\omega(\mathbf{x}|\mathbf{z})$ with respect to a suitable dominating measure m. Typically m will be a product measure with some components in the product being Lebesgue measure, corresponding to error-prone continuous covariates, some components being counting measure, corresponding to error-prone discrete covariates, and some components being point masses, corresponding to covariates measured without error. Actually, with the same theory, we can work with the conditional distribution of $\mathbf{X}$ given $\mathbf{Z}$ and some additional observable random variable $\mathbf{W}$, with the corresponding conditional density being $\omega(\mathbf{x}|\mathbf{z}, \mathbf{w})$. This extension permits incorporation of auxiliary variables that aid prediction of the true covariate value and allows for the possibility that the covariates are measured with varying precision in different subgroups. This possibility would arise in the case where there is a subsample of individuals for which the observed covariate vector is a composite of several successive measurements, and in the case where there are subsets of the population (for instance males and females) over which the precision of the observed covariate is known to vary. In cases where the auxiliary or subgrouping variable affects survival, it is assumed that this variable is included as element of the covariate vector $\mathbf{X}$. Our formulation also covers the case in which there is an internal validation sample with both $\mathbf{X}_i$ and $\mathbf{Z}_i$ measured, and for individuals in this subsample (with a slight abuse of notation) $\mathbf{Z}_i$ is replaced by $\mathbf{X}_i$ in the analysis. To simplify the exposition, we usually will suppress $\mathbf{W}$ from the notation, and just write $\omega(\mathbf{x}|\mathbf{z})$, and so on; i.e., notationally we will subsume $\mathbf{W}$ within $\mathbf{Z}$. At this point, we assume that $\omega(\mathbf{x}|\mathbf{z})$ is known. In Sections 4 and 5 we discuss the case in which $\omega(\mathbf{x}|\mathbf{z})$ is estimated.

We assume throughout that the random vector $(\mathbf{Z}, \mathbf{W})$ and the survival time T^0 are conditionally independent given $\mathbf{X}$. This assumption parallels Assumption (2) of Prentice (1982); note that in Prentice's notation $\mathbf{Z}$ is the true covariate and $\mathbf{X}$ is the proxy, whereas in our notation it is the reverse. This assumption in effect says that the measurement error process is unrelated to the survival time. Finally, we assume that the censoring time $T^\dagger$ is independent of all other random variables in the model.

Define $S_i^*(t) = \Pr(T_i^0 > t|\mathbf{Z}_i, \mathbf{W}_i)$. Under the foregoing setup we have

$$S_i^*(t) = \int \exp(-\Lambda_0(t)\psi(\mathbf{x};\boldsymbol{\beta}))\omega(\mathbf{x}|\mathbf{Z}_i)m(\mathbf{dx}). \tag{3}$$

The corresponding cumulative hazard and instantaneous hazard functions are given by $\Lambda_i^*(t) = -\log S_i^*(t)$ and

$$\lambda_i^*(t) = \frac{d}{dt}\Lambda_i^*(t) = \lambda_0(t)\exp(\phi(\boldsymbol{\beta}, \mathbf{Z}_i, \Lambda_0(t))), \tag{4}$$

where $\lambda_0(t) = \Lambda_0'(t)$, the prime denoting derivative, and

$$\begin{aligned}\phi(\boldsymbol{\beta}, \mathbf{z}, c) &= \log \int \exp(-c\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta})\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}) \\ &\quad - \log \int \exp(-c\psi(\mathbf{x};\boldsymbol{\beta}))\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}). \end{aligned} \tag{5}$$

Note that $\lambda_i^*(t) = \lambda_0(t)E[\psi(\mathbf{X}_i;\boldsymbol{\beta})|\mathbf{Z}_i, T_i^0 \geq t]$, as discussed in Prentice (1982, sec. 2).

2.2. The Procedure

Our procedure is a maximum pseudo partial likelihood estimation (MPPLE) procedure. Assuming temporarily that Λ_0 is known, the analogue of the Cox (1972, 1975) partial likelihood under the induced hazard model (4) is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[\frac{\lambda_i^*(T_i)}{\sum_{j=1}^n Y_j(T_i)\lambda_j^*(T_i)}\right]^{\delta_i} = \prod_{i=1}^n \left[\frac{\exp(\phi(\boldsymbol{\beta}, \mathbf{Z}_i, \Lambda_0(T_i)))}{\sum_{j=1}^n Y_j(T_i)\exp(\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0(T_i)))}\right]^{\delta_i}, \tag{6}$$

where $Y_j(t) = I(T_j \geq t)$, and $\lambda_0(t)$ cancels out. The corresponding normalized log likelihood function is

$$l(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \delta_i \left[\phi(\boldsymbol{\beta}, \mathbf{Z}_i, \Lambda_0(T_i)) - \log \sum_{j=1}^n Y_j(T_i)\exp(\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0(T_i)))\right]. \tag{7}$$

This partial likelihood was discussed in Prentice's (1982) seminal paper. Prentice noted that inference based on this partial likelihood is complicated by its dependence on the unknown baseline cumulative hazard function $\Lambda_0(t)$. Prentice focused on approximate inference when $\boldsymbol{\beta}$ is small, the event is rare, or the covariate error is small (so that $\omega(\mathbf{x}|\mathbf{z})$ is concentrated around $\mathbf{z}$). Prentice did not attempt to develop a general inference procedure based on the partial likelihood for the case where the foregoing approximations do not apply.

Here we propose to substitute for $\Lambda_0(t)$ an estimate $\hat{\Lambda}_0(t,\boldsymbol{\beta})$. By analogy to the standard Breslow (1974) cumulative hazard function estimator for the classical Cox regression model, we propose to estimate Λ_0 (given $\boldsymbol{\beta}$) as a step function with jumps at the ordered observed event times $\tau_k, k = 1,\dots K$, using the equation

$$\Delta\hat{\Lambda}_0(\tau_k) = \frac{d_k}{\sum_{i=1}^n Y_i(\tau_k)\exp(\phi(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1})))}, \tag{8}$$

where d_k is the number of events at time τ_k. In theory, d_k should equal one for all k since T_i^0 is assumed continuous, but we allow here for occasional tied event times. Thus, the quantities $\hat{\Lambda}_0(\tau_1),\dots,\hat{\Lambda}_0(\tau_K)$ are obtained by a simple non-iterative forward recursion rather than by solving a huge K-dimensional system of equations.

The pseudo partial likelihood score equations are given by substituting $\hat{\Lambda}_0(t,\boldsymbol{\beta})$ for $\Lambda_0(t)$ in (7) and setting the derivatives with respect to $\boldsymbol{\beta}$ to zero. This gives the estimating equations $\mathbf{U}(\boldsymbol{\beta},\hat{\Lambda}_0) = \mathbf{0}$, where the normalized score vector $\mathbf{U}$ is given by

$$U_r(\boldsymbol{\beta},\hat{\Lambda}_0) = \frac{1}{n}\sum_{i=1}^n \delta_i \left[\xi_r(\boldsymbol{\beta},\mathbf{Z}_i,T_i) - \frac{\sum_{j=1}^n Y_j(T_i)\xi_r(\boldsymbol{\beta},\mathbf{Z}_j,T_i)e^{\phi(\boldsymbol{\beta},\mathbf{Z}_j,\hat{\Lambda}_0(T_i))}}{\sum_{j=1}^n Y_j(T_i)e^{\phi(\boldsymbol{\beta},\mathbf{Z}_j,\hat{\Lambda}_0(T_i))}}\right], \tag{9}$$

where

$$\xi_r(\boldsymbol{\beta},\mathbf{z},t) = \frac{\partial}{\partial\beta_r}\phi(\boldsymbol{\beta},\mathbf{z},\hat{\Lambda}_0(t,\boldsymbol{\beta})) = \alpha_r(\boldsymbol{\beta},\mathbf{z},\hat{\Lambda}_0(t,\boldsymbol{\beta})) + \nu(\boldsymbol{\beta},\mathbf{z},\hat{\Lambda}_0(t,\boldsymbol{\beta}))Q_r(t,\boldsymbol{\beta}), \tag{10}$$

with

$$\begin{aligned}
\alpha_r(\boldsymbol{\beta},\mathbf{z},c) &= \frac{\partial}{\partial\beta_r}\phi(\boldsymbol{\beta},\mathbf{z},c),\\
\nu(\boldsymbol{\beta},\mathbf{z},c) &= \frac{\partial}{\partial c}\phi(\boldsymbol{\beta},\mathbf{z},c),\\
Q_r(t,\boldsymbol{\beta}) &= \frac{\partial}{\partial\beta_r}\hat{\Lambda}_0(t,\boldsymbol{\beta}).
\end{aligned}$$

The quantities α_r and ν are obtained by straightforward differentiation, and the expressions are given in Appendix Sec. A.1. As for $Q_r(t,\boldsymbol{\beta})$, differentiating the equation (8) gives the following expression for $\Delta Q_r(\tau_k,\boldsymbol{\beta})$:

$$\begin{aligned}
\Delta Q_r(\tau_k,\boldsymbol{\beta}) = -d_k &\left(\sum_{i=1}^n Y_i(\tau_k)\exp(\phi(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1},\boldsymbol{\beta}))\right)^{-2}\\
&\times\left\{\sum_{i=1}^n Y_i(\tau_k)\alpha_r(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1},\boldsymbol{\beta}))\exp(\phi(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1},\boldsymbol{\beta})))\right.\\
&\left.+ Q_r(\tau_{k-1},\boldsymbol{\beta})\left(\sum_{i=1}^n Y_i(\tau_k)\nu(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1},\boldsymbol{\beta}))\exp(\phi(\boldsymbol{\beta},\mathbf{Z}_i,\hat{\Lambda}_0(\tau_{k-1},\boldsymbol{\beta})))\right)\right\}.
\end{aligned} \tag{11}$$

For the theoretical development it is convenient to define, analogously to (10), the quantity $\xi_r(\boldsymbol{\beta}, \mathbf{z}, t, \Lambda)$ for a general function Λ by

$$\xi_r(\boldsymbol{\beta}, \mathbf{z}, t, \Lambda) = \alpha_r(\boldsymbol{\beta}, \mathbf{z}, \Lambda(t)) + \nu(\boldsymbol{\beta}, \mathbf{z}, \Lambda(t)) Q_r(t, \boldsymbol{\beta}, \Lambda),$$

where $Q_r(t, \boldsymbol{\beta}, \Lambda)$ is defined as in (11) (starting from $Q_r(0, \boldsymbol{\beta}, \Lambda) = 0$) with $\hat{\Lambda}_0$ replaced by Λ. We may then define $U_r(\boldsymbol{\beta}, \Lambda)$ analogously to (9) with $\hat{\Lambda}_0$ replaced by Λ and $\xi_r(\boldsymbol{\beta}, \mathbf{z}, t)$ replaced by $\xi_r(\boldsymbol{\beta}, \mathbf{z}, t, \Lambda)$.

Obviously, in the case of the classical Cox regression model with no covariate error, the foregoing procedure reduces to the classical Cox partial likelihood procedure.

A related pseudo full likelihood approach is also possible, but several reasons favored the pseudo partial likelihood approach. First, the pseudo partial likelihood approach seemed likely to be more robust to the estimation of Λ_0. Second, as will be seen, the pseudo partial likelihood approach has an appealing asymptotic theory similar to that of the classical Cox regression model. Third, the partial likelihood may be justified by the same invariance considerations as made in Cox's paper, in that our model is invariant to monotone transformation of the time axis. Finally, in initial explorations the pseudo full likelihood procedure was found to have convergence problems and to yield estimates with higher variance than the pseudo partial likelihood procedure. Therefore, the pseudo partial likelihood approach has been followed here.

2.3. Asymptotic Properties

The Appendix presents an analysis of the asymptotic properties of the procedure. We find, under conditions stated in Appendix Sec. A.2, that the estimate of $\boldsymbol{\beta}$ is strongly consistent and that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is asymptotically mean-zero normal. The normality is established by setting forth the decomposition

$$\begin{aligned} \mathbf{0} &= \mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}})) \\ &= \mathbf{U}(\boldsymbol{\beta}, \Lambda_0(\cdot)) + [\mathbf{U}(\boldsymbol{\beta}, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta})) - \mathbf{U}(\boldsymbol{\beta}, \Lambda_0(\cdot))] \\ &\quad + [\mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}})) - \mathbf{U}(\boldsymbol{\beta}, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}))] \qquad (12) \end{aligned}$$

and analyzing the three terms on the right hand side one by one. The first two terms can be approximated by martingale processes that are independent of each other, while the third term can be approximated by $-\mathbf{V}(\boldsymbol{\beta}, \Lambda_0)(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, where

$$\begin{aligned}\mathbf{V}(\boldsymbol{\beta}, \Lambda) &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{\sum_{j=1}^{n} Y_j(T_i)\boldsymbol{\xi}(\boldsymbol{\beta}, \mathbf{Z}_j, T_i, \Lambda)^{\otimes 2} e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}{\sum_{j=1}^{n} Y_j(T_i) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}\right. \\ &\quad \left. - \left(\frac{\sum_{j=1}^{n} Y_j(T_i)\boldsymbol{\xi}(\boldsymbol{\beta}, \mathbf{Z}_j, T_i, \Lambda) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}{\sum_{j=1}^{n} Y_j(T_i) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}\right)^{\otimes 2}\right], \qquad (13)\end{aligned}$$

with $\mathbf{w}^{\otimes 2} = \mathbf{w}\mathbf{w}^T$. The asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ may be consistently estimated by $\hat{\mathbf{V}}^{-1} + \hat{\mathbf{V}}^{-1}\hat{\mathbf{H}}\hat{\mathbf{V}}^{-1}$, where $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0)$ and $\hat{\mathbf{H}}$ is given in the Appendix (eqn. (30)). Thus, the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is equal to a matrix of a form similar to that which arises for the classical Cox regression model plus an additional term which arises from the estimation of Λ_0. The first term tends to be the dominant term. This first term also reflects the estimation of Λ_0 in that it involves the quantities $Q_r(T_i, \hat{\boldsymbol{\beta}})$.

2.4. Remarks on Extension to Transformation Models

It is possible to consider extension to a survival models of the form

$$S(t|\mathbf{x}) = \exp(-g(h(\Lambda_0(t))\psi(\mathbf{x}; \boldsymbol{\beta}))), \qquad (14)$$

where g is a known strictly increasing function with $g(0) = 0$, and h is the inverse function of g. For example, the choice $g(y) = \log(1 + y)$ gives the proportional odds model of Bennett (1983), which has been studied by several workers. More generally, the model (14) is equivalent to the transformation model of Cheng, Wei, and Ying (1995): we have

$$\Pi(T^0) = -\log\psi(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon \quad \Leftrightarrow \quad \tilde{\Pi}(T^0) = \psi(\mathbf{x}; \boldsymbol{\beta})^{-1}\tilde{T}^0,$$

where $\Pi(t) = \log h(\Lambda_0(t))$, $\tilde{\Pi}(t) = h(\Lambda_0(t))$, $\tilde{T}^0 = \exp(\varepsilon)$, and g is the cumulative hazard function of $\tilde{T}^0$.

For the case of the proportional odds model, our development readily goes through, after suitable modifications in the definitions of ϕ, α_r, and ν, as indicated in Appendix

Sec. A.6. The same applies to the generalized odds rate family of Dabrowska and Doksum (1988), for which $g(y) = r^{-1}\log(1 + ry)$ for some $r > 0$. In the case where r is known, the extension is immediate. In the case where r is unknown, it is necessary to add one more estimating equation, obtained by differentiating the pseudo partial likelihood with respect to r. The development then can be carried through in a straightforward manner. For transformation models with $g'(0) = 0$, which includes many cases of interest, some technical problems arise and a more intricate analysis is needed. Preliminary simulation results for the extended version of our method for the lognormal case have yielded encouraging results. It is hoped to follow up on the extension to more general g in future work.

3. SIMULATION STUDIES

This section presents two simulation studies evaluating the performance of the method. The computations were performed using Fortran programs written by the author. IMSL routines were used for optimization and computation of Gaussian quadrature points and weights. For generating uniform pseudorandom numbers, the IMSL routine DRNUNF (Option 2) was used. In all simulations, the number of simulation replications was 5,000. In both simulation studies, the method converged in 99-100% of the simulation replications.

3.1. Cox Model With Error-Prone Binary Covariate

The setup here is that of Zucker and Spiegelman (2004, sec. 6): the standard Cox model with a single 0–1 binary risk factor subject to misclassification, with error rates assumed to be known. The study horizon is 5 years, the sample size is 2,000, and the 5-year cumulative incidence rate among those unexposed to the risk factor is 25%. The baseline survival distribution was taken to be Weibull, with baseline hazard function $\lambda_0(t) = \vartheta\mu(\mu t)^{\vartheta-1}$. The power parameter ϑ was taken equal to 5, which is typical of many types of cancer (Armitage and Doll, 1961; Breslow and Day, 1993, sec. 6.3). The scale parameter μ was chosen so as to yield the specified cumulative incidence rate for the unexposed population. Censoring was taken to be exponential with a rate of 1% per year. For brevity of presentation, the false positive rate $\Pr(Z = 1|X = 0)$ and the false negative rate $\Pr(Z = 0|X = 1)$ were taken to be equal to a common classification error rate. A range of values was used for the prevalence of the risk factor (5%, 25%, 40%), the classification error rate (1%, 5%, 10%, 20%), and the true relative risk (1.5, 2.0). For comparison, we also present the simulation results given by Zucker and Spiegelman for the naive Cox (1972) partial likelihood estimator ignoring the measurement error and for the parametric log relative risk estimator obtained by maximizing the full Weibull log likelihood under the relevant measurement error model.

Table 1 shows the results. The naive Cox estimator is typically badly biased except under 1% misclassification with exposure prevalence of 25% or 40%. By contrast, our method exhibits excellent performance, comparable to that of the fully parametric Weibull estimator. For the cases with an exposure prevalence of 25% or 40%, our method yields nearly zero bias in the estimated log relative risk, accurate estimates of the variance of the estimated log relative risk, and accurate confidence interval coverage. With an exposure prevalence of 5%, the performance of all three estimators under consideration is degraded. This finding is not surprising given that the expected number of events in the exposed group in these cases is only on the order of 25–50 and that with an exposure prevalence of 5% and a misclassification rate of 5% or more the predictive value of an observed positive exposure is low. The naive Cox estimator is drastically biased. Our estimator and the Weibull estimator are dramatically less biased, though they do exhibit some degree of bias. This bias is due in part to outlying values; for both our estimator and the Weibull estimator, the deviation between the median value of the estimate and the true log relative risk is noticeably lower than the deviation between the mean estimated value and the true value. Overall, in terms of mean square error, the performance of our estimator is typically essentially identical to that of Weibull estimator. In a few cases our estimator is better, reflecting the fact that, in a finite-sample situation, it is possible for the asymptotically optimal parametric MLE to be outperformed by an alternate estimator.

3.2. Cox Model With Error-Prone Normal Covariate

The setup here is the standard Cox model with a single continuous covariate X distributed $N(0, \sigma_X^2)$, with the measured value Z given by $Z = X + \epsilon$, where the error term ϵ is independent of X and distributed $N(0, \sigma_\epsilon^2)$. This setup has been explored by a number of other researchers (e.g., Prentice, 1982; Hughes, 1993; Huang and Wang, 2000; Xie, Wang, and Prentice, 2001). Technically, this setup violates the assumption of bounded X and Z made in Appendix Sec. A.2, but obviously a finite range can be defined such that the probability that X or Z in this setup lie outside this range is negligible. The various integrals involved in the estimation procedure were evaluated by 20-point Gauss-Hermite quadrature. The configurations studied here are patterned after those of Xie, Wang, and Prentice's Table 1: n equal to 150 or 300, unit exponential baseline survival, common censoring at time 1, β equal to 0, 0.6931 (= log 2), or 1.3863 (= log 4), $\sigma_X^2 = 1$, and σ_ϵ^2 equal to 0.25, 0.50, or 1.00. Xie et al. considered n=150. The error variance values of 0.25 and 0.50 are parallel to the Xie et al. setup, which involved the average of four replicate measurements with error variance of 1 or 2. The error variance value of 1 represents an extreme case where the error variance of the covariate value entering into the analysis is equal to the population variance of the true covariate. In our simulations, we constrain the estimate of β to be of absolute value less than 3, and consider cases where the boundary is hit as cases of nonconvergence. This constraint effects less than 1% of the simulation replications. When the constraint is relaxed, the performance of the method is generally similar, except for some isolated outlying values that effect the variance and the accuracy of the variance estimation for n=150 and $\beta = \log 4$.

Tables 2 and 3 present, for $n = 150$ and $n = 300$, respectively, results under the foregoing scenarios for the naive Cox estimator and our estimator. In cases where comparison is possible, our results on the bias of the naive Cox estimator agree with those presented in Xie et al.'s Table 1. We also include relative mean square error results relative to parametric maximum likelihood for the full parametric exponential model and for a Weibull model (ratio of the MSE of the comparison estimator to the MSE of our estimator).

As in the preceding simulation study, the naive Cox estimator is typically badly biased, whereas our estimator generally exhibits little bias. For $n = 150$ and $\beta = \log 4$, the bias of our estimator is somewhat more pronounced, but still not more than 5% of the true value in the worst case. The variance of the estimated coefficient was generally well estimated, and the confidence interval coverage was quite accurate. The relative efficiency of our estimator relative to the exponential MLE was nearly 100% for $\beta = 0$, in the range 70-90% for $\beta = 0.6931$, and in the range 25-60% for $\beta = 1.3863$ (for reference, in the case of no measurement error, we found the relative efficiency of the Cox partial likelihood estimator to the exponential MLE to be about 95% for $\beta = 0.6931$ and about 77% for $\beta = 1.3863$). The relative efficiency of our estimator relative to the Weibull MLE was nearly 100% in most cases. In cases where comparison with Xie et al.'s estimate was possible from results in Xie et al.'s Table 1, our estimator and the Xie et al. estimator performed comparably, with our estimator having somewhat lower bias is some cases but somewhat higher variance in some cases. In Sec. 5 below, we sharpen the comparison with the results of Xie et al. by considering the case where σ_X^2 and σ_ϵ^2 are estimated from replicate measurements as in Xie et al. rather than being known.

4. ESTIMATED $\omega(\mathbf{x}|\mathbf{z})$

We now consider the case where $\omega(\mathbf{x}|\mathbf{z})$ is estimated. We work under the typical framework of a parametric model $\omega(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is an unknown parameter vector to be estimated. An estimate of $\boldsymbol{\theta}$ may be obtained from any one of the following three types of data (Carroll, Ruppert, and Stefanski, 1995, sec. 1.4):

(a) An external validation sample of size n^* with both $\mathbf{X}_i$ and $\mathbf{Z}_i$ measured.

(b) An internal validation sample, consisting of a subset of size n^* of the main survival study in which both $\mathbf{X}_i$ and $\mathbf{Z}_i$ are measured.

(c) A sample of size n^* (either external or internal) with replicate measurements of $\mathbf{Z}_i$.

A number of studies have made use of internal validation data. One example is the Nurses' Heath Study (Hu et al., 1997), where fat intake was measured using the surrogate Food Frequency Questionnaire in all subjects and using the more accurate Diet Recall interview in a subsample of subjects. Another example is the SOLVD heart failure survival trial discussed by Zhou and Pepe (1995), where left ventricular ejection fraction was measured using an error-prone non-standardized method in all patients and using a more accurate standardized method in a subset of patients. External validation data were used in the Pooling Project on Diet and Cancer (Hunter et al., 1996): here data from the Nurses' Health Study validation sample were used as the validation study for the Iowa Women's Health Study and the New York State Study. As an example of a study with replicate measurements, we mention the AIDS trial discussed by Huang and Wang (2000), which made use of replicate baseline CD4 measurements.

If external data are used, it is of course necessary that the relevant parameters be the same (to a reasonable approximation) for the external population as for the main study population (see Carroll, Ruppert, and Stefanski, 1995, secs. 1.3.2 and 1.3.3). In many cases, it will be desirable to estimate some elements of $\boldsymbol{\theta}$ (e.g. a measurement error variance) from external data and other elements (e.g. the mean and variance of the measured covariate $\mathbf{Z}$) from the data in the main study. A number of cohort studies with low event rate and sample size of several thousand or more have incorporated an internal validation sample with a sample size in the hundreds. For these studies, the internal validation sample functions in effect as an external validation study, with a sample that may readily be regarded as representative of the main study population. The aforementioned Nurses' Health Study falls into this category.

In Sec. 5 below we give a detailed treatment of the replicate data setup under the normal error model of Sec. 3.2.

Whatever method is used to estimate $\boldsymbol{\theta}$, in general we will have an estimate $\hat{\boldsymbol{\theta}}$ that is asymptotically normal with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Omega}/n^*$, along with an estimator of the matrix $\boldsymbol{\Omega}$. For example, for the case of a single 0–1 binary covariate, the estimates of $\theta_k = \Pr(X = k-1|Z = k-1), k = 1, 2$, are given by the obvious sample proportions, and $\boldsymbol{\Omega}$ is a 2×2 diagonal matrix with $\Omega_{kk} = \theta_k(1-\theta_k)/\pi_k$, where π_k is the probability that $Z = k-1$ in the validation study. For the asymptotics we assume that n^* and n are of the same order of magnitude, i.e., $n^*/n \to \varrho$ for some constant ϱ as $n \to \infty$; otherwise the error in $\boldsymbol{\theta}$ will either be dominated by or will dominate the other components of error in $\hat{\boldsymbol{\beta}}$. Typically ϱ will be between 0 and 1.

To account for the error in $\boldsymbol{\theta}$, we have to add to the representation (12) the additional term $\mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}}), \hat{\boldsymbol{\theta}}) - \mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}}), \boldsymbol{\theta})$. This term can be approximated by $\mathbf{F}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\}$, where F_{rs} is the limiting value of $\partial U_r/\partial \theta_s$. We then get

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \;\doteq\;& \mathbf{V}(\boldsymbol{\beta}, \Lambda_0)^{-1}\Big[\mathbf{U}(\boldsymbol{\beta}, \Lambda_0(\cdot), \boldsymbol{\theta}) \\ &+ [\mathbf{U}(\boldsymbol{\beta}, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}), \boldsymbol{\theta}) - \mathbf{U}(\boldsymbol{\beta}, \Lambda_0(\cdot), \boldsymbol{\theta})] + \mathbf{F}\{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\}\Big]. \qquad (15)\end{aligned}$$

Obviously if $\boldsymbol{\theta}$ is estimated entirely from external data, the last term in the brackets in (15) will be independent of the first two terms. In the case where internal data are used, it comes out that this last term is asymptotically independent of the first two terms as well. This result follows from the fact that the first two terms can be approximated by martingale processes, as indicated in the Appendix, whereas $\hat{\boldsymbol{\theta}}$ is based on $\mathbf{Z}$'s and $\mathbf{X}$'s, or replicate $\mathbf{Z}$'s, which are known at time zero. Thus, in either case, the correction needed to account for the error in $\boldsymbol{\theta}$ is to add to the estimated asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ the term $\varrho^{-1}\hat{\mathbf{V}}^{-1}\hat{\mathbf{F}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{F}}^T\hat{\mathbf{V}}^{-1}$. By an argument similar to that made in the Appendix in connection with the second and third terms of (12), it is seen that F_{rs} may be consistently estimated by $\hat{F}_{rs}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0)$, where

$$\hat{F}_{rs}(\boldsymbol{\beta}, \Lambda) = -\frac{1}{n}\sum_{i=1}^{n}\delta_i\left[\frac{\sum_{j=1}^{n} Y_j(T_i)\xi_r(\boldsymbol{\beta}, \mathbf{Z}_j, T_i, \Lambda)\kappa_s(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}{\sum_{j=1}^{n} Y_j(T_i)e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}\right.$$

$$- \left(\frac{\sum_{j=1}^n Y_j(T_i) \xi_r(\boldsymbol{\beta}, \mathbf{Z}_j, T_i, \Lambda) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}{\sum_{j=1}^n Y_j(T_i) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}} \right) \left(\frac{\sum_{j=1}^n Y_j(T_i) \kappa_s(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i)) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}}{\sum_{j=1}^n Y_j(T_i) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(T_i))}} \right) \Bigg],$$

where

$$\begin{aligned} \kappa_s(\boldsymbol{\beta}, \mathbf{z}, c) &= \frac{\partial}{\partial \theta_s} \phi(\boldsymbol{\beta}, \mathbf{z}, c) \\ &= \frac{\int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \psi(\mathbf{x};\boldsymbol{\beta}) \dot{\omega}_{is}(\mathbf{x}|\mathbf{z}) m(\mathbf{dx})}{\int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \psi(\mathbf{x};\boldsymbol{\beta}) \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx})} - \frac{\int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \dot{\omega}_{is}(\mathbf{x}|\mathbf{z}) m(\mathbf{dx})}{\int e^{-c\psi(x;\boldsymbol{\beta})} \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx})}, \end{aligned}$$

with $\dot{\omega}_{is}(\mathbf{x}|\mathbf{z})$ denoting the partial derivative of $\omega(\mathbf{x}|\mathbf{z}, \boldsymbol{\theta})$ with respect to θ_s.

5. NORMAL ERROR MODEL WITH REPLICATE MEASUREMENTS

Here we illustrate the methodology of the preceding section in the case of the normal error model with replicate measurements. For concreteness, we consider the case of replicate measurements on all individuals in the main survival study, though replicate measurements on a subsample of the main study or on an external sample also can be readily handled.

5.1. Setup and procedure

We assume the true covariate X_i has a $N(\mu, \sigma_X^2)$ distribution, and that for each individual we have two replicate surrogate measurements $\tilde{Z}_{i1}$ and $\tilde{Z}_{i2}$. We take $\tilde{Z}_{ij} = X_i + \epsilon_{ij}$, where the ϵ_{ij}'s are distributed $N(0, \sigma_\epsilon^2)$, independently of each other and of X_i. We define the overall surrogate value of X_i for individual i to be $Z_i = \frac{1}{2}(\tilde{Z}_{i1} + \tilde{Z}_{i2})$. We set $\sigma_Z^2 = \sigma_X^2 + \frac{1}{2}\sigma_\epsilon^2$ and $\sigma_{\bar{\epsilon}}^2 = \frac{1}{2}\sigma_\epsilon^2$.

Standard normal theory shows that the conditional distribution of X given $Z = z$ is normal with mean $\mu(x|z) = \mu + a(z - \mu)$ and variance $\varsigma^2 = \sigma_X^2(1-a) = \sigma_{\bar{\epsilon}}^2(1 - \sigma_{\bar{\epsilon}}^2/\sigma_Z^2)$, where $a = \sigma_X^2/\sigma_Z^2 = (1 - \sigma_{\bar{\epsilon}}^2/\sigma_Z^2)$. Thus

$$\omega(x|z) = (2\pi\varsigma^2)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\varsigma^2}(x - \mu(x|z))^2 \right).$$

The three parameters that determine this conditional distribution are $\theta_1 = \mu$, $\theta_2 = \sigma_Z^2$, and $\theta_3 = \sigma_\epsilon^2$. The obvious estimators of these parameters are

$$\begin{aligned}
\hat{\mu} &= \bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i \\
\hat{\sigma}_Z^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2 \\
\hat{\sigma}_{\bar{\epsilon}}^2 &= \frac{1}{4n}\sum_{i=1}^{n}(Z_{1i} - Z_{2i})^2.
\end{aligned}$$

By standard theory, these three estimators are independent and unbiased, with the $\sqrt{n}$-normalized difference between the estimates and the true parameter values being asymptotically normal with respective variances $\sigma_Z^2, 2\sigma_Z^4$, and $2\sigma_{\bar{\epsilon}}^4$. The derivatives $\dot{\omega}_s(x|z) = (\partial/\partial\theta_s)\omega(x|z)$ are as follows.

$$\begin{aligned}
\frac{\partial}{\partial\mu}\,\omega(x|z) &= \varsigma^{-2}(1-a)(x-\mu(x|z))\omega(x|z), \\
\frac{\partial}{\partial\sigma_Z^2}\,\omega(x|z) &= \Bigg[-\frac{1}{2}\left(\frac{1}{\varsigma^2}\right) + \frac{1}{2}\left(\frac{1}{\varsigma^2}\right)^2 (x-\mu(x|z))^2 \\
&\quad + \left(\frac{1}{\varsigma^2}\right)\left(\frac{1}{\sigma_{\bar{\epsilon}}^2}\right)(x-\mu(x|z))(z-\mu)\Bigg]\left(\frac{\sigma_{\bar{\epsilon}}^2}{\sigma_Z^2}\right)^2 \omega(x|z), \\
\frac{\partial}{\partial\sigma_{\bar{\epsilon}}^2}\,\omega(x|z) &= \Bigg[-\frac{1}{2}\left(\frac{1}{\varsigma^2}\right)\left(1-\frac{2\sigma_{\bar{\epsilon}}^2}{\sigma_Z^2}\right) + \frac{1}{2}\left(\frac{1}{\varsigma^2}\right)^2\left(1-\frac{2\sigma_{\bar{\epsilon}}^2}{\sigma_Z^2}\right)(x-\mu(x|z))^2 \\
&\quad - \left(\frac{1}{\varsigma^2}\right)\left(\frac{1}{\sigma_Z^2}\right)(x-\mu(x|z))(z-\mu)\Bigg]\omega(x|z).
\end{aligned}$$

Given the foregoing results, the procedure described in the preceding section, including the variance correction, may be implemented in a straightforward manner.

5.2. Simulation results

As in Sec. 3.2, the simulation scenarios are patterned after those of Xie et al. (2001, Table 1). We take n equal to 150 or 300, unit exponential baseline survival, common censoring at time 1, β equal to 0, 0.6931 (= log 2), or 1.3863 (= log 4), $\mu = 0$, $\sigma_X^2 = 1$, and σ_ϵ^2 equal to 0.5, 1, or 2. The first two error variance values we consider are parallel to the Xie et al. setup: Xie et al. dealt with four replicate measurements with error variance of 1 or 2, while our first two cases involve two replicate measurements with error variance of 0.5 or 1. The estimation procedure converged in 99-100% of the simulation replications in all configurations except that with $n = 150, \sigma_\epsilon^2 = 2$, and $\beta = \log 4$, in which the percent convergence was 93%.

Table 4 presents the results. Our method generally performs very well. In most cases, the bias is low and the variance estimation is reasonably accurate. With n=150, $\beta = \log 4$, and the extreme error variance of 2, the variance estimate is noticeably off the mark, but this problem clears up for n=300. The 95% confidence interval coverage is generally reasonably accurate. In comparison with the results of Sec. 3.2, the variances are slightly larger, as expected due to the estimation of the error distribution parameters. Comparing our results with those in Xie et al.'s Table 1, we see that our method performs in a generally similar manner. With $\beta = \log 4$, our method tends to have slightly lower bias. For $\beta = \log 4$ and an $\sigma_Z^2 = 0.50$, our estimate has higher variance. For $\beta = \log 4$, our method has more accurate 95% confidence interval coverage.

6. EXAMPLE

We illustrate our method on a data set from the Framingham Heart Study (Gordon and Kannel, 1968) similar to that used by Xie et al. (2001) in their example, but with event times that are not so heavily tied. The data set comprises 664 men aged 35-44 with no history of high blood pressure or cardiovascular disease at the beginning of the study. We examine the effect of a subject's long-term underlying systolic blood pressure (SBP) level on the risk of cardiovascular disease death, using follow-up data up to Framingham Study Exam 24, for a maximum follow-up of about 48 years. There were 208 events. Following Xie et al., the covariate X we use is a transformed version of SBP suggested by Cornfield (1962) and defined by TSBP=log((SBP-75)/25), which has been found to be approximately normally distributed. The true underlying TSBP is unknown, and the proxy Z for X we use is the average of the TSBP values from the first two exams, which were 2 years apart. The analysis is conducted within a framework similar to that of Sec. 3.2, with X taken to be distributed $N(\mu, \sigma_X^2)$ and Z assumed to be given by $Z = X + \epsilon$, with the error term ϵ being independent of X and distributed $N(0, \sigma_\epsilon^2)$. From the two initial TSBP values, μ is estimated to be equal to 0.71, σ_X^2 is estimated to be equal to 0.045, and σ_ϵ^2 is estimated to be equal to 0.013 (about 29% of the estimated σ_X^2). For purposes of this illustration, we shall regard these values as known values, which is reasonable since the data set is large. The relative risk function used is the classical one, $\psi(x; \beta) = e^{\beta x}$. The table below presents the estimated regression coefficients and corresponding standard errors for the following methods: (1) the naive approach in which a standard Cox analysis is run with X replaced by Z, (2) the "ordinary" regression calibration (ORC) approach in which a standard Cox analysis is run with X replaced by $E[X|Z]$ (cf. Wang et al., 1997), (3) the Xie et al. risk regression calibration (RRC) method, (4) the Huang and Wang (2000) nonparametric corrected score estimate (NPCSE), and (5) our MPPLE analysis. For the naive, ORC, and RRC methods, the standard errors were computed in a simple fashion using the standard Cox variance formula (with the relevant imputed covariate value) rather than the more complex rigorous formulas of Xie et al.

	Naive	ORC	RRC	NPCSE	MPPLE
Estimate	1.764	2.277	2.234	2.411	2.315
Std Error	0.319	0.412	0.408	0.373	0.429

We find that the naive estimate is considerably lower than the estimates produced by the methods that correct for measurement error. The methods correcting for measurement error give roughly similar results, consistent with the moderate degree of measurement error in this data set, but there are nonetheless nontrivial differences in the values obtained. The MPPLE estimate lies about halfway between the RRC estimate and the NPCSE estimate.

7. DISCUSSION

We have presented an estimator for the regression coefficient vector in the Cox (1972) proportional hazards model with covariate error. The procedure involves evaluating the relative risk function given the observed covariate value and maximizing a likelihood-type function similar in form to the Cox (1972, 1975) partial likelihood. Here it is necessary to plug in an estimator of the baseline cumulative hazard function, which in the present context enters into the partial likelihood function.

To estimate the cumulative baseline hazard, we use a simple and easily computed Breslow-type estimator. The resulting pseudo partial likelihood method exhibits very reasonable performance. In particular, our simulation results showed that our procedure exhibits a high degree of efficiency relative to parametric maximum likelihood estimation under a known parametric model, except for the narrow one-parameter exponential model. Thus, it does not appear that an alternate method for estimating the cumulative hazard would offer any substantial advantage. This feeling is borne out by some recent work the author has been involved with in the setting of frailty models (Gorfine, Zucker, and Hsu, 2005). In this work, a pseudo-likelihood approach with a Breslow-type plug-in cumulative hazard estimator was found to exhibit essentially the same performance as full semiparametric maximum likelihood estimation.

As indicated in the introduction, the method overcomes certain limitations of other methods proposed in the literature for this problem and has a distinctive flexibility and broadness of applicability. In particular, the method is not restricted to classical additive error models, but is capable of handling general covariate error structures. Accommodation of general error structures is of particular importance for handling binary and categorical covariates that are subject to error. It is also of relevance in cases involving a continuous error-prone covariate where the error variance can depend on the underlying true covariate value (e.g., a disease marker may be harder to measure among patients in worse condition). Initial steps toward extending the method to general transformation survival models have been presented; in particular the proportional odds model and generalized odds rate models are covered.

The method presented here works with the conditional density $\omega(\mathbf{x}|\mathbf{z})$ of the true covariate $\mathbf{X}$ given the surrogate $\mathbf{Z}$. We use a parametric model to describe this distribution. The relevant parameters can be estimated using external validation data, internal validation data, or replicate measurements. General theory covering all three cases is presented. The fact that our approach accommodates external validation data is a significant advance, since there are few methods in the literature dealing with this case, and apparently none with the theoretical rigor and and broad generality that our method possesses.

The assumption that $\omega(\mathbf{x}|\mathbf{z})$ is known up to a finite number of specified parameters admittedly imposes a restriction on the applicability of our method. The methods of Zhou and Pepe (1995), Zhou and Wang (2000), and Huang and Wang (2000) do not require this assumption. At the same time, however, these methods require other assumptions that our method does not. The Zhou-Pepe method and the Zhou-Wang method require internal validation data, while the Huang-Wang method requires replicate measurements and is restricted to an independent additive error model. As noted above, the restriction to the independent additive error model is a substantial limitation.

Parametric models of the kind we assume are commonly used in practice in the measurement error setting. Appropriate steps can be taken to check the parametric model. When validation data with true $\mathbf{X}$ values are available, the model can be developed and checked directly by regression methods, without need to deal with the marginal distribution of $\mathbf{X}$ or $\mathbf{Z}$. In the replicate measures case, parametric forms can be assumed for the marginal distribution of $\mathbf{X}$ and the conditional distribution of $\mathbf{Z}$ given $\mathbf{X}$, and the resulting marginal distribution of $\mathbf{Z}$ can be derived. The parametric form for the marginal distribution of $\mathbf{Z}$ can be checked against the observed $\mathbf{Z}$ data, while the parametric form for the distribution of $\mathbf{Z}$ given $\mathbf{X}$ can be checked by analyzing differences among the replicate data. Moreover, the restrictiveness of the parametric modelling setup can be alleviated by using a flexible parametric model. Also, in the case of a binary or categorical error-prone covariate with unstructured misclassification matrix, a case of interest in a number of epidemiological applications, our parametric model setup automatically holds.

In principle, one might also consider nonparametric estimation of the conditional density $\omega(\mathbf{x}|\mathbf{z})$. Zhou and Wang (2000) used a kernel smoothing approach in the context of a method designed for an internal validation setup. They showed that, even with the nonparametric smoothing embedded in their procedure, their estimator for the regression coefficient vector converges to the true value at the classical $O(n^{-\frac{1}{2}})$ rate and is asymptotically normal. It should be possible to adapt their techniques of proof to obtain a similar result for our approach. This is a potential topic for further work.

Insofar as the approach presented here involves numerical integration over the distribution of the true covariate given the observed covariate, it requires a certain degree of computational effort. However, when the number of continuous covariates subject to error is not too large, as is the case in typical applications, the computational burden is relatively modest. For example, for the Framingham example presented in Section 6, with a single continuous covariate, 668 subjects, and 208 events, the method ran on a SunOS 5.8 mainframe in about 1 minute of real time. In principle, when there are many continuous covariates that are subject to error, Markov Chain Monte Carlo methods could be used to evaluate the integrals.

Simulation studies showed that the procedure typically produces estimates with low bias and confidence intervals with accurate coverage rates. The convergence performance was excellent: in the simulation studies presented, the method typically converged in 99-100% of the simulation replications. Efficiency comparisons relative to fully parametric maximum likelihood were also undertaken. In the simulation scenarios studied, except in cases where the parametric model used was a narrowly-specified one-parameter family and the covariate effect was moderate to large, our estimator was found to have good efficiency relative to the fully parametric maximum likelihood estimator.

APPENDIX: ASYMPTOTIC THEORY OF THE ESTIMATOR

A.1. Preliminaries

We begin with some definitions. Throughout, for emphasis, we denote the true value of $\boldsymbol{\beta}$ by $\boldsymbol{\beta}^\circ$, the true $\Lambda_0(t)$ by $\Lambda_0^\circ(t)$, and the true $\lambda_0(t)$ by $\lambda_0^\circ(t)$ We denote the maximum follow-up time by τ. With ϕ as in (5) we have

$$\begin{aligned}
\alpha_r(\boldsymbol{\beta}, \mathbf{z}, c) &= \frac{\partial}{\partial \beta_r} \phi(\boldsymbol{\beta}, \mathbf{z}, c) = B_2^{-1} B_1 + B_4^{-1} B_3, \\
\nu(\boldsymbol{\beta}, \mathbf{z}, c) &= \frac{\partial}{\partial c} \phi(\boldsymbol{\beta}, \mathbf{z}, c) = B_2^{-1} \tilde{B}_1 + B_4^{-1} B_2,
\end{aligned}$$

where, with $\dot{\psi}_r(\mathbf{x}; \boldsymbol{\beta})$ denoting the derivative of $\psi(\mathbf{x}; \boldsymbol{\beta})$ with respect to β_r,

$$\begin{aligned}
B_1 &= \int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \dot{\psi}_r(\mathbf{x}; \boldsymbol{\beta})[1 - c\psi(\mathbf{x}; \boldsymbol{\beta})] \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx}), \\
\tilde{B}_1 &= -\int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \psi(\mathbf{x}; \boldsymbol{\beta})^2 \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx}), \\
B_2 &= \int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \psi(\mathbf{x}; \boldsymbol{\beta}) \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx}), \\
B_3 &= c \int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \dot{\psi}_r(\mathbf{x}; \boldsymbol{\beta}) \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx}), \\
B_4 &= \int e^{-c\psi(\mathbf{x};\boldsymbol{\beta})} \omega(\mathbf{x}|\mathbf{z}) m(\mathbf{dx}).
\end{aligned}$$

Further, we define $N_i(t) = I(T_i \le t, \delta_i = 1)$ and let $M_i(t)$ be the corresponding counting process martingale, given by

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{\phi(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, \Lambda_0^\circ(s))} \lambda_0^\circ(s) ds.$$

We then may write

$$\mathbf{U}(\boldsymbol{\beta}, \Lambda) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{\xi}(\boldsymbol{\beta}, \mathbf{Z}_i, t, \Lambda) - \frac{\sum_{j=1}^n Y_j(t) \boldsymbol{\xi}(\boldsymbol{\beta}, \mathbf{Z}_j, t, \Lambda) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(t))}}{\sum_{j=1}^n Y_j(t) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda(t))}} \right] dN_i(t),$$

and, as in Andersen and Gill (1982, p. 1103), we have

$$\mathbf{U}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left[\boldsymbol{\xi}(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, t, \Lambda_0^\circ) - \frac{\sum_{j=1}^n Y_j(t) \boldsymbol{\xi}(\boldsymbol{\beta}^\circ, \mathbf{Z}_j, t, \Lambda_0^\circ) e^{\phi(\boldsymbol{\beta}^\circ, \mathbf{Z}_j, \Lambda_0^\circ(t))}}{\sum_{j=1}^n Y_j(t) e^{\phi(\boldsymbol{\beta}^\circ, \mathbf{Z}_j, \Lambda_0^\circ(t))}} \right] dM_i(t). \tag{16}$$

A.2. Assumptions

We make the following assumptions (for completeness, we repeat the assumptions stated in the text):

I. The setup is i.i.d.

II. $(\mathbf{Z}, \mathbf{W})$ and T^0 are conditionally independent given $\mathbf{X}$, and $T^\dagger$ is independent of all other random variables in the model.

III. The function $\Lambda_0^\circ(t)$ is strictly increasing and differentiable with derivative $\lambda_0^\circ(t)$.

IV. $\mathbf{X}$ and $\mathbf{Z}$ are bounded and ω is bounded. The parameter $\boldsymbol{\beta}$ lies in a compact set $\boldsymbol{\mathcal{B}}$ of $\mathrm{I\!R}^p$ which includes an open neighborhood of $\boldsymbol{\beta}^\circ$.

V. The function $\psi(\mathbf{x};\boldsymbol{\beta})$ satisfies $\psi(\mathbf{x};\mathbf{0}) = 1$ for all $\mathbf{x}$, is twice continuously differentiable with respect to $\boldsymbol{\beta}$ over $\boldsymbol{\mathcal{B}}$, and is bounded from below by a positive number ψ_{min} for all $\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}$ and all $\mathbf{x}$ in the relevant bounded domain of $\mathbf{x}$ values.

The continuous differentiability condition implies that there exists also an upper bound ψ_{max} on $\psi(\mathbf{x};\boldsymbol{\beta})$ over the relevant domain of $\boldsymbol{\beta}$ and $\mathbf{x}$.

VI. $y^* = \Pr(Y_i(\tau) = 1) > 0$.

VII. The limiting value $\mathbf{v}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ)$ of the matrix $\mathbf{V}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ)$ defined in (13) is positive definite (existence of the limit is justified by the considerations at the beginning of the next section).

VIII. The baseline hazard function $\lambda_0^\circ(t)$ is bounded over $[0, \tau]$ by some constant λ_{max}.

IX. The censoring distribution has at most a finite number of jumps on $[0, \tau]$.

The foregoing assumptions imply that the functions $\phi(\boldsymbol{\beta}, \mathbf{z}, c)$, $\alpha_r(\boldsymbol{\beta}, \mathbf{z}, c)$, and $\nu(\boldsymbol{\beta}, \mathbf{z}, c)$ are bounded over $\boldsymbol{\beta}$, $\mathbf{z}$, and c for values of these arguments in the relevant domain (for c this domain is taken to be $[0, \Lambda_{max}]$, where we define $\Lambda_{max} = 1.01/(y^*\psi_{min})$).

Moreover, these functions are Lipschitz continuous with respect to c uniformly in $\boldsymbol{\beta}$, $\mathbf{z}$, and c over this domain. It is easy to check that

$$\psi_{min} \leq e^{\phi(\boldsymbol{\beta},\mathbf{z},c)} \leq \psi_{max}, \tag{17}$$

$$\left|\frac{\partial}{\partial c} e^{\phi(\boldsymbol{\beta},\mathbf{z},c)}\right| \leq \psi_{max}^2, \tag{18}$$

for $\boldsymbol{\beta}$ and $\mathbf{z}$ in the relevant domain and all $c \in \mathbb{R}$.

A.3. Consistency

To begin, consider the quantities

$$\begin{aligned}
A_0(\boldsymbol{\beta}, t, c) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t) e^{\phi(\boldsymbol{\beta},\mathbf{Z}_i,c)}, \\
A_{1r}(\boldsymbol{\beta}, t, c) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t) \alpha_r(\boldsymbol{\beta}, \mathbf{Z}_i, c) e^{\phi(\boldsymbol{\beta},\mathbf{Z}_i,c)}, \\
A_2(\boldsymbol{\beta}, t, c) &= \frac{1}{n}\sum_{i=1}^{n} Y_i(t) \nu(\boldsymbol{\beta}, \mathbf{Z}_i, c) e^{\phi(\boldsymbol{\beta},\mathbf{Z}_i,c)},
\end{aligned}$$

and denote by a_0, a_{1r}, a_2 the corresponding expectations. The development of the consistency and asymptotic normality results involves empirical processes such as the A's above and other processes of similar form. By the functional strong law of large numbers given in Andersen and Gill (1982, Appendix III), we find that, uniformly over $\boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}$, $t \in [0, \tau]$, and $c \in [0, \Lambda_{max}]$, such processes converge to their corresponding expectations almost surely.

We now address the asymptotic behavior of $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ for given $\boldsymbol{\beta}$. Denote, for a general function Λ,

$$\begin{aligned}
\Upsilon_n(t, \boldsymbol{\beta}, \Lambda) &= \int_0^t \frac{n^{-1}\sum_{i=1}^{n} dN_i(s)}{A_0(\boldsymbol{\beta}, s, \Lambda(s-))}, \\
\Upsilon(t, \boldsymbol{\beta}, \Lambda) &= \int_0^t \frac{E[Y_i(s) e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}]}{a_0(\boldsymbol{\beta}, s, \Lambda(s-))} \lambda_0^\circ(s) ds.
\end{aligned}$$

Then $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ is the solution to the equation $\Lambda(t) = \Upsilon_n(t, \boldsymbol{\beta}, \Lambda)$ subject to $\Lambda(0) = 0$. Using (17) and (18), plus the boundedness of λ_0°, we find that the function

$$q_{\boldsymbol{\beta}}(s, c) = \frac{E[Y_i(s) e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}]}{a_0(\boldsymbol{\beta}, \mathbf{Z}_i, c)} \lambda_0(s)$$

satisfies the following two conditions: (a) boundedness over s and c, and (b) Lipschitz continuity with respect to c over all $c \in \mathbb{R}$, with Lipschitz constant that is independent of s. Hence, by classical differential equations theory (Henrici 1962, Theorem 1.1; Hartman 1973, Theorem 1.1), there exists a unique solution $\Lambda_0(t, \boldsymbol{\beta})$ to the functional equation $\Lambda = \Upsilon(t, \boldsymbol{\beta}, \Lambda)$ subject to $\Lambda(0) = 0$. (The theorems as stated by Henrici and Hartman require $q_{\boldsymbol{\beta}}(s, c)$ to be continuous with respect to s, but a close look at Hartman's proof reveals that this condition can be dispensed with. The conditions (a) and (b) above are sufficient.) We claim that $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ converges uniformly a.s. to $\Lambda_0(t, \boldsymbol{\beta})$. While this claim possibly can be proved directly by mimicing Henrici's proof of his Theorem 1.1, we shall use a more indirect, but shorter, argument.

We define $\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})$ as a modified version of $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ defined by linear interpolation between the jumps, where we have added the superscript n for emphasis. We have, with probability one, $\sup_{t,\boldsymbol{\beta}} |\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta}) - \hat{\Lambda}_0(t, \boldsymbol{\beta})| \to 0$ and $\sup_{t,\boldsymbol{\beta}} |\Upsilon_n(t, \boldsymbol{\beta}, \tilde{\Lambda}_0(t, \boldsymbol{\beta})) - \Upsilon_n(t, \boldsymbol{\beta}, \hat{\Lambda}_0(t, \boldsymbol{\beta}))| \to 0$. We will show that, for suitable n', the family $\boldsymbol{\mathcal{L}} = \{\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta}),\ n \geq n'\}$ is uniformly bounded and equicontinuous. It will then follow, from the Arzela-Ascoli theorem, that the closure of $\boldsymbol{\mathcal{L}}$ in $C([0, \tau] \times \boldsymbol{\mathcal{B}})$ is compact.

We reason as follows. We have $\hat{\Lambda}_0(t, \boldsymbol{\beta}) \leq \psi_{min}^{-1}[n^{-1} \sum_{i=1}^{n} Y_i(\tau)]^{-1}$. Because $n^{-1} \sum_{i=1}^{n} Y_i(\tau) \to y^*$ a.s. as $n \to \infty$, with probability one there exists some n' such that for all $n \geq n'$, we have $n^{-1} \sum_{i=1}^{n} Y_i(\tau) \geq 0.999 y^*$. We then have

$$\hat{\Lambda}_0(t, \boldsymbol{\beta}) \leq \Lambda_{max} \quad \text{for all } n \geq n'. \tag{19}$$

The same holds for the interpolated functions $\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})$, and thus $\boldsymbol{\mathcal{L}}$ is uniformly bounded.

Similarly, $\Lambda_0(t, \boldsymbol{\beta}) \leq \Lambda_{max}$. Thus, in considering empirical processes such as $A_0(\boldsymbol{\beta}, t, c)$, we can restrict attention to $c \in [0, \Lambda_{max}]$. As discussed above, $A_0(\boldsymbol{\beta}, t, c)$ converges a.s. uniformly to $a_0(\boldsymbol{\beta}, t, c)$ over the relevant range of $\boldsymbol{\beta}, t$, and c.

We now turn to the equicontinuity. Write $\bar{N}(t) = n^{-1} \sum_{i=1} N_i(t)$. We have that

$\bar{N}(t) \to E[N_i(t)]$ a.s. as $n \to \infty$, uniformly in t, with

$$E[N_i(t)] = \int_0^t E[Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ, Z_i, \Lambda_0^\circ(s))}]\lambda_0(s)ds.$$

Hence, for any given $\epsilon > 0$, with probability one we can find $n''(\epsilon) \geq n'$ such that $|\bar{N}(t) - E[N_i(t)]| \leq \epsilon/(4\Lambda_{max})$ for all t whenever $n \geq n''(\epsilon)$. In consequence, for all t and u with $u < t$, we have

$$\hat{\Lambda}_0(t, \boldsymbol{\beta}) - \hat{\Lambda}_0(u, \boldsymbol{\beta}) \leq \Lambda_{max}\psi_{max}\lambda_{max}(t-u) + \frac{1}{2}\epsilon \quad \text{for all } n \geq n''(\epsilon). \tag{20}$$

Moreover, from the boundedness of $\alpha_r(\boldsymbol{\beta}, \mathbf{z}, c)$, it follows that $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ satisfies a Lipschitz condition with respect to $\boldsymbol{\beta}$.

These results imply that $\boldsymbol{\mathcal{L}}$ is equicontinuous. (For given ϵ, we need to find δ_1^* and δ_2^* such that $|\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta}) - \tilde{\Lambda}_0^{(n)}(u, \boldsymbol{\beta})| \leq \epsilon$ whenever $|t - u| \leq \delta_1^*$ and $|\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta}) - \tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta}')| \leq \epsilon$ whenever $\|\boldsymbol{\beta} - \boldsymbol{\beta}'\| \leq \delta_2^*$. The latter is easily obtained using the Lipschitz continuity of $\hat{\Lambda}_0(t, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. As for the former, for $n \geq n''(\epsilon)$ this can be accomplished using (20), while for n in the finite set $n' \leq n < n''(\epsilon)$ this can be accomplished using the fact that the function $\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})$ is uniformly continuous on $[0, \tau]$ for every given n.) Hence, by the Arezla-Ascoli theorem, the closure of $\boldsymbol{\mathcal{L}}$ in $C([0, \tau] \times \boldsymbol{\mathcal{B}})$ is compact.

Now we have noted that $A_0(\boldsymbol{\beta}, t, c)$ converges a.s. to $a_0(\boldsymbol{\beta}, t, c)$ uniformly in $\boldsymbol{\beta}, t, c$, and that $\bar{N}(t)$ converges a.s. to $E[N_i(t)]$ uniformly in t. We want to infer from this that $\Upsilon_n(t, \boldsymbol{\beta}, \Lambda)$ converges to $\Upsilon(t, \boldsymbol{\beta}, \Lambda)$ uniformly over $t \in [0, \tau], \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}$, and $\Lambda \in \boldsymbol{\mathcal{L}}$. Given the equicontinuity of $\boldsymbol{\mathcal{L}}$, the argument of Aalen (1976, Lemma 6.1) can be easily adapted to obtain this result. It is here that we use Assumption IX; the adaptation of Aalen's argument requires $a_0(\boldsymbol{\beta}, t, c)$ to be piecewise continuous with finite left and right limits at each point of discontinuity.

Finally, since $\sup_{t,\boldsymbol{\beta}} |\tilde{\Lambda}_0(t, \boldsymbol{\beta}) - \Upsilon_n(t, \boldsymbol{\beta}, \tilde{\Lambda}_0(t, \boldsymbol{\beta}))|$ converges to zero, any limit point of $\{\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})\}$ must satisfy the equation $\Lambda = \Upsilon(t, \boldsymbol{\beta}, \Lambda)$. Since $\Lambda_0(t, \boldsymbol{\beta})$ is the unique solution of this equation, it is the unique limit point of $\{\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})\}$. Thus $\{\tilde{\Lambda}_0^{(n)}(t, \boldsymbol{\beta})\}$

is a sequence in a compact set with unique limit point $\Lambda_0(t,\boldsymbol{\beta})$. Accordingly $\tilde{\Lambda}_0^{(n)}(t,\boldsymbol{\beta})$, and thus also $\hat{\Lambda}_0(t,\boldsymbol{\beta})$ converges a.s. uniformly in t and $\boldsymbol{\beta}$ to $\Lambda_0(t,\boldsymbol{\beta})$.

We now can establish the consistency result. First, from the uniform convergence of $\hat{\Lambda}_0(t,\boldsymbol{\beta})$ to $\Lambda_0(t,\boldsymbol{\beta})$ and the structure of $U(\boldsymbol{\beta},\Lambda)$ (relying principally the previously-noted uniform Lipschitz continuity of $\phi(\boldsymbol{\beta},\mathbf{z},c)$, $\alpha_r(\boldsymbol{\beta},\mathbf{z},c)$, and $\nu(\boldsymbol{\beta},\mathbf{z},c)$ with respect to c), we find that $\sup_{\boldsymbol{\beta}}|\mathbf{U}(\boldsymbol{\beta},\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}))-\mathbf{U}(\boldsymbol{\beta},\Lambda_0(\cdot,\boldsymbol{\beta}))|\to 0$ as $n\to\infty$. Next, by the functional SLLN cited earlier, $\mathbf{U}(\boldsymbol{\beta},\Lambda_0(\cdot,\boldsymbol{\beta}))$ converges a.s. uniformly in $\boldsymbol{\beta}$ to a limit $\mathbf{u}(\boldsymbol{\beta},\Lambda_0(\cdot,\boldsymbol{\beta}))$. Thus, $\mathbf{U}(\boldsymbol{\beta},\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}))$ converges uniformly to $\mathbf{u}(\boldsymbol{\beta},\Lambda_0(\cdot,\boldsymbol{\beta}))$. Now the estimator $\hat{\boldsymbol{\beta}}$ is obtained by solving $\mathbf{U}(\hat{\boldsymbol{\beta}},\hat{\Lambda}_0(\cdot,\hat{\boldsymbol{\beta}}))=\mathbf{0}$. It is easily seen that $\Lambda_0(t,\boldsymbol{\beta}^\circ)=\Lambda_0^\circ(t)$ and that $\mathbf{u}(\boldsymbol{\beta}^\circ,\Lambda_0^\circ)=0$. Further, by arguments in Section A.5 below, the derivative matrix of $\mathbf{u}(\boldsymbol{\beta},\Lambda_0(\cdot,\boldsymbol{\beta}))$ with respect to $\boldsymbol{\beta}$ evaluated at $\boldsymbol{\beta}^\circ$ is equal to $\mathbf{v}(\boldsymbol{\beta}^\circ,\Lambda_0^\circ)$, which by assumption is positive definite. Consequently, by arguments along the lines of Foutz (1977), with probability one there exists a unique consistent root to the pseudo partial likelihood equation $\mathbf{U}(\hat{\boldsymbol{\beta}},\hat{\Lambda}_0(\cdot,\hat{\boldsymbol{\beta}}))=\mathbf{0}$.

A.4. Martingale Representation of $\hat{\Lambda}_0(t,\boldsymbol{\beta}^\circ)-\Lambda_0^\circ(t)$

By Taylor expansion we have

$$\begin{aligned}\hat{\Lambda}_0(t,\boldsymbol{\beta}^\circ) \;&\doteq\; \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}\\ &\quad-\int_0^t \frac{\sum_{i=1}^n Y_i(s)\nu(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}{(\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))})^2}(\hat{\Lambda}_0(s,\boldsymbol{\beta}^\circ)-\Lambda_0^\circ(s))\sum_{i=1}^n dN_i(s),\end{aligned}$$

using the fact that $\hat{\Lambda}_0(s,\boldsymbol{\beta}^\circ)-\hat{\Lambda}_0(s-,\boldsymbol{\beta}^\circ)$ is asymptotically negligible. Recalling $dN_i(s)=Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}\lambda_0^\circ(s)ds+dM_i(s)$, this yields

$$\begin{aligned}\hat{\Lambda}_0(t,\boldsymbol{\beta}^\circ)-\Lambda_0(t) \;&\doteq\; \int_0^t \frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}\\ &\quad-\int_0^t \frac{\sum_{i=1}^n Y_i(s)\nu(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}{(\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))})^2}(\hat{\Lambda}_0(s,\boldsymbol{\beta}^\circ)-\Lambda_0^\circ(s))\sum_{i=1}^n dN_i(s).\end{aligned}$$

As in Yang and Prentice (1999, Appendix C), the foregoing equation has solution

$$\hat{\Lambda}_0(t,\boldsymbol{\beta}^\circ)-\Lambda_0^\circ(t)\doteq\frac{1}{\hat{P}(t)}\int_0^t \hat{P}(s-)\frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}, \tag{21}$$

where

$$\hat{P}(t) = \prod_{s \le t} \left(1 + \frac{\sum_{i=1}^n Y_i(s) \nu(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, \Lambda_0^\circ(s)) e^{\phi(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, \Lambda_0^\circ(s))}}{(\sum_{i=1}^n Y_i(s) e^{\phi(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, \Lambda_0^\circ(s))})^2} \sum_{i=1}^n dN_i(s) \right).$$

A.5. Asymptotic Distribution of $\hat{\boldsymbol{\beta}}$

The starting point for developing the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is the relation (12), which we repeat here for convenience:

$$\begin{aligned} \mathbf{0} &= \mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}})) \\ &= \mathbf{U}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot)) + [\mathbf{U}(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}^\circ)) - \mathbf{U}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot))] \\ &\quad + [\mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}})) - \mathbf{U}(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}^\circ))]. \end{aligned} \tag{22}$$

We analyze the three terms on the right hand side one by one.

First, from the representation (16), it follows by standard martingale arguments as in Andersen and Gill (1982) that $n^{\frac{1}{2}}\mathbf{U}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot))$ is asymptotically mean-zero multivariate normal. Using further arguments of Andersen and Gill, along with consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$, it may be shown that the asymptotic covariance matrix may be consistently estimated by the matrix $\mathbf{V}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0)$ with $\mathbf{V}(\boldsymbol{\beta}, \Lambda)$ defined as in (13).

Next, $\mathbf{U}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}_0(\cdot, \hat{\boldsymbol{\beta}})) - \mathbf{U}(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}^\circ)) = \mathbf{D}(\tilde{\boldsymbol{\beta}})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\circ)$, where $\tilde{\boldsymbol{\beta}}$ lies between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^\circ$, and $D_{rs}(\boldsymbol{\beta})$ is the partial derivative of $U_r(\boldsymbol{\beta}, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}))$ with respect to β_s. A straightforward calculation shows that

$$D_{rs}(\boldsymbol{\beta}) = -V_{rs}(\boldsymbol{\beta}, \hat{\Lambda}_0) + \Gamma_{rs}(\boldsymbol{\beta}), \tag{23}$$

where

$$\Gamma_{rs}(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\gamma_{rs}(\boldsymbol{\beta}, \mathbf{Z}_i, u) - \frac{\sum_{j=1}^n Y_j(u) \gamma_{rs}(\boldsymbol{\beta}, \mathbf{Z}_j, u) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \hat{\Lambda}_0(u, \boldsymbol{\beta}))}}{\sum_{j=1}^n Y_j(s) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \hat{\Lambda}_0(u, \boldsymbol{\beta}))}} \right) dN_i(u), \tag{24}$$

with $\gamma_{rs}(\boldsymbol{\beta}, \mathbf{z}, t)$ denoting the partial derivative of $\xi_r(\boldsymbol{\beta}, \mathbf{z}, t)$ with respect to β_s. By consistency of $\hat{\boldsymbol{\beta}}$ and $\hat{\Lambda}_0$, the quantity $D_{rs}(\boldsymbol{\beta})$ is asymptotically equivalent to the expression

(23) with $\boldsymbol{\beta}^\circ$ in place of $\boldsymbol{\beta}$ and Λ_0° in place of $\hat{\Lambda}_0$. After making these substitutions in (24), we find, as with $\mathbf{U}$, that the dN_i in (24) may be replaced by dM_i and consequently $\Gamma_{rs}(\boldsymbol{\beta}^\circ)$ tends in the limit to zero. It follows that $\mathbf{D}(\tilde{\boldsymbol{\beta}})$ is asymptotically equivalent to $-\mathbf{V}(\boldsymbol{\beta}^\circ, \Lambda_0^\circ)$ and thus also to $-\mathbf{V}(\hat{\boldsymbol{\beta}}, \hat{\Lambda}(\cdot, \hat{\boldsymbol{\beta}}))$.

Finally, we have to deal with $U_r(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot))$. Define, for $\sigma \in \mathbb{R}$, $U_r^*(\boldsymbol{\beta}, \sigma) = U_r(\boldsymbol{\beta}, \Lambda_0^*(t, \boldsymbol{\beta}, \sigma))$, where

$$\Lambda_0^*(t, \boldsymbol{\beta}, \sigma) = \Lambda_0(t, \boldsymbol{\beta}) + \sigma[\hat{\Lambda}_0(t, \boldsymbol{\beta}) - \Lambda_0(t, \boldsymbol{\beta})].$$

By the mean value theorem

$$U_r(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot, \boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot)) = \frac{\partial}{\partial \sigma} U_r^*(\boldsymbol{\beta}^\circ, \sigma)\Big|_{\sigma=\tilde{\sigma}} \tag{25}$$

for some $\tilde{\sigma} \in [0, 1]$. Now define

$$\begin{aligned}
\zeta_r(\boldsymbol{\beta}, \mathbf{z}, u, \sigma) &= \frac{\partial}{\partial \sigma} \xi_r(\boldsymbol{\beta}, \mathbf{z}, u, \Lambda_0^*(\cdot, \boldsymbol{\beta}, \sigma)) \\
&= \dot{\alpha}_r(\boldsymbol{\beta}, \mathbf{z}, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma))(\hat{\Lambda}_0(u, \boldsymbol{\beta}) - \Lambda_0(u, \boldsymbol{\beta})) \\
&+ \dot{\nu}(\boldsymbol{\beta}, \mathbf{z}, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma)) Q_r(u, \boldsymbol{\beta}, \Lambda_0^*(\cdot, \boldsymbol{\beta}, \sigma))(\hat{\Lambda}_0(u, \boldsymbol{\beta})) - \Lambda_0(u, \boldsymbol{\beta})) \\
&+ \nu(\boldsymbol{\beta}, \mathbf{z}, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma)) \int_0^u \Xi(s, \Lambda_0^*(s-, \boldsymbol{\beta}, \sigma))(\hat{\Lambda}_0(s-, \boldsymbol{\beta}) - \Lambda_0(s, \boldsymbol{\beta})) d\bar{N}(s),
\end{aligned} \tag{26}$$

where

$$\begin{aligned}
\dot{\alpha}_r(\boldsymbol{\beta}, \mathbf{z}, c) &= \frac{\partial}{\partial c} \alpha_r(\boldsymbol{\beta}, \mathbf{z}, c), \\
\dot{\nu}(\boldsymbol{\beta}, \mathbf{z}, c) &= \frac{\partial}{\partial c} \nu(\boldsymbol{\beta}, \mathbf{z}, c),
\end{aligned}$$

and $\Xi(s, c)$ is a lengthy expression of magnitude $O(1)$. (The last term in (26) arises from differentiating $Q_r(u, \boldsymbol{\beta}, \Lambda_0^*(\cdot, \boldsymbol{\beta}, \sigma))$ with respect to σ. This involves differentiating the analogue of (11) with respect to σ and solving the forward recursion.) In addition, define

$$\begin{aligned}
C_r(u, \boldsymbol{\beta}, \sigma) = & \frac{\sum_{j=1}^n Y_j(u) \xi_r(\boldsymbol{\beta}, \mathbf{Z}_j, u, \Lambda_0^*(\cdot, \boldsymbol{\beta}, \sigma)) \nu(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma)) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma))}}{\sum_{j=1}^n Y_j(u) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma))}} \\
& - \Bigg[\left(\frac{\sum_{j=1}^n Y_j(u) \xi_r(\boldsymbol{\beta}, \mathbf{Z}_j, u, \Lambda_0^*(\cdot, \boldsymbol{\beta}, \sigma)) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma))}}{\sum_{j=1}^n Y_j(u) e^{\phi(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u, \boldsymbol{\beta}, \sigma))}} \right)
\end{aligned}$$

$$\times \left(\frac{\sum_{j=1}^n Y_j(u)\nu(\boldsymbol{\beta}, \mathbf{Z}_j, \Lambda_0^*(u,\boldsymbol{\beta},\sigma))e^{\phi(\boldsymbol{\beta},\mathbf{Z}_j,\Lambda_0^*(u,\boldsymbol{\beta},\sigma))}}{\sum_{j=1}^n Y_j(s)e^{\phi(\boldsymbol{\beta},\mathbf{Z}_j,\Lambda_0^*(u,\boldsymbol{\beta},\sigma))}} \right) \Bigg]. \qquad (27)$$

Then straightforward calculation from (25) yields

$$\begin{aligned} U_r(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot)) &= -\frac{1}{n}\sum_{i=1}^n \int_0^\tau C_r(u,\boldsymbol{\beta}^\circ,\tilde{\sigma})(\hat{\Lambda}_0(u,\boldsymbol{\beta}^\circ) - \Lambda_0^\circ(u))dN_i(u) \\ +\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left(\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, u, \tilde{\sigma}) - \frac{\sum_{j=1}^n Y_j(u)\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{Z}_j, u,\tilde{\sigma})e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_j,\Lambda_0^*(u,\boldsymbol{\beta}^\circ,\tilde{\sigma}))}}{\sum_{j=1}^n Y_j(u)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_j,\Lambda_0^*(u,\boldsymbol{\beta}^\circ,\tilde{\sigma}))}} \right) dN_i(u). \quad (28) \end{aligned}$$

Next, define $\zeta_r(\boldsymbol{\beta}, \mathbf{z}, u) = \zeta_r(\boldsymbol{\beta}, \mathbf{z}, u, 0)$ and $C_r(u,\boldsymbol{\beta}) = C_r(u,\boldsymbol{\beta},0)$. We then have

$$\begin{aligned} \zeta_r(\boldsymbol{\beta}, \mathbf{z}, u, \tilde{\sigma}) &= \zeta_r(\boldsymbol{\beta}, \mathbf{z}, u) + O(\|\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}) - \Lambda_0(\cdot,\boldsymbol{\beta})\|^2), \\ C_r(u,\boldsymbol{\beta},\tilde{\sigma}) &= C_r(u,\boldsymbol{\beta}) + O(\|\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}) - \Lambda_0(\cdot,\boldsymbol{\beta})\|). \end{aligned}$$

Hence

$$\begin{aligned} U_r(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot)) &\doteq -\frac{1}{n}\sum_{i=1}^n \int_0^\tau C_r(u,\boldsymbol{\beta}^\circ)(\hat{\Lambda}_0(u,\boldsymbol{\beta}^\circ) - \Lambda_0^\circ(u))dN_i(u) \\ +\frac{1}{n}\sum_{i=1}^n \int_0^\tau \left(\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{Z}_i, u) - \frac{\sum_{j=1}^n Y_j(u)\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{Z}_j, u)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_j,\Lambda_0^\circ(u))}}{\sum_{j=1}^n Y_j(u)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_j,\Lambda_0^\circ(u))}} \right) dN_i(u). \quad (29) \end{aligned}$$

The error in the above approximation is of order $O(\|\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ) - \Lambda_0^\circ\|^2)$ and hence negligible for present purposes.

We can now deal with the second term in (29) in the same way that we dealt with $\Gamma_{rs}(\boldsymbol{\beta})$ in the third term of (22). Using $dN_i(u) = \lambda_0^\circ(u)Y_i(u)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(u))} + dM_i(t)$ and re-arranging terms, we find that the $dN_i(u)$ in the second term of (29) can be replaced by $dM_i(u)$. Now, since $\Delta\hat{\Lambda}_0(u,\boldsymbol{\beta}^\circ)$, $\Delta Q_r(u,\boldsymbol{\beta},\Lambda_0^\circ)$, and $d\bar{N}(u)$ are of order $O(n^{-1})$, we can replace $\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{z}, u)$ by $\zeta_r(\boldsymbol{\beta}^\circ, \mathbf{z}, u-)$ to obtain a predictable integrand in the second term of (29). The resulting martingale term is of order $O_p(n^{-\frac{1}{2}}\|\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ) - \Lambda_0^\circ(\cdot)\|)$. In sum, the second term in (29) is negligible in comparison with the first term.

Discarding this second term and substituting the representation given in Section A.4 for $\hat{\Lambda}_0(u,\boldsymbol{\beta}^\circ) - \Lambda_0^\circ(u)$ into the first term, we obtain

$$U_r(\boldsymbol{\beta}^\circ, \hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ, \Lambda_0^\circ(\cdot))$$

$$\doteq -\frac{1}{n}\int_0^\tau C_r(t,\boldsymbol{\beta}^\circ)\left[\frac{1}{\hat{P}(t)}\int_0^t \hat{P}(s-)\frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}}\right]\sum_{j=1}^n dN_j(t).$$

Interchanging the order of integration, we get

$$\begin{aligned}
&U_r(\boldsymbol{\beta}^\circ,\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ)) - U_r(\boldsymbol{\beta}^\circ,\Lambda_0^\circ(\cdot)) \\
&\quad\doteq\ -\int_0^\tau\left[\frac{1}{n}\int_s^\tau \frac{C_r(t,\boldsymbol{\beta}^\circ)}{\hat{P}(t)}\sum_{j=1}^n dN_j(t)\right]\hat{P}(s-)\frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}} \\
&\quad=\ -\int_0^\tau G_r(s,\boldsymbol{\beta}^\circ)\hat{P}(s-)\frac{\sum_{i=1}^n dM_i(s)}{\sum_{i=1}^n Y_i(s)e^{\phi(\boldsymbol{\beta}^\circ,\mathbf{Z}_i,\Lambda_0^\circ(s))}},
\end{aligned}$$

where

$$G_r(s,\boldsymbol{\beta}) = \frac{1}{n}\int_s^\tau \frac{C_r(t,\boldsymbol{\beta})}{\hat{P}(t)}\sum_{j=1}^n dN_j(t).$$

By the asymptotic stability of the empirical process terms and the martingale central limit theorem, we now find that $n^{\frac{1}{2}}[\mathbf{U}(\boldsymbol{\beta}^\circ,\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}^\circ)) - \mathbf{U}(\boldsymbol{\beta}^\circ,\Lambda_0^\circ(\cdot))]$ is asymptotically mean-zero multivariate normal with covariance matrix that may be consistently estimated by

$$\hat{H}_{rs} = \int_0^\tau G_r(u,\hat{\boldsymbol{\beta}})G_s(u,\hat{\boldsymbol{\beta}})\hat{P}(u-)^2\frac{n\sum_{i=1}^n dN_i(u)}{(\sum_{i=1}^n Y_i(u)e^{\phi(\hat{\boldsymbol{\beta}},\mathbf{Z}_i,\hat{\Lambda}_0(u))})^2}. \tag{30}$$

Moreover, by an argument as in Andersen and Gill (1982, p. 1104), we find that $n^{\frac{1}{2}}[\mathbf{U}(\boldsymbol{\beta}^\circ,\hat{\Lambda}_0(\cdot,\boldsymbol{\beta}) - \mathbf{U}(\boldsymbol{\beta}^\circ,\Lambda_0^\circ(\cdot))]$ and $n^{\frac{1}{2}}\mathbf{U}(\boldsymbol{\beta}^\circ,\Lambda_0^\circ(\cdot))$ are asymptotically independent.

Putting the foregoing results together, we obtain the asymptotic distribution result stated in Section 2.4.

A.6. Modifications for Transformation Models

Below we indicate the modifications to ϕ, α_r, and ν for transformation models of the form (14) with $g'(0) > 0$. We put $\rho(c_1,c_2) = g'(h(c_1)c_2)/g'(h(c_1))$ and we let $\dot{\rho}_l(c_1,c_2)$ denote the partial derivative of $\rho(c_1,c_2)$ with respect to l for $l = 1,2$. Note that $(\partial/\partial c_1)g(h(c_1)c_2) = \rho(c_1,c_2)$. We further let $\dot{\psi}_r(\mathbf{x};\boldsymbol{\beta})$ denote the partial derivative of $\psi(\mathbf{x};\boldsymbol{\beta})$ with respect to β_r. With this notation the appropriate modified definitions are as follows:

$$
\begin{aligned}
\phi(\boldsymbol{\beta},\mathbf{z},c) &= \log\int \exp(-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta})))\rho(c,\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta})\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}) \\
&\quad -\log\int \exp(-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta})))\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
\alpha_r(\boldsymbol{\beta},\mathbf{z},c) &= \frac{\partial}{\partial\beta_r}\phi(\boldsymbol{\beta},\mathbf{z},c) = B_2^{-1}B_1 + B_4^{-1}B_3, \\
\nu(\boldsymbol{\beta},\mathbf{z},c) &= \frac{\partial}{\partial c}\phi(\boldsymbol{\beta},\mathbf{z},c) = B_2^{-1}\tilde{B}_1 + B_4^{-1}\tilde{B}_3, \\
B_1 &= \int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}\dot{\psi}_r(\mathbf{x};\boldsymbol{\beta})[-h(c)g'(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))\rho(c,\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta}) \\
&\quad +\dot{\rho}_2(h(c),\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta}) + \rho(h(c),\psi(\mathbf{x};\boldsymbol{\beta}))]\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
\tilde{B}_1 &= \int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}[-\rho(c,\psi(\mathbf{x};\boldsymbol{\beta}))^2\psi(\mathbf{x};\boldsymbol{\beta})^2 + \dot{\rho}_1(c,\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta})]\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
B_2 &= \int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}\rho(c,\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta})\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
B_3 &= h(c)\int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}g'(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))\dot{\psi}_r(\mathbf{x};\boldsymbol{\beta})\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
\tilde{B}_3 &= \int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}\rho(c,\psi(\mathbf{x};\boldsymbol{\beta}))\psi(\mathbf{x};\boldsymbol{\beta})\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}), \\
B_4 &= \int e^{-g(h(c)\psi(\mathbf{x};\boldsymbol{\beta}))}\omega(\mathbf{x}|\mathbf{z})m(\mathbf{dx}).
\end{aligned}
$$

REFERENCES

Aalen, O. (1976). "Nonparametric Inference in Connection with Multiple Decrement Models," *Scandinavian Journal of Statistics*, 3, 15-27.

Anderson, P. K., and Gill, R. D. (1982). "Cox's Regression Model for Counting Processes: A Large Sample Study," *Annals of Statistics*, 10, 1100-1120.

Armitage, P. and Doll, R. (1961). "Stochastic Models for Carcinogenesis," in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability: Biology and Problems of Health*, pp. 19-38, Berkeley, CA: University of California Press.

Bennett, S. (1983). "Analysis of Survival Data by the Proportional Odds Model," *Statistics in Medicine*, 2, 273-277.

Breslow, N. (1974). "Covariance Analysis of Censored Survival Data," *Biometrics*, 30, 89-99.

Breslow, N. and Day, N. E. (1993). *Statistical Methods in Cancer Research. Vol. II: The Design and Analysis of Cohort Studies*, Oxford: Oxford University Press.

Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, London: Chapman and Hall.

Chen, K., Jin, Z., and Ying, Z. (2002). "Semiparametric Analysis of Transformation Models With Censored Data," *Biometrika*, 89, 659-668.

Chen, Y. H. (2002). "Cox Regression in Cohort Studies with Validation Sampling," *Journal of the Royal Statistical Society, Series B*, 64, 51-62.

Cheng, S. C., Wei, L. J., and Ying Z. (1995). "Analysis of Transformation Models With Censored Data," *Biometrika*, 82, 835-845.

Cheng, S. C. and Wang, N. (2001). "Linear Transformation Models for Failure Time Data With Covariate Measurement Error," *Journal of the American Statistical Association*, 96, 706-716.

Cox, D. R. (1972). "Regression Models and Life-Tables" (with discussion), *Journal of the Royal Statistical Society, Series B*, 34, 187-220.

Cox, D. R. (1975). "Partial Likelihood," *Biometrika*, 62, 269-276.

Dabrowska, D. M. and Doksum, K. A. (1988). "Estimating and Testing in a Two-Sample Generalized Odds-Rate Model," *Journal of the American Statistical Association*, 83, 744-749.

Foutz, R. V. (1977). "On the Unique Consistent Solution to the Likelihood Equations," *Journal of the American Statistical Association*, 72, 147-148.

Gordon, T., and Kannel, W. E. (1968). *The Framingham Study, Introduction and General Background on the Framingham Study*, Bethesda: National Heart, Lung, and Blood Institute.

Gorfine, M., Zucker, D. M., and Hsu, L. (2005). "Prospective Survival Analysis With a General Semiparametric Shared Frailty Model: A Pseudo Full Likelihood Approach," Technical Report, Bar-Ilan University.

Hartman, P. (1973). *Ordinary Differential Equations*, 2nd ed. (reprinted, 1982), Boston: Birkhäuser.

Henrici, P. (1962). *Discrete Variable Methods in Ordinary Differential Equations*, New York: Wiley.

Hu, C., and Lin, D. Y. (2002). "Cox Regression with Covariate Measurement Error," *Scandinavian Journal of Statistics*, 29, 637-655.

Hu, F. B., Stampfer, M. J., Manson, J. E., Rimm, E., Colditz, G. A., Rosner, B. A., Hennekens, C. H., and Willett, W. C. (1997). "Dietary Fat Intake and the Risk of Coronary Heart Disease in Women," *New England Journal of Medicine*, 337, 1491-1499.

Hu, P., Tsiatis, A. A., and Davidian, M. (1998)." Estimating the Parameters in the Cox Model when Covariate Variables are Measured With Error," *Biometrics*, 54, 1407-1419.

Huang, Y. and Wang, C. Y. (2000). "Cox Regression with Accurate Covariates Unascertainable: a Nonparametric-Correction Approach," *Journal of the American Statistical Association*, 95, 1209-1219.

Hughes, M. D. (1993). "Regression Dilution in the Proportional Hazards Model," *Biometrics*, 49, 1056–1066.

Hunter, D. J., Spiegelman, D., Adami, H. O., Beeson, L., van den Brandt, P. A., Folsom, A. R., Fraser, G. E., Goldbohm, R. A., Graham, S., Howe, G. R., Kushi, L. H., Marshall, J. R., McDermott, A., Miller, A. B., Speizer. F. E., Wolk, A., Yaun, S. S., and Willett, W. (1996). "Cohort Studies of Fat Intake and Risk of Breast Cancer: A Pooled Analysis," *New England Journal of Medicine*, 334, 356-361.

Kong, F. H., and Gu, M. (1999). "Consistent Estimation in Cox's Proportional Hazards Model with Covariate Measurement Errors," *Statistica Sinica*, 9, 953-969.

Nakamura, T. (1990). "Corrected Score Function of Errors-in-variables Models: Methodology and Application to Generalized Linear Models," *Biometrika*, 77, 127-137.

Nakamura, T. (1992). "Proportional Hazards Model with Covariates Subject to Measurement Error," *Biometrics*, 48, 829-838.

Prentice, R. (1982). "Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model," *Biometrika*, 69, 331-342.

Stefanski, L. (1989). "Unbiased Estimation of a Nonlinear Function of a Normal Mean with Application to Measurement-error Models," *Communications in Statistics, Theory and Methods*, 18, 4335-4358.

Thomas, D. C. (1981). "General Relative-Risk Models for Survival Time and Matched Case-Control Analysis," *Biometrics*, 37, 673-686.

Tsiatis, A. A. (1981). "A Large Sample Study of Cox's Regression Model," *Annals of Statistics*, 9, 93-108.

Wang, C. Y., Hsu, L., Feng, Z. D., and Prentice, R. L. (1997). "Regression Calibration in Failure Time Regression," *Biometrics*, 53, 131-145.

Xie, S. X., Wang, C. Y., and Prentice, R. L. (2001). "A Risk Set Calibration Method for Failure Time Regression by Using a Covariate Reliability Sample. *Journal of the Royal Statistical Society, Series B*, 63, 855-870.

Yang, S. and Prentice, R. L. (1999). "Semiparametric Inference in the Proportional Odds Regression Model," *Journal of the American Statistical Association*, 94, 125-136.

Zhou, H., and Pepe, M. (1995). "Auxilliary Covariate Data in Failure Time Regression," *Biometrika*, 82, 139-149.

Zhou, H., and Wang, C. Y. (2001). "Failure Time Regression With Continuous Covariates Measured With Error," *Journal of the Royal Statistical Society, Series B*, 62, 657-665.

Zucker, D. M., and Spiegelman, D. (2004). "Inference for the Proportional Hazards Model with Misclassified Discrete-Valued Covariates," *Biometrics*, 60, 324-334.

Zucker, D. M., and Yang, S. (2005). "Inference for a Family of Survival Models Encompassing the Proportional Hazards and Proportional Odds Model," *Statistics in Medicine* (to appear).

Table 1

Simulation Results for the Case of a Single Binary Covariate
Sample Size = 2,000, Unexposed Cumulative Incidence = 25%

Percent Exposed	Error Rate	True RR	Percent Bias In Estimated Log RR Naive Cox	MPPLE	FWMLE	Empirical Variance	% Bias In Variance	95% CI Coverage	MSE Ratio
5 %	1 %	1.5	-15.89	-2.68	-1.92	3.77	3.63	95.34	1.01
5 %	1 %	2.0	-14.00	-0.74	-0.43	3.18	-1.17	95.64	0.97
5 %	5 %	1.5	-48.45	-5.73	-4.77	6.72	4.45	95.96	1.03
5 %	5 %	2.0	-46.08	-2.91	-2.16	5.37	2.19	96.30	0.99
5 %	10 %	1.5	-66.02	-7.50	-6.99	11.47	12.96	96.82	1.09
5 %	10 %	2.0	-64.16	-4.11	-2.90	9.62	-0.84	96.20	1.01
5 %	20 %	1.5	-82.24	-16.71	-23.81	28.33	153.27	96.40	1.50
5 %	20 %	2.0	-80.92	-10.19	-9.87	23.91	28.87	97.65	1.21
25 %	1 %	1.5	-3.03	-0.33	-0.02	0.94	-1.86	95.00	0.99
25 %	1 %	2.0	-3.01	-0.18	-0.10	0.81	-1.80	94.62	1.00
25 %	5 %	1.5	-14.09	-0.26	0.06	1.14	-0.91	94.80	0.99
25 %	5 %	2.0	-13.89	-0.32	-0.22	0.99	-1.42	94.96	1.00
25 %	10 %	1.5	-26.61	-0.33	0.04	1.52	-1.93	95.22	0.99
25 %	10 %	2.0	-26.14	-0.41	-0.30	1.25	2.43	95.24	1.00
25 %	20 %	1.5	-48.41	-1.09	-0.49	2.77	1.57	95.54	1.00
25 %	20 %	2.0	-47.42	-0.33	-0.15	2.46	-2.00	94.82	0.99
40 %	1 %	1.5	-2.43	-0.41	-0.32	0.77	-1.63	94.64	1.00
40 %	1 %	2.0	-2.11	-0.06	0.01	0.66	1.96	95.20	1.00
40 %	5 %	1.5	-10.03	0.32	0.43	0.91	-1.13	94.64	1.00
40 %	5 %	2.0	-10.35	0.00	0.07	0.83	-2.78	94.92	0.99
40 %	10 %	1.5	-20.63	-0.08	0.05	1.15	-1.28	94.98	1.01
40 %	10 %	2.0	-20.40	0.27	0.30	1.05	-1.81	95.14	0.99
40 %	20 %	1.5	-41.04	-0.38	-0.33	1.94	5.15	95.84	1.01
40 %	20 %	2.0	-40.66	0.33	0.34	1.87	-0.81	94.86	0.98

MPPLE = maximum pseudo partial likelihood estimate
FWMLE = full Weibull maximum likelihood estimate

Table 2

Simulation Results for the Case of a Single Normal Covariate, n=150
Error Variance Assumed Known
Evaluation of the Maximum Pseudo Likelihood Estimate (MPPLE)

Error	True	Bias in Estimated β		Empirical	% Error	95% CI	MSE Ratio	
Variance	e^{β}	Naive Cox	MPPLE	Variance × 100	In Variance	Coverage	Vs Expnl	Vs Weibull
0.25	1	0.00	0.00	1.40	-1.87	95.62	0.99	1.01
0.25	2	-0.15	0.01	2.10	1.62	95.98	0.86	0.99
0.25	4	-0.42	0.03	5.51	-0.21	96.18	0.54	0.99
0.50	1	-0.00	0.00	1.67	-0.87	95.50	1.01	1.02
0.50	2	-0.26	0.02	2.97	-0.54	95.90	0.82	0.98
0.50	4	-0.64	0.05	9.20	1.95	95.93	0.40	0.96
1.00	1	0.00	0.00	2.37	-4.66	95.48	0.98	1.01
1.00	2	-0.37	0.02	4.56	3.00	96.28	0.71	0.99
1.00	4	-0.87	0.07	13.97	24.04	95.49	0.27	0.96

Table 3

Simulation Results for the Case of a Single Normal Covariate, n=300
Error Variance Assumed Known
Evaluation of the Maximum Pseudo Likelihood Estimate (MPPLE)

Error	True	Bias in Estimated β		Empirical	% Error	95% CI	MSE Ratio	
Variance	e^{β}	Naive Cox	MPPLE	Variance × 100	In Variance	Coverage	Vs Expnl	Vs Weibull
0.25	1	0.00	0.00	0.67	0.98	95.62	1.00	1.01
0.25	2	-0.16	0.01	1.02	1.12	95.42	0.90	0.98
0.25	4	-0.42	0.01	2.53	1.00	95.96	0.58	0.97
0.50	1	0.00	0.00	0.84	-3.34	94.46	1.00	1.01
0.50	2	-0.26	0.01	1.38	1.94	95.48	0.83	0.98
0.50	4	-0.64	0.02	4.04	2.88	95.98	0.44	0.95
1.00	1	0.00	0.00	1.10	-1.76	95.34	1.00	1.01
1.00	2	-0.38	0.01	2.14	0.98	95.20	0.73	0.97
1.00	4	-0.87	0.03	7.86	1.34	95.50	0.28	0.92

Table 4
Simulation Results for the Case of a Single Normal Covariate
Replicate Measurements Setup - Two Replicates Per Individual
Evaluation of the Maximum Pseudo Likelihood Estimate (MPPLE)

Error Variance	True e^β	Bias in Estimated β n=150	n=300	Empirical Variance* n=150	n=300	% Error in Variance n=150	n=300	95% CI Coverage n=150	n=300
0.50	1	0.00	0.00	1.35	0.65	3.40	4.10	95.72	95.78
0.50	2	0.02	0.01	2.41	1.09	-2.72	0.89	95.80	95.56
0.50	4	0.04	0.02	6.36	3.01	1.48	-1.90	95.74	95.96
1.00	1	0.00	0.00	1.70	0.88	3.17	-5.53	96.28	94.92
1.00	2	0.03	0.01	3.76	1.69	3.14	0.77	96.04	95.90
1.00	4	0.07	0.04	13.15	5.97	13.72	0.50	94.80	95.76
2.00	1	0.00	0.00	2.95	1.14	11.06	4.37	97.34	96.34
2.00	2	0.07	0.03	10.39	3.55	32.27	8.09	94.31	95.16
2.00	4	0.08	0.08	23.33	14.69	91.07	25.08	92.34	93.43

* Table shows empirical variance times 100.