mathematics

MDPI

Article

Estimation and Model Misspecification for Recurrent Event Data with Covariates Under Measurement Errors

Ravinath Alahakoon [1], Gideon K. D. Zamba [2], Xuerong Meggie Wen [3] and Akim Adekpedjou [3,*]

[1] Gary W. Rollins College of Business, The University of Tennessee-Chattanooga, Chattanooga, TN 37403, USA; ravinath-alahakoon@utc.edu
[2] Department of Biostatistics, The University of Iowa, Iowa City, IA 52242, USA; gideon-zamba@uiowa.edu
[3] Department of Mathematics and Statistics, Missouri University of Science and Technology, Rolla, MO 65409, USA; wenx@mst.edu
* Correspondence: akima@mst.edu

Abstract: For subject i, we monitor an event that can occur multiple times over a random observation window $[0, \tau_i)$. At each recurrence, p concomitant variables, $\mathbf{x}_i$, associated to the event recurrence are recorded—a subset ($q \leq p$) of which is measured with errors. To circumvent the problem of bias and consistency associated with parameter estimation in the presence of measurement errors, we propose inference for corrected estimating equations with well-behaved roots under an additive measurement errors model. We show that estimation is essentially unbiased under the corrected profile likelihood for recurrent events, in comparison to biased estimations under a likelihood function that ignores correction. We propose methods for obtaining estimators of error variance and discuss the properties of the estimators. We further investigate the case of misspecified error models and show that the resulting estimators under misspecification converge to a value different from that of the true parameter—thereby providing a basis for bias assessment. We demonstrate the foregoing correction methods on an open-source rhDNase dataset gathered in a clinical setting.

Keywords: recurrent events; covariate measurement errors; model misspecification; Kullback–Leibler divergence; corrected score

MSC: 62N02



Received: 11 November 2024
Revised: 17 December 2024
Accepted: 24 December 2024
Published: 30 December 2024

Citation: Alahakoon, R.; Zamba, G.K.D.; Wen, X.M.; Adekpedjou, A. Estimation and Model Misspecification for Recurrent Event Data with Covariates Under Measurement Errors. *Mathematics* **2025**, *13*, 113. https://doi.org/10.3390/math13010113

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A recurrent event process is a process that repeatedly generates events serially. It is encountered in many fields such as biomedical science, epidemiology, social science, reliability, and actuarial science, just to name a few. The literature on methods and models that address various scientific questions pertaining to recurrent event processes is well established. Regardless of the problem of interest, analyses of recurrent events can be broadly classified as either gap-time analysis, or time-to-event analysis. Under these two paradigms, the focus can be put on estimating the intensity function $\lambda(\cdot)$, the survivor function $\bar{F}(\cdot)$, the mean rate function $\mu(\cdot)$, or other functionals of these unknowns in the presence of possible time-dependent covariates. Because covariates play an important role in better understanding time to failure, incorporating them in modeling has always been beneficial.

Denote $\mathbf{x}$ a p-dimensional vector of possibly time varying covariates, $\boldsymbol{\beta}$ a p-dimensional regressor, and $\lambda_0(\cdot)$ a baseline hazard function. For a vector $\mathbf{a}$, we call $\mathbf{a}'$ its transpose. Models that account for covariates include the multiplicative intensity model

$\lambda(s) = \lambda_0(s)\exp(\boldsymbol{\beta}'\mathbf{x}(s))$, the additive model $\lambda(s) = \lambda_0(s) + \boldsymbol{\beta}'\mathbf{x}(s)$, the mean-rate model $\mu(s) = \mu_0(s)\exp(\boldsymbol{\beta}'\mathbf{x}(s))$, the additive–multiplicative model $\lambda(s) = g(\boldsymbol{\beta}'\mathbf{x}(s)) + \lambda_0(s)h(\boldsymbol{\beta}'\mathbf{x}(s))$, and the accelerated failure-time model $\lambda(s) = \mathbf{x}(s)'\boldsymbol{\beta} + w$, as a few examples. With specific choices of $g(\cdot)$ and $h(\cdot)$ in the additive–multiplicative model, one can easily retrieve the Cox and the additive models. The choice of a model depends on the area of applications, the area of research interest, and the goodness-of-fit. What has been abundant in the literature are model specifications assuming all covariates are perfectly measured data during collection. However, the notion that all covariates are perfectly measured is farfetched in the majority of research fields. A simple fact, well-known in medicine, is that blood pressure measured in healthcare facilities, for example, suffers from the psychological effect of subjects being in a doctor's office even if the instrument used to measure this vital sign appears to be well-calibrated. The same is true for measuring household lead level, an error-prone process usually influenced by environmental factors including air quality, dust movement, and soil quality. Similarly, the process of measuring nutrient intake has been a long-documented error-prone process with measurement errors that can negatively impact health. Covariates measurement errors can be associated with the mechanism by which the measurements are taken, the situations under which they are measured, environmental factors, human errors, missingness, or many other lurking factors. Ignoring errors in modeling could lead to biased parameter estimates, distorted inferences, and inaccurate conclusions. These are the technical consequences. There are practical consequences as well, specifically in biomedical studies where doctors' directives are given based on the accurate reading of patient diagnostics. If these readings are erroneous, inaccurate prescriptions and directives will be given, leading to serious health consequences. There has been an extensive body of work on the topic including classical, censored, truncated, and uncensored data, resulting in numerous correction methods in models that include measurement errors. The correction methods are either parametric or nonparametric, wherein the error term acts on the true value of the covariates in an additive, multiplicative, or other manner. The parametric approach to the problem assumes a distribution on the error terms, while the nonparametric approach relaxes that assumption and uses replicate surrogates and instrumental variables for error correction. Regardless of the approach and data type, the following references are noteworthy: [1–11]. A comprehensive review of methods for measurement errors can be found in [12]. Textbooks dealing with measurement errors include [12,13], and, most recently, [14].

Although there has been an abundance of literature with measurement error models and correction methods for single events, there is a paucity of literature on recurrent events. Turnbull [15], under a normal assumption for the errors, proposed a moment-based method for correcting a naive estimator, while Hu and Lin [16] corrected a partial score function under a symmetric distribution for the errors. A few manuscripts on the topic with recurrent events have appeared in the last decade. Yi and Lawless [17], using the correction from [4] and the simulation extrapolation (Simex) of [18], present methods for modeling time-to-events that account for measurement errors under a wide class of models for the hazards. In their approach, parameters were estimated using likelihood-based tools and estimating equations. More recently, some authors have focused on the measurement errors problem in recurrent events while simultaneously dealing with informative censoring. For instance, Yu et al. [19] developed a regression-calibration and moment-corrected approach to adjust for measurement errors while accounting for informative censoring. Yu et al. [19] modeled time-to-event data and incorporated informative censoring using a shared frailty model. Yu et al. [20], on the other hand, modeled informative censoring using a shared frailty, as in [19], but relaxed the distributional assumption on the errors and proposed a general nonparametric missing-at-random model to account for the errors. To the best of

our knowledge, the most recent manuscript dealing with the topic of recurrent events is that of [21]. They investigated another aspect of failure-time data, namely, left truncation. Since right-censored and left-truncated data are quite prevalent in practice, Chen and Yi [21] developed models to simultaneously handle both features of the data while modeling measurement errors using moment correction.

To correct the bias induced by measurement errors in estimating model parameters with failure-time data, researchers have relied on the so-called induced hazard rates, which are defined as the conditional hazard given the observed covariates and the events history. The idea is to construct an unbiased score function, called *corrected score*, upon which estimation and inference are based rather than naive score (i.e., the score that ignores measurement errors), which uses an error-prone covariate and tends to yield spurious results of no practical value. The main contributors to this idea in single-event settings are [1,4,5]. Many authors have shown that the corrected score is not the gradient of a corrected likelihood, which led [5] to introduce the concept of *approximately corrected partial likelihood*. Later, Augustin [22] justified that the corrected scores proposed in [5,23] are exact and that their corresponding estimators are consistent. In light of the recent clarifications and based on the results in [22], we take the approach of corrected partial likelihood and consider a gap-time modeling of the intensity function with recurrent events when one or more covariates are measured with errors. We operate under the classical additive measurement errors model that is known to have a broad applicability in scientific research. We make the blanket assumption that the errors fluctuate around the covariates. Other general measurement error models such as regression calibration in [24,25] and Chapter 4 of [12] could also be used. In our current development, we do not impose any distributional assumption on the errors other than their variance–covariance matrix being time-independent and possessing a consistent estimator. We propose a corrected partial likelihood score process with root consistent estimators that follow a multivariate normal large sample distribution when properly standardized. Our results generalize those of [23] in two ways: (1) They are applied to recurrent event data by focusing on their dynamic feature, and (2) the errors do not follow any particular distribution in our approach. Kong and Gu [23] deals with single-event models with assumed normally distributed errors. We augment our work by deriving a corrected baseline cumulative hazard with recurrent events and its asymptotic properties. We add a discussion on misspecified error models and develop properties of the estimators under error misspecification. We show, in that case, that the estimator converges to a value different from the true parameter, enabling an evaluation of the bias and its magnitude. The correction methods proposed here will result in improved estimators from the technical point of view. There are numerous advantages, practically, such as better doctor directives, whether or not covariates are measured with errors in biomedical studies, and better products warranty in reliability and engineering, to name a few.

Our manuscript is organized as follows. In Section 2, we state the stochastic setting for our recurrent event model and define some key concepts that guide our theoretical development along with preliminary results. In Section 3, we propose the corrected estimator including the corrected estimating equations and develop their key asymptotic properties. In Section 4, we follow with a simulation study that serves as a validation of the theoretical properties, as we also illustrate our findings in an applied setting. In Section 5, we propose an estimator for the corrected baseline hazard and its asymptotic properties. In Section 6, we focus on the issues of misspecification of the error model and asymptotic bias. We finally conclude our work with a section on the conclusion, discussion, and recommendations.

2. Preliminaries

2.1. Dynamic Modeling and Observables

Consider n units that are monitored for the occurrences of an event that can recur up to a random time τ_i for each unit i. For unit i, let $S_{i,j}$ be the time of occurrence of the j^{th} event, and $T_{i,j}$ the gap between the $(j-1)$ and the j^{th} occurrence. For every unit i, a p-dimensional time varying covariate $\mathbf{x}_i$ is recorded. We assume, for $j = 1, ..., K_i$, that the $T_{i,j}$s have a distribution function $F_i(\cdot)$, and a hazard function $\lambda_i(t|\mathbf{x}) \equiv \lambda(t \mid \mathbf{x})$ that is a function of the covariates, with K_i being the total number of events per unit satisfying the sum quota scheme. For unit i, the observables at the censoring time τ_i are

$$\boldsymbol{\mathcal{O}}_i = (K_i, \tau_i, T_{i,1}, \ldots, T_{i,K_i}, \tau_i - S_{i,K_i}, \mathbf{x}_{i,1}(s_{i1}), ..., \mathbf{x}_{i,K_i}(s_{iK_i})), \tag{1}$$

which define an aggregate vector $\boldsymbol{\mathcal{O}} = (\boldsymbol{\mathcal{O}}_1, ..., \boldsymbol{\mathcal{O}}_n)$ over the observed sample of size n. In what follows, s represents calendar time whereas t represents gap time. The relevant counting processes from the data on which estimation will be conducted are $N_i^\dagger = \{N_i^\dagger(s) : s \leq \tau_i\}$, $Y_i^\dagger = \{Y_i^\dagger(s) : s \geq 0\}$, where $N_i^\dagger(s) = \sum_{j=1}^{\infty} I\{S_{i,j} \leq s \wedge \tau_i\}$, and $Y_i^\dagger(s) = I\{\tau_i \geq s\}$. The process $N_i^\dagger(s)$ determines, for subject i, the event occurrences up to time s, whereas the $Y_i^\dagger(s)$ process determines if the unit is at risk for future recurrences. We write the effective age of the unit i at time s as $\varphi_i(s) = s - S_{i,N_i^\dagger(s^-)}$. Observe that $0 \leq \varphi(s) \leq s$, and is viewed as a process that keeps track of the time elapsed since the last occurrence of an event. In φ, not only do we track the time elapsed but, also, we verify whether the time being tracked is a true event time versus another time with the hopes of a future event. Following [26], we provide a link among the effective age process, the regressors, and the multiplicative intensity process by $\lambda(s|\mathbf{x}_i(s)) = \lambda_0(\varphi_i(s)) \exp(\boldsymbol{\beta}'\mathbf{x}_i(s))$, where $\boldsymbol{\beta} \in \mathcal{R}^p$ is a set of regressor parameters. The corresponding survivor function would then be $\bar{F}_i(s|\mathbf{x}_i(s))$. The utility of the effective age rendering of the Cox model is that it allows to explicitly model the effect of an intervention or treatment performed just after an event occurrence, an adaptation that is more in harmony with settings where treatments are administered during an observation window such as in health science. From the theory of stochastic integration, the compensator process of $N_i^\dagger(s)$ is $A_i^\dagger(s)$ given by $A_i^\dagger(s|\boldsymbol{\beta}) = \int_0^s Y_i^\dagger(v)\lambda_0[\varphi_i(v)] \exp[\boldsymbol{\beta}'\mathbf{x}_i(v)]dv$. Due to the randomness of the argument $\varphi_i(v)$ in $A_i^\dagger(s|\boldsymbol{\beta})$, one usually works with the doubly-indexed processes $N_i(s,t)$ and $A_i(s,t|\boldsymbol{\beta})$ that are functions of both the calendar and gap times in order to handle the random argument in the integrand. The $N_i(s,t)$ and $A_i(s,t|\boldsymbol{\beta})$ processes are the number of events that occurred by calendar time s for unit i whose effective age is, at most, gap time t and their associated doubly-indexed compensator. Hence, for fixed t, $M_i(s,t|\boldsymbol{\beta}) = N_i(s,t) - A_i(s,t|\boldsymbol{\beta})$ is a zero-mean square integrable martingale. More information the notations and details on the stochastic processes formulation in this section can be found in [27].

2.2. Measurement Errors Notations

Under the additive model for errors, for unit i, at calendar time s, let $\mathbf{x}_i(s)$ be the p-dimensional covariates, of which possibly $q \leq p$ are measured with errors. If $\boldsymbol{\epsilon}_i(s) = (\epsilon_{i1}(s), ..., \epsilon_{ip}(s))$ is the p-dimensional errors on $\mathbf{x}_i(s) = (x_{i1}(s), ..., x_{ip}(s))$, then

$$\mathbf{x}_i(s) = \mathbf{z}_i(s) + \boldsymbol{\epsilon}_i(s),$$

where the $\mathbf{z}_i(s)$ are the true and unobserved covariates. We do not impose any distributional assumption on the $\boldsymbol{\epsilon}_i(s)$, other than having a zero mean and a variance covariance matrix $\boldsymbol{\Xi}$ that is time-independent. We assume the existence of a consistent estimator $\hat{\boldsymbol{\Xi}}$ of $\boldsymbol{\Xi}$ that can be derived using validation methods or replicates (cf. [12], Chapter 4). It is to be noted that some authors have assumed the errors to have a multivariate normal distribution with

error variance obtained using validation data and sample variance formulas; cf. section 2.1 of [17]. See also section 4.7 of [12] on ways to derive estimators for the error variance. We now introduce some notation in the sequel, to be used throughout this manuscript. All random entities are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The space $D[0,t]$ denotes the cadlag functions on $[0,t]$ equipped with the supremum norm $\|\cdot\|_\infty$, and all asymptotic results are taken as $n \to \infty$. The notations $\xrightarrow{d}$, $\xrightarrow{as}$, and $\xrightarrow{p}$, respectively, denote convergence in distribution, almost sure convergence, and convergence in probability. For a vector $\mathbf{a} = (a_1, \ldots, a_p)$, $\mathbf{a}'$ is its transpose; if $\mathbf{b}$ is also a p-dimensional vector, then $\mathbf{a} \otimes \mathbf{b}$ is the $p \times p$ matrix $\mathbf{a}\mathbf{b}'$ with $(k,l)^{th}$ element $a_k b_l$. In addition, $\mathbf{a}^{\otimes 2} = \mathbf{a} \otimes \mathbf{a}$, and for the vector $\mathbf{a}$, $\|\mathbf{a}\| = \sup_k |a_k|$. Finally, we define the gradient operator $\nabla_{\boldsymbol{\beta}}$ by the vector of partial derivatives $\nabla_{\boldsymbol{\beta}} = \partial/\partial\boldsymbol{\beta} \equiv (\partial/\partial\beta_l, l = 1, 2, \ldots, p)'$; where $\boldsymbol{\beta}$ a p-dimensional vector.

3. Corrected Scores and Estimators

3.1. Preliminary

Some preliminary results are contained in [28]; consequently, we will not provide a full exposition here again. With $s^\star = \max_{1 \le i \le n} \tau_i$, the generalized at-risk process is defined as

$$\begin{aligned} Y_i(s, \mathbf{x}(t)|\boldsymbol{\beta}) &= \left[\sum_{j=1}^{N_i^\dagger((s \wedge \tau_i)-)} I(T_{i,j} \ge t) + I((s \wedge \tau_i) - S_{i,N_i^\dagger((s\wedge\tau_i)-)} \ge t) \right] \cdot \exp(\boldsymbol{\beta}'\mathbf{x}_i(t)) \\ &:= h_i(s,t) \cdot \exp(\boldsymbol{\beta}'\mathbf{x}_i(t)). \end{aligned}$$

We write $S^{(0)}(s, \mathbf{x}(t)|\boldsymbol{\beta}) := n^{-1} \sum_{i=1}^n Y_i(s, \mathbf{x}(t)|\boldsymbol{\beta})$, and its k^{th} order derivative with respect to $\boldsymbol{\beta}$ is written as

$$\begin{aligned} S^{(k)}(s, \mathbf{x}(t)|\boldsymbol{\beta}) &= \frac{1}{n} \boldsymbol{\nabla}_{\boldsymbol{\beta}}^{(k)} \sum_{i=1}^n h_i(s,t) \cdot \exp(\boldsymbol{\beta}'\mathbf{x}_i(t)) \\ &= \frac{1}{n} \sum_{i=1}^n h_i(s,t) \cdot \mathbf{x}_i(t)^{\otimes k} \exp(\boldsymbol{\beta}'\mathbf{x}_i(t)). \end{aligned}$$

Taking $k = 0, 1, 2$, write

$$E(s, \mathbf{x}(t)|\boldsymbol{\beta}) = S^{(1)}(s, \mathbf{x}(t)|\boldsymbol{\beta}) / S^{(0)}(s, \mathbf{x}(t)|\boldsymbol{\beta})$$

and

$$V(s, \mathbf{x}(t)|\boldsymbol{\beta}) = [S^{(2)}(s, \mathbf{x}(t)|\boldsymbol{\beta}) / S^{(0)}(s, \mathbf{x}(t)|\boldsymbol{\beta})] - E(s, \mathbf{x}(t)|\boldsymbol{\beta})^{\otimes 2}.$$

The log partial likelihood is written as

$$l_p(\boldsymbol{\beta}, \mathbf{x}(s^\star)) = \sum_{i=1}^n \int_0^{s^\star} \left[\boldsymbol{\beta}'\mathbf{x}_i(v) - \log[Y_i(s^\star, \varphi_i(v)|\boldsymbol{\beta})] \right] N_i^\dagger(dv), \tag{2}$$

where the integral is obtained over the calendar time process. Because $\varphi_i(v)$ is a random variable, a change of variable $\varphi_i(v) = w$ in the integrand leads to a likelihood profile $l_P(\boldsymbol{\beta}, \mathbf{x}(s^\star))$ and an associated score process $\mathbf{U}(\boldsymbol{\beta}, \mathbf{x}(s^\star)) = \nabla_{\boldsymbol{\beta}} l_P(\boldsymbol{\beta}, \mathbf{x}(s^\star))$ given by

$$\mathbf{U}(\boldsymbol{\beta}, \mathbf{x}(t^\star)) = \sum_{i=1}^n \int_0^{t^\star} [\mathbf{x}_i(\varphi_i^{-1}(w)) - E(s^\star, w)] N_i(s^\star, dw), \tag{3}$$

where $t^\star = \max_{i,j} T_{i,j}$ is the maximum gap time. The estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$, say, is the solution to $\mathbf{U}(\boldsymbol{\beta}, \mathbf{x}(t^\star)) = \mathbf{0}$, with predictors assumed error-free. Note that $\mathbf{x}_i(s)$ is not the true

covariate; consequently, the solution $\hat{\boldsymbol{\beta}}$ is biased by virtue of the biasedness of $\mathbf{U}(\boldsymbol{\beta}, \mathbf{x}(t^\star))$. As $\mathbf{U}(\hat{\boldsymbol{\beta}}, \mathbf{x}(t^\star))$ does not equate zero in the presence of measurement errors, the corresponding likelihood and score which are functions of $S^{(k)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$, $k = 0, 1$ are also biased, and cannot induce reasonable estimating equations. We provide a corrected expression for $\mathbf{U}(\boldsymbol{\beta}, \mathbf{x}(t^\star))$, which is inextricably linked to corrections of $S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$ and $S^{(1)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$. We denote $E_{\boldsymbol{\epsilon}}(\cdot|\boldsymbol{\mathcal{O}})$, the conditional expectation under the distribution of $\boldsymbol{\epsilon}$ with respect to the observables $\boldsymbol{\mathcal{O}}$. In the spirit of corrected likelihoods, we seek estimating functions $\breve{\mathbf{U}}(\boldsymbol{\beta}, \mathbf{x}(t))$, expressed in terms of the observed data and satisfying

$$E_{\boldsymbol{\epsilon}}\big(\breve{\mathbf{U}}(\boldsymbol{\beta}, \mathbf{x}(t)|\boldsymbol{\mathcal{O}})\big) = \mathbf{U}(\boldsymbol{\beta}, \mathbf{z}(t)). \tag{4}$$

This will be handled through the use of the first-order approximation of the ratio of expectations. We also operate under some regularity conditions that are detailed in Appendix A. Under these regularity conditions, $\boldsymbol{\epsilon}_i$s are independent zero-means with a time-independent covariance structure; all-order moment-generating functions of the error distribution exist, the integrated hazard is finite, the covariates are uniformly bounded and cannot escape to infinity, and uniform continuity of $S^{(k)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$ and its expectations is guaranteed.

Proposition 1. *The following results holds:*

$$E(s^\star, \mathbf{x}(t)|\boldsymbol{\beta}) = \frac{S^{(1)}(s^\star, \mathbf{z}(t))}{S^{(0)}(s^\star, \mathbf{z}(t))} + \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ln(\phi(\boldsymbol{\beta})).$$

Proof. See Appendix A □

Remark 1. *There are many ways of dealing with measurement errors in estimating parameters that connect event time to covariates; specifically,*

(*i*) *likelihood-based, where a specification of the distribution of the covariates is usually conjectured;*
(*ii*) *methods based on unbiased estimating functions;*
(*iii*) *methods based on correcting a naive parameter estimate, i.e., the one obtained ignoring the errors.*

Each of these approaches has its pros and cons. The estimating-functions approach has the advantage of bypassing the difficult and unrealistic, if not counterintuitive and circular, step of specifying the distribution of the variable measured with error as required by (*i*). The approach in (*iii*) is primarily a simulation-based approach that does not serve us well for a rigorous mathematical derivation of asymptotic properties. In addition, (*iii*) relies on the use of bridge functions to connect the naive estimator in-probability properties to the corrected ones, beyond its requirement of the known measurement errors covariance structure needed for the simulation extrapolation's best performance. By settling for (*ii*), we maintain the statistical constructs needed for a valid estimating function, i.e., the ease of computability *vis-à-vis* the measurement error variable and the ability to produce consistent regressor parameters with asymptotic normality under suitable regularity conditions. As an example, similar to the argument about exact corrected scores and root consistency as those put forth in [5,22,23], we show that the results hold for recurrent events as well. We also note that when $\boldsymbol{\epsilon}(t) \equiv \mathbf{0}$ in (A1), one recovers the no-measurement-errors covariates as in [28].

3.2. Large Sample Properties Under Correction

In order to establish the convergence in distribution of the standardized version of $\breve{\boldsymbol{\beta}}$, we would need the large sample properties of the scaled corrected score $\sqrt{n}^{-1}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)$.

Assume $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, and $\mathbf{z}_i(t)$ is the true covariate free of measurement errors; then, the process

$$M_i(s,t) = N_i(s,t) - \int_0^t Y_i(s,\mathbf{z}(w)|\boldsymbol{\beta}_0)\lambda_0(w)dw$$

is, for fixed t, a zero-mean martingale with respect to the calendar time filtration $\mathfrak{F} = \{\mathfrak{F}_s : s \geq 0\}$, with $\mathfrak{F}_s$ the σ-field generated by $\{[(N_i^\dagger(s), Y_i^\dagger(s+)) : s \geq 0], i = 1,2,\ldots,n\}$. The first result in this section pertains to the consistency of the corrected maximum likelihood $\breve{\boldsymbol{\beta}}$.

Theorem 1. *As $n \to \infty$, the sequence of solutions $\breve{\boldsymbol{\beta}}_n = (\breve{\beta}_{1n},...,\breve{\beta}_{np})$ to the equations $\breve{\mathbf{U}}_n(\boldsymbol{\beta}; s^\star, t) = \mathbf{0}$, $n = 1,2,...$ is consistent.*

Proof. See the Appendix A. □

The moment-generating function of the $\boldsymbol{\epsilon}_i(s)$, $\phi_i(\boldsymbol{\beta}) = E(e^{\boldsymbol{\beta}'\boldsymbol{\epsilon}_i(s)})$ plays a key role in the asymptotic properties of corrected estimators, the corrected score, and corrected Fisher information. In the next few lines, we discuss its large sample property. For every i, note that $\boldsymbol{\beta}'\boldsymbol{\epsilon}_i(s) = \sum_{j=1}^p \beta_j \epsilon_{ij}(s)$. The function $\phi_i(\boldsymbol{\beta})$ is assumed to exist for $\boldsymbol{\beta}$ in the neighborhood of zero and is twice differentiable with respect to $\boldsymbol{\beta}$. If a consistent estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ exists, the following lemma asserts the consistency of an estimator of $\phi(\boldsymbol{\beta})$, namely, $\hat{\phi}(\boldsymbol{\beta}) := \phi(\hat{\boldsymbol{\beta}})$, the empirical moment-generating function.

Lemma 1. *Let $\hat{\boldsymbol{\beta}}$ be the corrected and consistent maximum likelihood estimator of $\boldsymbol{\beta}$. The empirical moment-generating function of $\phi(\boldsymbol{\beta})$ based on $\{\phi_i(\boldsymbol{\beta}) : i = 1,...,n\}$ is defined by*

$$\hat{\phi}(\boldsymbol{\beta}) = \phi(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_{i=1}^n \phi_i(\hat{\boldsymbol{\beta}}),$$

and is a consistent estimator of $\phi(\boldsymbol{\beta})$.

Proof. The result follows from the functional continuous mapping theorem, the uniform law of large numbers, and the consistency of $\hat{\boldsymbol{\beta}}$. □

As indicated earlier, to establish the distribution limit of the properly standardized corrected maximum likelihood estimator, we will first establish the asymptotic properties of the corrected score $\sqrt{n}^{-1}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)$.

Theorem 2. *As $n \to \infty$, the process $\left\{\frac{1}{\sqrt{n}}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) : t \in [0, t^\star]\right\}$ converges weakly on $D[0, t^\star]$ to a zero-mean Gaussian process $\boldsymbol{U}^\infty(\boldsymbol{\beta}_0, s^\star, t)$ with covariance matrix given by $\Psi(s, t_1 \wedge t_2|\boldsymbol{\beta}_0))$ that can be estimated by $\breve{\Psi}(s, t_1 \wedge t_2|\breve{\boldsymbol{\beta}})) = E([V_i(s, t_1 \wedge t_2)^{\otimes 2}])$, with $V_i(s,t)$ given in the proof below.*

Proof. See the Appendix A. □

With the limiting distribution of $\sqrt{n}^{-1}\breve{\mathbf{U}}_n(\boldsymbol{\beta}; s^\star, t)$ established, we can now proceed with that of the properly standardized corrected maximum likelihood estimator. To that end, we need a corrected Fisher information matrix and its in-probability limit. Recall that the Fisher information matrix is given by

$$\boldsymbol{\mathcal{I}}_n(\breve{\boldsymbol{\beta}}; s^\star, t) = -\frac{1}{n}\boldsymbol{\nabla}_{\boldsymbol{\beta}'}\breve{\mathbf{U}}(\breve{\boldsymbol{\beta}}; s^\star, t).$$

We have

$$\begin{aligned}\boldsymbol{\mathcal{I}}_n(\boldsymbol{\beta}; s^\star, t) &= -\frac{1}{n}\nabla_{\boldsymbol{\beta}'}\breve{\mathbf{U}}(\boldsymbol{\beta}; s, t) \\ &= -\frac{1}{n}\int_0^t \left\{\nabla_{\boldsymbol{\beta}'}[\nabla_{\boldsymbol{\beta}} \ln \phi(\boldsymbol{\beta})] - \frac{S^{(2)}(s^\star, t|\boldsymbol{\beta})}{S^{(0)}(s^\star, t|\boldsymbol{\beta})} + E^{\otimes 2}(s^\star, t|\boldsymbol{\beta})\right\}\sum_{i=1}^n N_i(s^\star, dw).\end{aligned}$$

To obtain the corrected Fisher information, we need to find the corrected expression of the second and third terms in the integrand. The third corrected expression has been provided in Proposition 1. The variance–covariance matrix of $\boldsymbol{\epsilon}_i$ is assumed to exist and is independent of t, that is, $\boldsymbol{\Xi} = E(\boldsymbol{\epsilon}_i^{\otimes 2})$, $i = 1, ..., n$ is independent of t. The next proposition pertains to the corrected expression of the second term; i.e.,

$$S^{(2)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})[S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})]^{-1}.$$

Proposition 2. *The corrected expression of* $S^{(2)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})[S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})]^{-1}$ *can be approximated by*

$$\begin{aligned}\frac{S^{(2)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})}{S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})} \approx & \frac{S^{(2)}(s^\star, \mathbf{z}(t)|\boldsymbol{\beta})}{S^{(0)}(s^\star, \mathbf{z}(t)|\boldsymbol{\beta})} + [E(s^\star, \mathbf{z}(t)|\boldsymbol{\beta})(\boldsymbol{\beta}\boldsymbol{\Xi})]' \\ & + [E(s^\star, \mathbf{z}(t)|\boldsymbol{\beta})(\boldsymbol{\beta}\boldsymbol{\Xi})] + \boldsymbol{\Xi} + [(\boldsymbol{\Xi}\boldsymbol{\beta})(\boldsymbol{\Xi}\boldsymbol{\beta})'].\end{aligned}$$

Proof. See the Appendix A. □

In light of the above proposition, the corrected Fisher information matrix is given by

$$\boldsymbol{\mathcal{I}}_n(\boldsymbol{\beta}; s^\star, t^\star) = -\frac{1}{n}\int_0^{t\star}\left[E(s^\star, \mathbf{x}(w)|\boldsymbol{\beta})^{\otimes 2} - \frac{S^{(2)}(s^\star, \mathbf{z}(w)|\boldsymbol{\beta})}{S^{(0)}(s^\star, \mathbf{z}(w)|\boldsymbol{\beta})}\right]\sum_{i=1}^{n} N_i(s^\star, dw).$$

It can be seen that the corrected Fisher information based on the true values of the covariates $\mathbf{z}$ no longer contains any information about the errors terms. This concurs with the results in [23] and those in [22]. Moreover, the Fisher information matrix can, therefore, be estimated by $\frac{1}{n}\check{\mathbf{I}}(\check{\boldsymbol{\beta}}; \mathbf{x}(t), s^\star)$, which is a consistent estimator of $\frac{1}{n}\check{\mathbf{I}}(\check{\boldsymbol{\beta}}; \mathbf{z}(t), s^\star)$ by virtue of the consistency of $\check{\boldsymbol{\beta}}$ and the corrected expression. The final result in this section is regarding the asymptotic property of the properly standardized corrected maximum partial likelihood estimators.

Theorem 3. *As* $n \to \infty$, $\sqrt{n}(\check{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0))$, *where* $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}_0; s^\star, t^\star)$ *can be consistently estimated by*

$$\check{\boldsymbol{\Sigma}}^{-1} = \left([\boldsymbol{\mathcal{I}}_n(\check{\boldsymbol{\beta}}; \mathbf{x}(t), s^\star)]^{-1}\right)'\check{\Psi}(\check{\boldsymbol{\beta}}; s, t)[\boldsymbol{\mathcal{I}}_n(\check{\boldsymbol{\beta}}; \mathbf{x}(t), s^\star)]^{-1}.$$

The proof follows from the usual Taylor expansion of the corrected score around $\boldsymbol{\beta}_0$, the limiting distribution of $\sqrt{n}^{-1}\check{\mathbf{U}}_n(\boldsymbol{\beta}; s, t)$, and the consistency of the corrected Fisher information matrix. The limiting variance is finally obtained using multivariate distribution theory.

3.3. Measurement Errors Variance Estimation

Estimating the measurement errors variance is useful in determining whether a simplified approach that ignores it would be acceptable. The size of the error variance is also a factor that is associated with bias in parameter estimates. In order to address this estimation, it is not uncommon to proceed by bootstrapping. Two approaches are typically used to achieve this goal: the method proposed by [29], which involves bootstrapping the empirical distribution, and that of [30,31], which relies on bootstrap resampling of the observables data. If the empirical distribution is the basis for bootstrapping, then, given $K_i = l_i$, one can obtain the empirical function of the $\mathbf{x}_i$ as

$$F_n(w|\mathbf{x}) = \frac{1}{\sum_{i=1}^{n} l_i}\sum_{i=1}^{n}\sum_{j=1}^{l_i}\mathrm{I}\{\mathbf{x}_i(j) \leq w\}$$

and obtain draws from the empirical function. One could also deal directly with the observables, target the within-subject observations as replicates, and derive an estimated variance–covariance structure $\hat{\Xi}$ by restricting to subjects with multiple events and obtain

$$\hat{\Xi} = \frac{\sum_{i=1}^{n} I\{N_i^{\dagger}((s \wedge \tau_i)-) > 0\} \sum_{j=1}^{N_i^{\dagger}((s\wedge\tau_i)-)} (\mathbf{x}_{i,j}(s_{i,j}) - \bar{\mathbf{x}}_{i\cdot})^{\otimes 2}}{\sum_{i=1}^{n} [N_i^{\dagger}((s \wedge \tau_i)-) - 1]},$$

$$\bar{\mathbf{x}}_{i\cdot} = \frac{\sum_{j=1}^{N_i^{\dagger}((s\wedge\tau_i)-)} \mathbf{x}_{i,j}(s_{i,j})}{N_i^{\dagger}((s \wedge \tau_i)-)},$$

and re-sample this estimate for smoothing purposes. Regardless of the approach taken, since the empirical distribution function is a consistent estimator of the true distribution function, the individual components of $\hat{\Xi}$ converge almost surely to their true values.

Theorem 4. *The components of $\hat{\sigma}^2_{qr}$ of $\hat{\Xi}$ satisfy $\hat{\sigma}^2_{qr} \xrightarrow{as} \sigma^2_{qr}$ under the bootstrap probability. That is, the coordinate-wise bootstrap estimator converges in probability to its counterpart in the true variance of the errors.*

Proof. See the Appendix A. □

4. Simulation and Application

We perform a simulation study to assess the accuracy of parameter estimation with respect to the corrected likelihood process. We operate under a Weibull model and simulate a wide range of inter-event times that fit recurrent processes in real-life settings. Among these settings, we explore reliability deterioration in time (Weibull$(1, \theta_2) : \theta_2 > 1$, which corresponds to shortened inter-event times or a high frequency of occurrence), reliability improvement in time (Weibull$(1, \theta_2) : \theta_2 < 1$, which corresponds to lengthier inter-event times), and constant reliability where $\theta_2 = 1$. We operate under a two-covariate CPH model $\mathbf{z} = (z_1, z_2)$, with $z_1 = z_1(s)$ a continuous time-varying covariate and z_2 an equiprobability Bernoulli process indicating a grouping factor. On z_1, we superpose a Gaussian noise with variance σ^2 to create x_1, and we explore the effect of the measurement error variance on the Cox regression parameters β_1, β_2 in $\beta_1 x_1 + \beta_2 z_2 \equiv \beta_1 x_1 + \beta_2 x_2 \equiv \boldsymbol{\beta}'\mathbf{x}$. We henceforth refer to the analytic model that overlooks the measurement errors as a *naive* fit, and to the one that incorporates the measurement error in estimation *via* corrected partial likelihood as the *corrected* fit. Since a simulation is a controlled environment, the true β values are known by design. We explore both small and large sample scenarios with $n = (40, 80, 100, 200)$. The data are simulated for an average of four recurrent events per subject. We keep $\beta_2 = 1$ fixed and varied $\beta_1 \in (-1, -0.8, -0.6, -0.4, -0.2, -0.1, 0.1, 0.2, 0.4, 0.6, 0.8, 1)$. We report the results for Weibull$(1, \theta_2 = 1.5)$ inter-event times, though the results carry similar information across Weibull$(1, \theta_2 > 1)$. We next design another simulation setting where the Cox regression coefficients are kept at $(\beta_1, \beta_2) = (-1, 1)$, while we vary the measurement error $\sigma \in (0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5)$ to aid our understanding of the size of a measurement error, which has the potential to seriously affect the estimation of the Cox parameters.

4.1. Simulation Results

The simulation results are summarized in Figure 1 and Tables 1–3. The pictorial representation in Figure 1 encapsulates the advantages of using estimation based on the corrected partial score over the naive approach in parameter estimation for predictors recorded with measurement error, while in the naive case, bias increases with the magnitude of the effect of the predictor on the hazard of recurrence. We clearly see that the corrected

partial likelihood has created an estimation scheme with minimal, if not nonexistent, bias. This observation holds irrespective of n. Overall, the standard error trivially decreases with n, and the parameter estimate for the predictor without measurement error has been estimated with minimal bias in both the partial and the corrected partial scores. We measure the accuracy of the estimation of the parameter without measurement error by using the mean square error. We call `MSE(naive)` the mean squared error for the estimation based on the partial likelihood, and `MSE(corrected)` the mean squared error for the estimation based on the corrected partial likelihood. Expectedly, these values are negligible and asymptotically similar. It is expected that correction would have minimal and negligible effects on this covariate. The results appear to be in line with the published literature on single events, where the corrected likelihood outperforms the naive approach.

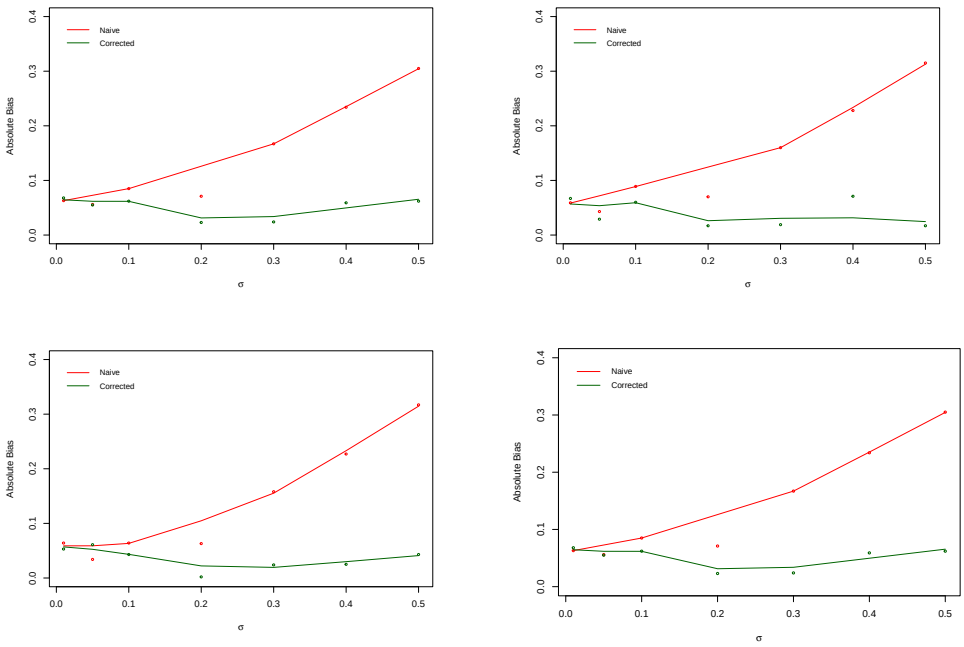


Figure 1. Bias assessment comparing the naive approach (in red) to the corrected (in green). The points are joined by the lowess estimating curve. From left to right, top graphs are $n = 40$ and 80, and bottom graphs are $n = 100$ and 200, respectively.

Table 1. Weibull intensity: $\bar{F}(t;\boldsymbol{\theta}) = e^{-(\theta_2 t)^{\theta_1}}; \theta_2 = 1; \theta_1 = 1.5\ \sigma = 0.5$.

						$\mathcal{N}(0, 0.1)$				
n	β_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_{1c}$	$se(\hat{\beta}_{1c})$	β_2	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_{2c}$	$se(\hat{\beta}_{2c})$
	−1.0	−0.839	0.095	−0.980	0.143		0.978	0.238	1.036	0.240
	−0.8	−0.678	0.109	−0.836	0.158		1.039	0.214	1.112	0.217
	−0.4	−0.365	0.117	−0.415	0.106		1.054	0.238	1.049	0.240
	−0.2	−0.177	0.117	−0.230	0.128		1.062	0.249	1.040	0.247
40	−0.1	−0.078	0.080	−0.081	0.120	1	1.002	0.248	1.040	0.230
	0.1	0.096	0.085	0.093	0.124		1.052	0.238	1.063	0.193
	0.2	0.170	0.079	0.183	0.125		1.019	0.260	1.028	0.207
	0.4	0.347	0.101	0.411	0.134		1.013	0.208	1.024	0.256
	0.8	0.709	0.138	0.773	0.150		1.067	0.260	1.079	0.244
	1.0	0.859	0.116	1.021	0.169		0.968	0.196	1.111	0.270

Table 1. *Cont.*

					$\mathcal{N}(0, 0.1)$					
n	β_1	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_{1c}$	$se(\hat{\beta}_{1c})$	β_2	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_{2c}$	$se(\hat{\beta}_{2c})$
	−1.0	−0.840	0.079	−0.947	0.093		0.991	0.153	1.008	0.170
	−0.8	−0.687	0.079	−0.801	0.088		1.018	0.166	1.052	0.187
	−0.4	−0.342	0.064	−0.395	0.084		1.005	0.171	1.030	0.156
	−0.2	−0.187	0.062	−0.187	0.079		1.035	0.165	1.031	0.158
80	−0.1	−0.083	0.068	−0.105	0.059	1	1.019	0.164	0.976	0.154
	0.1	0.075	0.062	0.090	0.078		1.026	0.177	1.051	0.131
	0.2	0.195	0.062	0.187	0.056		1.035	0.128	1.024	0.160
	0.4	0.338	0.064	0.385	0.089		1.006	0.139	0.980	0.139
	0.8	0.665	0.075	0.784	0.093		0.972	0.156	1.046	0.146
	1.0	0.831	0.070	0.959	0.100		1.002	0.137	1.029	0.158
	−1.0	−0.826	0.069	−0.984	0.100		0.960	0.132	1.058	0.140
	−0.8	−0.674	0.060	−0.807	0.086		0.984	0.127	1.017	0.127
	−0.4	−0.354	0.063	−0.388	0.066		1.040	0.151	1.007	0.119
	−0.2	−0.177	0.060	−0.199	0.068		1.031	0.146	1.006	0.135
	−0.1	−0.086	0.064	−0.101	0.058		0.993	0.126	1.033	0.137
100	0.1	0.093	0.059	0.103	0.056	1	1.028	0.142	0.988	0.131
	0.2	0.183	0.063	0.211	0.072		1.057	0.140	1.029	0.115
	0.4	0.367	0.063	0.399	0.070		1.016	0.144	1.017	0.115
	0.8	0.685	0.070	0.784	0.083		1.011	0.141	1.026	0.166
	1.0	0.839	0.068	0.965	0.094		0.989	0.141	1.010	0.149
	−1.0	−0.821	0.039	−0.962	0.065		1.018	0.089	1.004	0.106
	−0.8	−0.668	0.053	−0.763	0.051		0.969	0.097	1.026	0.103
	−0.4	−0.338	0.039	−0.401	0.051		1.011	0.112	1.022	0.103
	−0.2	−0.182	0.046	−0.210	0.049		1.028	0.105	1.014	0.093
	−0.1	−0.092	0.045	−0.109	0.044		1.015	0.081	1.012	0.093
200	0.1	0.079	0.035	0.101	0.043	1	1.013	0.081	1.048	0.109
	0.2	0.177	0.042	0.208	0.047		1.016	0.099	1.022	0.089
	0.4	0.348	0.046	0.404	0.040		1.004	0.095	1.033	0.100
	0.8	0.678	0.043	0.758	0.050		0.976	0.107	1.001	0.099
	1.0	0.832	0.057	0.968	0.067		1.015	0.101	1.036	0.111

Table 2. MSE($\widehat{\beta}_2$): Error for the covariate with no measurement error.

n	40	80	100	200
`MSE(naive)`	0.0016	0.0005	0.0009	0.0004
`MSE(corrected)`	0.0041	0.0011	0.0007	0.0006

Table 3. Weibull intensity: $\bar{F}(t;\boldsymbol{\theta}) = e^{-(\theta_2 t)^{\theta_1}}; \theta_2 = 1; \theta_1 = 1.5; (\beta_1, \beta_2) = (-1, 1)$.

					$\mathcal{N}(0, \sigma)$				
n	σ	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_{1c}$	$se(\hat{\beta}_{1c})$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_{2c}$	$se(\hat{\beta}_{2c})$
	0.5	−0.704	0.085	−1.024	0.228	0.988	0.231	1.046	0.295
	0.4	−0.737	0.079	−0.980	0.193	1.012	0.224	1.129	0.265
	0.3	−0.848	0.094	−0.965	0.155	0.960	0.231	1.042	0.256
40	0.2	−0.909	0.107	−0.979	0.135	0.979	0.237	1.055	0.243
	0.1	−0.926	0.099	−0.911	0.116	1.029	0.227	1.036	0.193
	0.05	−0.947	0.122	−0.976	0.120	1.018	0.211	1.099	0.234
	0.01	−0.938	0.144	−0.928	0.134	0.956	0.177	1.044	0.227
	0.5	−0.685	0.071	−0.983	0.144	0.916	0.155	1.055	0.181
	0.4	−0.772	0.086	−0.929	0.086	0.966	0.152	1.049	0.162
	0.3	−0.840	0.077	−0.981	0.115	0.960	0.126	1.003	0.130
80	0.2	−0.930	0.091	−0.983	0.089	1.005	0.128	1.025	0.150
	0.1	−0.911	0.066	−0.940	0.082	0.989	0.135	1.056	0.138
	0.05	−0.957	0.081	−0.971	0.080	1.068	0.124	1.014	0.157
	0.01	−0.941	0.089	−0.933	0.071	1.005	0.132	1.001	0.161

Table 3. *Cont.*

n	σ	$\hat{\beta}_1$	$se(\hat{\beta}_1)$	$\hat{\beta}_{1c}$	$\mathcal{N}(0,\sigma)$ $se(\hat{\beta}_{1c})$	$\hat{\beta}_2$	$se(\hat{\beta}_2)$	$\hat{\beta}_{2c}$	$se(\hat{\beta}_{2c})$
100	0.5	−0.683	0.068	−0.957	0.139	0.956	0.133	1.015	0.173
	0.4	−0.773	0.056	−0.975	0.108	1.026	0.143	1.118	0.173
	0.3	−0.842	0.075	−0.976	0.090	0.971	0.127	1.058	0.133
	0.2	−0.937	0.074	−1.002	0.091	1.011	0.135	0.996	0.136
	0.1	−0.936	0.063	−0.957	0.078	0.997	0.137	1.042	0.124
	0.05	−0.966	0.066	−0.939	0.075	1.016	0.129	1.029	0.133
	0.01	−0.936	0.068	−0.947	0.071	1.011	0.154	1.024	0.120
200	0.5	−0.695	0.045	−0.938	0.065	0.950	0.084	1.022	0.104
	0.4	−0.766	0.042	−0.941	0.059	1.000	0.081	1.095	0.110
	0.3	−0.833	0.052	−0.976	0.057	0.995	0.084	1.021	0.085
	0.2	−0.929	0.051	−0.977	0.059	1.004	0.108	1.029	0.097
	0.1	−0.915	0.044	−0.938	0.047	0.978	0.076	1.004	0.094
	0.05	−0.944	0.055	−0.945	0.054	1.027	0.098	1.037	0.073
	0.01	−0.937	0.051	−0.932	0.049	1.013	0.109	1.010	0.087

4.2. Application: rhDNase Data

This is a pulmonary exacerbation study that has appeared in many publications, including [14,32], to name a few. This double-blind, randomized clinical trial aims to assess the effect of rhDNase, a recombinant deoxyribonuclease I enzyme, versus placebo, on respiratory exacerbations among patients with cystic fibrosis. Six hundred and forty-five patients were recruited for the trial, and each patient was followed up for about 169 days. We use a modified version of these data as an illustrative example. FEV measurements are well-known to be subject to measurement errors. This is why an argument was made regarding forced expiratory volume (FEV) measurements being recorded with errors, as two measurements taken a few minutes apart (FEV1 and FEV2) showed discrepancies. Even though the original dataset contains error-contaminated FEV values, it is of no use to us because the error variance is unavailable, both in the original dataset and in any literature where this dataset has been used. As we discussed in the theory part of this study, the error variance must be known to verify the accuracy of our proposed error variance estimator and the overall error-corrected parameter estimation procedure. Therefore, we first assumed that the original dataset was error-free. Later, the error contamination was simulated by making a modification to FEV values. Another issue with the original dataset was that the FEV covariate was time-independent. Since our study focuses on the more general case of time-dependent covariates, we made the FEV covariate time-dependent for each subject at each occurrence of respiratory exacerbation, using the subject-specific ordered measurements FEV1 and FEV2 according to a Uniform (FEV1, FEV2). The dataset with time-varying FEV is posted on our web domain https://www.myweb.uiowa.edu/gzamba/ (accessed on 10 November 2023) for a reproducibility check on our proposed methods. Since we assumed that the time-independent FEV values of the original dataset were error free, the modified time-dependent FEV values were also assumed to be error free. Next, to make time-dependent FEV values error-contaminated, we added a Gaussian noise with variance σ^2. We use $\sigma = 0$ as a reference value, which would be the basis for our comparison; degeneracy under both the corrected and uncorrected methods should yield similar results. We increase the error variance so that the resulting FEV value remains within the range of clinically acceptable and realistic pulmonary function values. By increasing the measurement error variance, we verify in this illustrative example how the naive approach would progressively lead us to a spurious inference about an FEV effect and how this inconvenience has been repaired through correction. We recognize that our method is not one-size-fits-all. The last row of Table 4 highlights points beyond which

estimation crumbles down for the corrected partial likelihood—a situation partly addressed by [23], showing that failure in estimation for the corrected partial score tends to occur when $|\beta_1\sigma| \geq 0.8$. However, in this example, estimation appears to remain reasonably stable up to $|\beta_1\sigma| \leq 1$. The breakdown observed in the corrected score coefficient estimates is due to the divergence of the iteration process.

Table 4. Regression parameter estimates.

σ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_{1c}$	$\hat{\beta}_{2c}$
0.000	−1.7025	−0.2734	−1.7004	−0.2729
0.050	−1.6201	−0.2695	−1.6884	−0.2678
0.100	−1.4338	−0.2665	−1.6763	−0.2644
0.150	−1.2153	−0.2646	−1.6612	−0.2603
0.200	−1.0102	−0.2636	−1.6446	−0.2566
0.300	−0.6944	−0.2633	−1.6125	−0.2467
0.400	−0.4910	−0.2640	−1.5873	−0.2370
0.500	−0.3610	−0.2649	−1.5774	−0.2240
0.600	−0.2752	−0.2657	−1.6094	−0.2438
0.700	−0.2166	−0.2665	421.8588	86.7149

5. Corrected Baseline Hazard and Properties

The baseline hazard plays a very important role in the Cox model. The covariates, whether time-varying or not, act multiplicatively on the baseline hazard and give real time failure rate at time t. The Breslow cumulative baseline hazard was developed based on error-free covariates. So, any errors in the covariates measurement hinder the interpretation of the baseline hazard. It has been shown in the lifetime literature that the baseline hazard estimator based on the true covariates is consistent and converges to Gaussian processes when properly standardized. However, when the covariates are error-prone, these large sample properties do not hold. As a consequence, a corrected baseline hazard needs to be developed. Some work has been done for right-censored data under the assumption of normally distributed errors, while large sample properties are derived such as in [23]. Below, we propose a corrected baseline hazard for recurrent events with covariates under measurement errors. Our corrected estimator generalizes the one in [22,23] to recurrent events and is not restricted to normally distributed errors. The cumulative baseline hazard estimator based on error-free covariates under the Cox model was derived in [28,33], and given by

$$\hat{\Lambda}_0(s^\star, \mathbf{x}(t)|\hat{\boldsymbol{\beta}}) = \int_0^t \frac{N(s^\star, dw)}{Y(s^\star, w|\hat{\boldsymbol{\beta}})},$$

where $\hat{\boldsymbol{\beta}}$ is the estimate based on error-free covariates. We seek a corrected estimator of $\Lambda_0(t)$, say, $\breve{\Lambda}_0(s, t; \mathbf{z}|\breve{\boldsymbol{\beta}})$ that satisfies

$$E_{\boldsymbol{\epsilon}}(\hat{\Lambda}_0(s^\star, \mathbf{x}(t)|\hat{\boldsymbol{\beta}})) = \breve{\Lambda}_0(s^\star, \mathbf{z}(t)|\breve{\boldsymbol{\beta}}).$$

To that end, we assume the interchangeability of the expectation and the integral, and we observe that from the Taylor expansion of the ratio

$$\begin{aligned} E_{\boldsymbol{\epsilon}}\left(\frac{1}{S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})}\right) &= \frac{1}{E_{\boldsymbol{\epsilon}}(S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta}))} \\ &= \frac{1}{\phi(\boldsymbol{\beta})S^{(0)}(s^\star, \mathbf{z}(t)|\boldsymbol{\beta})} + o_p(1). \end{aligned}$$

Then, it is not difficult to see that a corrected baseline cumulative hazard estimator for $\Lambda_0(t)$ based on the observables with covariates $\{\mathbf{x}_i(t) : i = 1, ..., n\}$ is given by

$$\breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}}) = \phi(\breve{\boldsymbol{\beta}}) \int_0^t \frac{N(s^\star, dw)}{S^{(0)}(s^\star, \mathbf{x}(t)|\breve{\boldsymbol{\beta}}))}, \tag{5}$$

where $\breve{\boldsymbol{\beta}}$ is the corrected estimate of $\boldsymbol{\beta}$. If the errors are assumed to be normally distributed, then we obtain a similarly corrected baseline hazard as those in [23]. The corrected cumulative hazard in (5) is also similar in form to that proposed by [22] in the discrete failure time case. The consistency of $\breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}})$ follows from the consistency of $\phi(\breve{\boldsymbol{\beta}})$ and $\hat{\Lambda}_0(s^\star, \mathbf{x}(t)|\hat{\boldsymbol{\beta}})$, as stated in Theorem 2 of [28].

Theorem 5. *As $n \to \infty$, the process $\{\sqrt{n}(\breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}}) - \Lambda_0(t))\}$ converges to a zero-mean Gaussian process with variance function given by $\Gamma(s, t_1 \wedge t_2)$.*

Proof. See the Appendix A. □

6. Misspecified Errors Model

A fundamental assumption underlying classical results on the properties of estimators obtained from the score process, namely, the maximum likelihood estimators, is that the distribution or model that determines the behavior under investigation is well specified. For example, thus far in this manuscript, we have assumed that the measurement errors have the additive form $\mathbf{x} = \mathbf{z} + \boldsymbol{\epsilon}$. However, the additive model for errors may be incorrect, and the true form may follow a different, unknown model. For instance, it might be the Berkson model $\mathbf{z} = \mathbf{x} + \boldsymbol{\epsilon}$, or a mixture of the Berkson and additive models, namely, $\mathbf{x} = \mathbf{u} + \boldsymbol{\epsilon}$, where $\mathbf{u}$ is a latent variable having a mean μ_u and a variance–covariance matrix Σ_u. The true error model can also be multiplicative or a transformed additive model. In many situations, it is hard to know how the errors act on the covariates. If the measurement errors model is not correctly specified, it is natural to ask what would result in properties such as consistency and the convergence in law of the properly standardized corrected partial likelihood estimator $\breve{\boldsymbol{\beta}}$. Does this estimator still converge asymptotically to a limit? If so, does this limit have any practical value? Does the asymptotic normality of the properly standardized $\breve{\boldsymbol{\beta}}$ still hold under the misspecified errors model? These questions can be answered using the theory of misspecified models. White and Domowitz [34] provides a unified framework on the properties of maximum likelihood estimators obtained under various types of misspecfication in different contexts.

Properties of Corrected Estimator Under Misspecified Errors

To set the stage for this subsection, let $l_P(\boldsymbol{\beta}; s^\star, t)$ be the log-profile likelihood process under the working errors model, and $\boldsymbol{\beta}_n^\star$ the solution to $l_P(\boldsymbol{\beta}; s^\star, t) = \mathbf{0}$, if one exists. Suppose our true measurement errors model is defined by some unknown function $\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})$. Further, let $\tilde{l}_P(\boldsymbol{\beta}; s^\star, t)$ be the corrected likelihood under the true measurement errors $\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})$. If $\boldsymbol{\beta}$ is in a compact subset $\mathcal{B}$ of $\Re^p$, the solution $\tilde{\boldsymbol{\beta}}_n$ to $\tilde{l}(\boldsymbol{\beta}; s^\star, t) = \mathbf{0}$ is defined by

$$\tilde{\boldsymbol{\beta}}_n = \text{argmax}_{\boldsymbol{\beta} \in \mathcal{B}} \tilde{l}_P(\boldsymbol{\beta}; s^\star, t).$$

Likewise,

$$\boldsymbol{\beta}_n^\star = \text{argmax}_{\boldsymbol{\beta} \in \mathcal{B}} l_P(\boldsymbol{\beta}; s^\star, t)$$

is the sequence of solution under the working likelihood. The first question is, does there exist a sequence of solutions under the working model assumption? The answer is yes, as given in the following Lemma 2 of [35].

Lemma 2. *Say $\mathcal{B}$ is a compact subset of $\Re^p$. Given the data $\boldsymbol{\mathcal{O}}$, there exists a sequence $\boldsymbol{\beta}_n^\star$ maximizing $l(\boldsymbol{\beta}; s^\star, t)$, that is,*

$$\boldsymbol{\beta}_n^\star = \text{argmax}_{\boldsymbol{\beta}\in\mathcal{B}} l_P(\boldsymbol{\beta}; s^\star, t).$$

Following [34], to describe the discrepancy between the sequence of estimators $\{\tilde{\boldsymbol{\beta}}_n : n = 1, 2, 3...\}$ and $\{\breve{\boldsymbol{\beta}}_n : n = 1, 2, 3...\}$, we use the Kullback–Leibler ([36]) information criterion (KLI). The KLI gives an idea about how far apart the two solutions are and is given by

$$\text{KLI}(\boldsymbol{\mathcal{O}}; \boldsymbol{\beta}) = E_{\boldsymbol{\kappa}(\mathbf{x},\boldsymbol{\epsilon})}\left[\frac{l_P(\boldsymbol{\beta}; s^\star, t)}{\tilde{l}_P(\boldsymbol{\beta}; s^\star, t)}\right].$$

The following theorem discusses the large sample behavior of the sequence $\breve{\boldsymbol{\beta}}_n$, the solution obtained using the working likelihood.

Theorem 6. *As $n \to \infty$, we have the following:*
(a) $\boldsymbol{\beta}_n^\star \xrightarrow{p} \boldsymbol{\beta}^\star \neq \boldsymbol{\beta}_0$
(b) $\sqrt{n}(\boldsymbol{\beta}_n^\star - \boldsymbol{\beta}^\star) \xrightarrow{d} N_p(\mathbf{0}, \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}^\star))$, where $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}^\star))$ is given by

$$\boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}^\star) = \left([\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}^\star; s^\star, t)]^{-1}\right)' \Psi(\boldsymbol{\beta}^\star; s^\star, t)[\boldsymbol{\mathcal{I}}(\boldsymbol{\beta}^\star; s^\star, t)]^{-1},$$

and can be consistently estimated with the observables data and expectation for the Fisher information taken under $\boldsymbol{\kappa}(\boldsymbol{x}, \boldsymbol{\epsilon})$. The true error model, $\Psi(\cdot; s^\star, t)$ is the same as in Theorem 2.

Remark 2. *The previous theorem gives us two important results when the model for errors is misspecified. First, the root of the corrected score under the misspecified model is still consistent—however, it does not converge to the true value of $\boldsymbol{\beta}$, and we still have convergence to a multivariate normal distribution when the misspecified estimator is standardized. The other very important result is the fact that we can quantify the magnitude of the inconsistency, if any, by looking at $\boldsymbol{\beta}_n^\star - \boldsymbol{\beta}_0$. To obtain $\boldsymbol{\beta}_n^\star$, one needs to numerically solve the equation*

$$\begin{aligned}
\tilde{\boldsymbol{U}}(\boldsymbol{\beta}; s^\star, t) &= \sum_{i=1}^{n} \int_0^t [\mathbf{x}_i(E_i^{-1}(w)) - E_{\boldsymbol{\kappa}(\mathbf{x},\boldsymbol{\epsilon})}(s^\star, \mathbf{x}(w))] N_i(s^\star, dw) \\
&= \sum_{i=1}^{n} \int_0^t \left[\mathbf{x}_i(E_i^{-1}(w)) - \frac{E_{\boldsymbol{\kappa}(\mathbf{x},\boldsymbol{\epsilon})}(S^{(1)}(s^\star, \mathbf{x}(w)|\boldsymbol{\mathcal{O}}))}{E_{\boldsymbol{\kappa}(\mathbf{x},\boldsymbol{\epsilon})}(S^{(0)}(s^\star, \mathbf{x}(w)|\boldsymbol{\mathcal{O}}))}\right] N_i(s^\star, dw) \\
&= \mathbf{0}.
\end{aligned} \tag{6}$$

Proof. See the Appendix A. □

7. Conclusions and Discussion

Through our exposition, we have demonstrated that, in the context of recurrent events, correcting the profile likelihood and estimating equations is a sensible approach when measurement errors in the Cox model covariates are suspected. We have further provided a mechanism for correcting the baseline integrated hazard in the presence of measurement errors, as it is a key component for inference in reliability and recurrence. We operated under the additive measurement errors model and have shown that the corrected estimators proposed are consistent and, when properly standardized, converge to Gaussian processes. We further utilized the Kullback–Leibler information criterion and misspecification theorems from [37] to show that when the errors model is misspecified by taking a form that differs from the additive, bias can be easily quantified and that the new

estimators obtained under this scenario converge when properly standardized, although not to the true parameter values. The latter observation provides an indication of the size of the bias. We provided simulation results in support of bias and estimation, and successfully applied our findings to an open-source rhDNase data. Among all else, the choice of our model, including the corrected score, is primarily guided by our desire to build a compact tool ranging from estimation to asymptotic convergence including a mechanism to investigate asymptotic bias. Our hope is that practitioners will not settle for models that ignore measurement errors, but instead will wisely question the data-generation process and consider whether bias may have influenced the results. We have provided a theoretical method for estimating the magnitude of errors, and we plan to release the current work as software, which will be published on the Comprehensive R Archive Network (CRAN) and tailored for practitioners with a more applied focus.

Author Contributions: Conceptualization, R.A., X.M.W. and A.A.; methodology, G.K.D.Z. and R.A.; software, G.K.D.Z.; validation, G.K.D.Z. and R.A.; formal analysis R.A., G.K.D.Z., X.M.W. and A.A.; writing—original draft preparation, R.A., A.A. and X.M.W.; writing—review and editing, A.A. and X.M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: We attest that all authors have no conflicts of interest or appointments with any institutions other than their respective academic institutions. We did not receive any funding for the development of this manuscript.

Appendix A. Technical Appendix

Regularity Conditions A:
I. The $\epsilon_i(t)$s are independent with mean zero and independent of $\mathcal{O}$.
II. $\mathrm{Var}(\boldsymbol{\epsilon}_i(t)) = \boldsymbol{\Xi}$ and is time-independent.
III. The moment-generating functions at all orders exist.
Assumptions I, II, and III are regular assumptions imposed on measurement errors. Assumption I is a trivial assumption in the sense that errors, in general, should not deviate much from the true value and should average to zero. In addition, errors on one unit are not an indication of errors on the next one, hence the independence between errors and observables. Assumption II is given to simplify calculation of the variance–covariance matrix of the large sample distribution of the corrected error properly standardized. As for III, it is the usual assumption on the moment-generating function. It is needed for the large sample properties of the corrected score since the asymptotic properties require the probability limit of the estimator of the moment-generating function.
Regularity Conditions B:
The Regularity Conditions I, II, III, and IV on page 6 of [28] with $\mathbf{x}_i$ replaced by $\mathbf{z}_i$ are in force, namely,
I. $\int_0^{t^\star} \lambda_0(w)dw < \infty$.
II. $(\mathbf{X}_i(s) : s \leq \tau_i)$ is uniformly bounded for $s \in [0, \tau_i]$ for all i.
III. For $k = 0, 1$, and 2 there exist functions $s^{(k)}(s^\star, t|\boldsymbol{\beta}) = E(S^{(k)}(s^\star, t|\boldsymbol{\beta}))$ and a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that the functions are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in $t \in [0, t^\star]$ and bounded on $[0, t^\star] \times \mathcal{B}$. Furthermore,

$$\sup_{\substack{t\in[0,t^\star]\\ \boldsymbol{\beta}\in\mathcal{B}}} \left|\left| S^{(k)}(s, t|\boldsymbol{\beta}) - s^{(k)}(s, t|\boldsymbol{\beta}) \right|\right| \xrightarrow{p} 0$$

and

$$s^{(1)}(s^\star, t|\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} s^{(0)}(s^\star, t|\boldsymbol{\beta}) \quad \text{and} \quad s^{(2)}(s^\star, t|\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^t} s^{(0)}(s^\star, t|\boldsymbol{\beta}).$$

IV. Let $\mathcal{B}$, $s^{(0)}(s^\star, t|\boldsymbol{\beta})$, $s^{(1)}(s^\star, t|\boldsymbol{\beta}), s^{(2)}(s^\star, t|\boldsymbol{\beta})$ be as in Condition III. For all $\boldsymbol{\beta} \in \mathcal{B}$ and $t \in [0, t^\star]$, define

$$e(s^\star, t|\boldsymbol{\beta}) = \frac{s^{(1)}(s^\star, t|\boldsymbol{\beta})}{s^{(0)}(s^\star, t|\boldsymbol{\beta})} \quad \text{and} \quad v(s^\star, t|\boldsymbol{\beta}) = \frac{s^{(2)}(s^\star, t|\boldsymbol{\beta})}{s^{(0)}(s^\star, t|\boldsymbol{\beta})} - e(s^\star, t|\boldsymbol{\beta})^{\otimes 2}.$$

Assume that $s^{(0)}(s^\star, t|\boldsymbol{\beta})$ is bounded away from 0 on $[0, t^\star] \times \mathcal{B}$; there exists a positive-definite matrix $\boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0; s^\star, t)$ such that

$$\sup_{t\in[0,t^\star]} \left\| \frac{1}{n}\sum_{i=1}^{n} \int_0^t \left[\mathbf{X}_i(\mathcal{E}_i^{-1}(w)) - E(s^\star, w|\boldsymbol{\beta}_0)\right]^{\otimes 2} S^{(0)}(s^\star, w|\boldsymbol{\beta}_0)\lambda_0(w)dw - \boldsymbol{\Sigma}_1(\boldsymbol{\beta}_0; s^\star, t) \right\|$$

converges in probability to 0; and the matrix

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}_0; s^\star, t^\star) = \int_0^{t^\star} v(s^\star, w|\boldsymbol{\beta}_0) s^{(0)}(s^\star, w|\boldsymbol{\beta}_0)\lambda_0(w)dw$$

is positive-definite. In Regularity Condition B, the $E(\cdot,\cdot)$, $V(\cdot,\cdot)$, and $S(\cdot,\cdot)$ functions are the regular functions that arise in the analysis of failure time data. The $E(\cdot,\cdot)$ and $V(\cdot,\cdot)$ are the expectation and variance, respectively, of the covariates when they are truly observed, whereas the $S^{(k)}(\cdot,\cdot)$, $k = 0, 1$, and 2 are normalizing constants.

Proof of Proposition 1. Under Regularity Condition I,

$$\begin{aligned} E_{\boldsymbol{\epsilon}}(S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\mathcal{O}})) &= E_{\boldsymbol{\epsilon}}\left(\frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\cdot \exp[\boldsymbol{\beta}'(\mathbf{z}_i(t) + \boldsymbol{\epsilon}_i(t))]\right) \\ &= E_{\boldsymbol{\epsilon}}\left(\exp(\boldsymbol{\beta}'\boldsymbol{\epsilon}_i(t))\right) \cdot E_{\boldsymbol{\epsilon}}\left(\frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\exp(\boldsymbol{\beta}'\mathbf{z}_i(t))\right) \\ &= \phi(\boldsymbol{\beta}) \cdot S^{(0)}(s^\star, \mathbf{z}(t)) + o_p(1). \end{aligned}$$

Likewise, using the fact that $\boldsymbol{\nabla}_{\boldsymbol{\beta}}\phi_i(\boldsymbol{\beta}) = E(\boldsymbol{\epsilon}_i(t)\exp(\boldsymbol{\beta}'\boldsymbol{\epsilon}_i(t))$, we obtain

$$E_{\boldsymbol{\epsilon}}(S^{(1)}(s^\star, \mathbf{x}(t)|\boldsymbol{\mathcal{O}}) = \phi(\boldsymbol{\beta}) \cdot S^{(1)}(s^\star, \mathbf{z}(t)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}}\phi(\boldsymbol{\beta}) S^{(0)}(s^\star, \mathbf{z}(t)).$$

Hence, applying a first order approximation of $E_{\boldsymbol{\epsilon}}(S^{(k)}(s^*, \mathbf{x}(t)|\boldsymbol{\mathcal{O}})$ for $k = 0, 1$ at $\left[E_{\boldsymbol{\epsilon}}(S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\mathcal{O}}), E_{\boldsymbol{\epsilon}}(S^{(1)}(s^\star, \mathbf{x}(t)|\boldsymbol{\mathcal{O}}))\right]$, we obtain

$$\begin{aligned} E_{\boldsymbol{\epsilon}}(E(s^\star, \mathbf{x}(t)|\mathcal{O})) &= \frac{E_{\boldsymbol{\epsilon}}\left(S^{(1)}(s^\star, \mathbf{x}(t))|\boldsymbol{\mathcal{O}}\right)}{E_{\boldsymbol{\epsilon}}\left(S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\mathcal{O}})\right)} \\ &= \frac{\phi(\boldsymbol{\beta}) \cdot S^{(1)}(s^\star, \mathbf{z}(t)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}}\phi(\boldsymbol{\beta}) S^{(0)}(s^\star, \mathbf{z}(t))}{\phi(\boldsymbol{\beta}) \cdot S^{(0)}(s^\star, \mathbf{z}(t))} \\ &= \frac{S^{(1)}(s^\star, \mathbf{z}(t))}{S^{(0)}(s^\star, \mathbf{z}(t))} + \boldsymbol{\nabla}_{\boldsymbol{\beta}}[\ln(\phi(\boldsymbol{\beta}))]. \end{aligned}$$

So, the corrected score to be used for unbiased estimating equations based on the observables is

$$\check{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) = \sum_{i=1}^{n} \int_0^t [\mathbf{x}_i(E_i^{-1}(w)) - E(s^\star \mathbf{x}(w)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ln(\phi(\boldsymbol{\beta}))] N_i(s^\star, dw). \tag{A1}$$

The solution $\check{\boldsymbol{\beta}}$ to the estimating equations $\check{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) = \mathbf{0}$ is the corrected maximum profile likelihood estimator. In a like manner, one can also derive the corrected information matrix as

$$\mathcal{I}_n(\check{\boldsymbol{\beta}}; s^\star, t) = -\boldsymbol{\nabla}_{\boldsymbol{\beta}}[\check{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)]\Big|_{\check{\boldsymbol{\beta}}}.$$

□

Proof of Theorem 1. The proof is based on a simplified functional uniform convergence argument for $M-$estimators as used in theorem 5.9 of [38], by showing two conditions needed to establish consistency, namely, convergence in probability (i) and separability of the root (ii). To show condition (i), observe that

$$\begin{aligned} \frac{1}{n}\check{\mathbf{U}}(\boldsymbol{\beta}; s^\star) &= \frac{1}{n}\sum_{i=1}^{n} \int_0^{s^\star} [\mathbf{x}_i(w) - E(s^\star, \varphi_i(w); \mathbf{x}(w)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ln(\phi(\boldsymbol{\beta}))] M_i(dw|\boldsymbol{\beta}_0) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \int_0^{s^\star} [\mathbf{x}_i(w) - E(s^\star, \varphi_i(w); \mathbf{x}(w)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}} \ln(\phi(\boldsymbol{\beta}))] A_i^\dagger(dw|\boldsymbol{\beta}_0) \\ &= T_1 + T_2. \end{aligned} \tag{A2}$$

We now use the time-transformed processes. Since $M_i(s^\star, dw|\boldsymbol{\beta}_0)$ is a zero-mean martingale for fixed t, transforming the first term in (A2), we see that it is $o_p(1)$. Next, replacing $\mathbf{x}_i(w)$ by $\mathbf{z}_i(w) + \boldsymbol{\epsilon}_i(w)$ and using the fact that $E(s^\star, \mathbf{x}(t)) = E(s^\star, \mathbf{z}(t)) + \boldsymbol{\nabla}_{\boldsymbol{\beta}}[\ln(\phi(\boldsymbol{\beta}))]$, the second term becomes

$$\begin{aligned} T_2 &= \frac{1}{n}\sum_{i=1}^{n} \int_0^{t^\star} [\mathbf{z}_i(\varphi_i^{-1}(w)) - E(s^\star, \mathbf{z}(w))] A_i^\dagger(s, dw|\boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n} \int_0^{t^\star} \boldsymbol{\epsilon}_i(\varphi_i^{-1}(w)) A_i^\dagger(s, dw|\boldsymbol{\beta}_0) \\ &\quad - \int_0^{s^\star} E(s^\star, \mathbf{z}(w)) S^{(0)}(s^\star, \mathbf{z}(w)) \lambda_0(w) dw. \end{aligned}$$

Using various stochastic processes techniques, it can easily be shown that

$$T_2 \xrightarrow{p} \int_0^{t^\star} [s^{(1)}(s^\star, \mathbf{z}(w)|\boldsymbol{\beta}) - e(s^\star, \mathbf{z}(w)) s^{(0)}(s^\star, \mathbf{z}(w)|\boldsymbol{\beta}_0)] \lambda_0(w) dw.$$

Hence, Condition (i) of Theorem 5.9 of [38] has been established. For Condition (ii), note that the corrected score $\check{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)$ is a continuous function and is $\mathbf{0}$ at $\boldsymbol{\beta}_0$. The partial derivative of the in-probability limit of T_2 is

$$-\int_0^{t^\star} s^{(0)}(s^\star, w|\boldsymbol{\beta}_0) \left[\frac{s^{(2)}(s^\star, w|\boldsymbol{\beta}_0)}{s^{(1)}(s^\star, w|\boldsymbol{\beta}_0)} - \left(\frac{s^{(1)}(s^\star, w|\boldsymbol{\beta})}{s^{(0)}(s^\star, w|\boldsymbol{\beta})} \right)^{\otimes 2} \right] \lambda_0(w) dw, \tag{A3}$$

which is negative definite at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Hence, $\check{\boldsymbol{\beta}}$ is a global maximum, thus concluding the consistency of $\boldsymbol{\beta}$. □

Proof of Theorem 2. $\sqrt{n}^{-1}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)$ can be written as

$$\begin{aligned}\frac{1}{\sqrt{n}}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \Big\{\mathbf{x}_i(\varphi_i^{-1}(v)) - E(s^\star, \mathbf{x}(v)) + \nabla_{\boldsymbol{\beta}}\ln[\phi(\boldsymbol{\beta}_0)]\Big\} M_i(s^\star, dv|\boldsymbol{\beta}_0) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t h_i(s^\star, v)\exp(\boldsymbol{\beta}_0'\mathbf{x}_i(v)) e^{-\boldsymbol{\epsilon}_i'(v)\boldsymbol{\beta}_0}\lambda_0(v)\phi(\boldsymbol{\beta}_0)\phi^{-1}(\boldsymbol{\beta}_0) e^{-\boldsymbol{\epsilon}_i(v)\boldsymbol{\beta}_0} dv.\end{aligned}$$

For large n, it can be seen that $\frac{1}{\sqrt{n}}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t)$ is equal to

$$\begin{aligned}\frac{1}{\sqrt{n}}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \Big\{\mathbf{x}_i(\varphi_i^{-1}(v)) - E(s^\star, \mathbf{x}(v)) + \nabla_{\boldsymbol{\beta}}[\ln\phi(\boldsymbol{\beta}_0)]\Big\} M_i(s^\star, dv|\boldsymbol{\beta}_0) \\ &\quad + \frac{1}{\sqrt{n}}\phi^{-1}(\boldsymbol{\beta}_0)\sum_{i=1}^{n}\int_0^t h_i(s^\star, v)\exp(\boldsymbol{\beta}_0'\mathbf{x}_i(v)) e^{-\boldsymbol{\epsilon}'(v)\boldsymbol{\beta}_0}\lambda_0(v) dv.\end{aligned} \tag{A4}$$

The display in (A4) can be viewed as the sum of n independent $V_i(s,t)$ processes with $V_i(s,t)$ given by

$$\begin{aligned}V_i(s,t) &= \frac{1}{\sqrt{n}}\int_0^t \Big\{\mathbf{x}_i(\varphi^{-1}(v)) - E(s, \mathbf{z}(v)) + \nabla_{\boldsymbol{\beta}}[\ln\phi(\boldsymbol{\beta}_0)]\Big\} M_i(s, dv|\boldsymbol{\beta}_0) \\ &\quad + \phi^{-1}(\boldsymbol{\beta}_0)\int_0^t h_i(s, v)\exp(\boldsymbol{\beta}_0'\mathbf{x}_i(v)) e^{-\boldsymbol{\epsilon}'(v)\boldsymbol{\beta}_0}\lambda_0(v) dv.\end{aligned}$$

We can then apply the functional central limit theorem to find its large sample properties. To apply the functional central limit theorem, we have to show conditions (a) to (e) of Theorem 11.16 of [39] pertaining to the manageability of the process and the existence of an in-probability limit of the variance covariance matrix of the standardized process. These conditions have been shown in numerous manuscripts dealing with recurrent event modeling when the covariates are not mismeasured. The addition of the constant term $\nabla_{\boldsymbol{\beta}}[\ln\phi(\boldsymbol{\beta}_0)]$ does not prevent those conditions from holding since $M_i(s, dv)$ is a zero-mean martingale. Therefore, by Theorem 11.16 of [39], the process

$$\Big\{\sqrt{n}^{-1}\breve{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) : t \in [0, t^\star]\Big\}$$

converges to a tight $\mathbf{U}^\infty(\boldsymbol{\beta}_0; s^\star, t)$ with a variance covariance matrix $\Psi(s, t_1 \wedge t_2|\boldsymbol{\beta}_0))$ that can be estimated by $\breve{\Psi}(s, t_1 \wedge t_2|\breve{\boldsymbol{\beta}}))$. □

Proof of Proposition 2. Replacing $\mathbf{x}$ by $\mathbf{z} + \boldsymbol{\epsilon}$, $S^{(2)}(s, \mathbf{x}(t)|\boldsymbol{\beta})$ can be written as

$$\begin{aligned}S^{(2)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta}) &= \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)(\mathbf{z}_i + \boldsymbol{\epsilon}_i)^{\otimes 2}\exp(\boldsymbol{\beta}'(\mathbf{z}_i + \boldsymbol{\epsilon}_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\mathbf{z}_i^{\otimes 2}\exp(\boldsymbol{\beta}'(\mathbf{z}_i + \boldsymbol{\epsilon}_i)) + \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\boldsymbol{\epsilon}_i^{\otimes 2}\exp(\boldsymbol{\beta}'(\mathbf{z}_i + \boldsymbol{\epsilon}_i)) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\mathbf{z}_i\boldsymbol{\epsilon}_i'\exp(\boldsymbol{\beta}'(\mathbf{z}_i + \boldsymbol{\epsilon}_i)) + \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\mathbf{z}_i'\boldsymbol{\epsilon}_i\exp(\boldsymbol{\beta}'(\mathbf{z}_i + \boldsymbol{\epsilon}_i)) \\ &= A_1 + A_2 + A_3 + A_4.\end{aligned}$$

A_1 can be written as

$$A_1 = \phi(\beta)\frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\mathbf{z}_i^{\otimes 2}\exp(\boldsymbol{\beta}'\mathbf{z}_i) + \frac{1}{n}\sum_{i=1}^{n} h_i(s,t)\mathbf{z}_i^{\otimes 2}\exp(\boldsymbol{\beta}'\mathbf{z}_i)\{\exp(\boldsymbol{\beta}'\boldsymbol{\epsilon}_i)\phi^{-1}(\boldsymbol{\beta}) - 1\}.$$

So that

$$E(A_1) = \phi(\boldsymbol{\beta}) s^{(2)}(s^\star, \mathbf{z}) + o_p(1).$$

Likewise,

$$\begin{aligned} A_2 &= \phi(\boldsymbol{\beta}) \frac{1}{n} \sum_{i=1}^{n} h_i(s,t) \exp(\boldsymbol{\beta}' \mathbf{z}_i) [\boldsymbol{\Xi} + (\boldsymbol{\Xi}\boldsymbol{\beta})(\boldsymbol{\Xi}\boldsymbol{\beta})'] \\ &\quad + \phi(\boldsymbol{\beta}) \frac{1}{n} \sum_{i=1}^{n} h_i(s,t) \exp(\boldsymbol{\beta}' \mathbf{z}_i) \left[\sum_{i=1}^{n} \boldsymbol{\epsilon}_i^{\otimes 2} \right] \\ &\quad + \frac{1}{n} \left[\left[\sum_{i=1}^{n} \boldsymbol{\epsilon}_i^{\otimes 2} \right] \boldsymbol{\beta} \right] \left[\sum_{i=1}^{n} \boldsymbol{\epsilon}_i^{\otimes 2} \right]' \left[e^{\boldsymbol{\beta}' \boldsymbol{\epsilon}_i} \phi(\boldsymbol{\beta}) - 1 \right]. \end{aligned}$$

Hence,

$$E(A_2) = \phi(\boldsymbol{\beta}) S^{(0)}(s^\star, \mathbf{z}(t)) [\boldsymbol{\Xi} + (\boldsymbol{\Xi}\boldsymbol{\beta})(\boldsymbol{\Xi}\boldsymbol{\beta})'] + o_p(1).$$

The terms A_3 and A_4 can also be corrected in the same way, leading to

$$E(A_3) = \phi(\boldsymbol{\beta}) [S^{(1)}(s^\star, \mathbf{z}(t))(\boldsymbol{\Xi}\boldsymbol{\beta})' + (\boldsymbol{\Xi}\boldsymbol{\beta}) s^{(1)}(s^\star, \mathbf{z}(t))'] + o_p(1).$$

Because of symmetry between A_3 and A_4, the expressions are the same, and the corrected expression of $S^{(2)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})[S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})]^{-1}$ follows. □

Proof of Theorem 4. This can be proved using the concept of statistical functionals. Define the Mallow metric $d_1(\cdot,\cdot)$ on L_1. Let X be an observable with probability measure P for which $E(X) < \infty$, likewise for Y with probability measure Q. Define the metric $d_1(\cdot,\cdot)$ between P and Q by $d_1(P,Q) = \inf E(|X - Y|)$. By Lemma 8.1 of [40], $d_1(\cdot,\cdot)$ is a metric. The empirical distribution function $F_n(w|\mathbf{x})$ converges in distribution to the true distribution function of the errors $F(t)$. By Lemma 8.3 of [40], $d_1(F_n(t), F(t)) \to 0$ as $n \to \infty$. Given $K_i = l_i$, define the functionals

$$\int \mathbf{x} d\bar{F}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \mathbf{x}_i(j), \quad \text{and} \quad \int \mathbf{x}^2 d\bar{F}_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K_i} \mathbf{x}_i^2(j).$$

Assume F_n and F are in L_1 and L_2, then, by [40], $d_1(F_n, F) \to 0$ implies $\int \|\mathbf{x}\| d\bar{F}_n \to E(\|\mathbf{x}\|)$ and $\int \|\mathbf{x}\|^2 d\bar{F}_n \to E(\|\mathbf{x}\|^2)$. Then, $\hat{\sigma}_{qr}^2 \xrightarrow{as} \sigma_{qr}^2$ in bootstrap probability, and consequently, $\hat{\boldsymbol{\Xi}} \xrightarrow{as} \boldsymbol{\Xi}$. □

Proof of Theorem 5. Write

$$\begin{aligned} \sqrt{n}(\check{\Lambda}_0(s^\star, t|\check{\boldsymbol{\beta}}) - \Lambda_0(t)) &= \sqrt{n}(\check{\Lambda}_0(s^\star, t|\check{\boldsymbol{\beta}}) - \check{\Lambda}_0(s^\star, t|\boldsymbol{\beta}_0)) \\ &\quad + \sqrt{n}(\check{\Lambda}_0(s^\star, t|\boldsymbol{\beta}_0) - \Lambda_0(t)) \\ &= C_1 + C_2. \end{aligned}$$

Consider C_1. Expanding C_1 using a Taylor series yields

$$\begin{aligned} C_1 &= (\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \left[\nabla_\beta \phi(\boldsymbol{\beta}^\star) \int_0^t \frac{\sum_{i=1}^{n} N_i(s^\star, du)}{S^{(0)}(s^\star, u|\boldsymbol{\beta}^\star)} \right] \\ &\quad - (\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \nabla_\beta \left[\phi(\boldsymbol{\beta}^\star) \int_0^t \frac{E(\boldsymbol{\beta}^\star, \mathbf{x}(u))}{S^{(0)}(s^\star, \mathbf{x}(u)|\boldsymbol{\beta}^\star)} \sum_{i=1}^{n} N_i(s^\star, du) \right] \\ &= -(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) H(\boldsymbol{\beta}^\star), \end{aligned}$$

where $\boldsymbol{\beta}^\star$ lies in the line segment between $\breve{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}^\star \xrightarrow{p} \boldsymbol{\beta}_0$ as $n \to \infty$. Next, using the corrected expression of $S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$ and $S^{(1)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta})$), it can be shown that, as $n \to \infty$, $H(\boldsymbol{\beta}^\star)$ converges to $-\int_0^t e(\boldsymbol{\beta}_0, u)\lambda_0(u)du$. Noting that $S^{(0)}(s^\star, \mathbf{x}(t)|\boldsymbol{\beta}_0) = \phi(\boldsymbol{\beta}_0)S^{(0)}(s^\star, \mathbf{z}(t)|\boldsymbol{\beta}_0)$, we have

$$\begin{aligned} C_2 &= \breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}}) - \Lambda_0(t) \\ &= \int_0^t \frac{dM(s, du)}{S^{(0)}(s^\star, \mathbf{z}(u))} + o_p(1). \end{aligned}$$

Hence, for large n

$$\begin{aligned} \sqrt{n}(\breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}}) - \Lambda_0(t)) &= \sqrt{n}H(\boldsymbol{\beta}^\star)(\breve{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &\quad + \sqrt{n}\int_0^t \frac{\sum_{i=1}^n M_i(s, du)}{S^{(0)}(s^\star, \mathbf{z}(u))} \\ &= \frac{1}{\sqrt{n}}\int_0^t \frac{dM(s, du)}{S^{(0)}(s^\star, \mathbf{z}(u))} \\ &\quad + \frac{1}{\sqrt{n}}H(\boldsymbol{\beta}_0)\left[\frac{1}{n}I_n(\boldsymbol{\beta}_0; s^\star, t)\right]^{-1}\mathbf{U}_n(\boldsymbol{\beta}_0, s^\star, \mathbf{x}(t)) \\ &= \sum_{i=1}^n W_i(\boldsymbol{\beta}_0; s^\star, t). \end{aligned}$$

Therefore, $\sqrt{n}(\breve{\Lambda}_0(s^\star, t|\breve{\boldsymbol{\beta}}) - \Lambda_0(t))$ can be viewed as the sum of n independent components, each of which is a p-vector. This is the sum of n independent random processes whose finite dimensional distribution converges to those of $W^\infty(s, t)$. By the functional central limit theorem, the sum converges weakly to a zero-mean Gaussian process with a covariance matrix $\Gamma(s, t_1 \wedge t_2)$ that can be estimated consistently by $\breve{\Gamma}(s, t_1 \wedge t_2) = E(W_i(s, t_1; \breve{\boldsymbol{\beta}})W_i(s, t_2; \breve{\boldsymbol{\beta}}))$. □

Proof of Theorem 6. (a) We apply Theorem 1.14 on page 28 of [17]. We have two tasks: (i) show that $l(\boldsymbol{\beta}; s^\star, t) = \ln[L(\boldsymbol{\beta}; s^\star, t)]$ is bounded by an integrable function with respect to the true errors distribution and (ii) KLI$(\boldsymbol{\mathcal{O}}; \boldsymbol{\beta})$ has a unique minimum at $\boldsymbol{\beta}_n^\star$ in $\mathcal{B}$, a compact subset of $\Re^p$. We begin with (i). Observe that the corrected score under the true errors model is

$$\tilde{\mathbf{U}}(\boldsymbol{\beta}; s^\star, t) := \sum_{i=1}^n \int_0^t \left[\mathbf{x}_i(\varphi_i^{-1}(w)) - \frac{E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})}(S^{(1)}(s^\star, w; \mathbf{x}|\boldsymbol{\mathcal{O}}))}{E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})}(S^{(0)}(s^\star, w; \mathbf{x}|\boldsymbol{\mathcal{O}}))}\right] N_i(s^\star, dw).$$

The concept of manageability can be used to show that it is bounded by an integrable function. We assume without loss of generality that $\mathbf{x}_{ik}$ is positive. Because x_{ik} has a total variation bounded by a constant, and $N_i(s, t)$ is a sum of indicator functions that are increasing functions for each i, we conclude that each of the p-components of $\tilde{l}_P(\boldsymbol{\beta}; s^\star, t)$ has a pseudodimension not exceeding 10 (cf. [41]) since the integrand has a pseudodimension of, at most, 1. Therefore, $\tilde{l}_P(\boldsymbol{\beta}; s^\star, t)$ is manageable, and hence, is bounded by an integrable function with respect to the true probability measure of the errors that also guarantees that KLI$(\boldsymbol{\mathcal{O}}; \boldsymbol{\beta})$ is well defined. To prove part (ii), let $l(\boldsymbol{\beta}; s^\star, t)$ be the true likelihood under the true errors model, and $\tilde{l}(\boldsymbol{\beta}; s^\star, t)$ be the working likelihood, that is, the one under the misspecified model. Define the random function $\mathcal{D}(\boldsymbol{\beta}; s^\star, t)$ in $\mathcal{B}$ by

$$\begin{aligned} \mathcal{D}(\boldsymbol{\beta}; s^\star, t) &= \frac{1}{n}E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})}[l(\boldsymbol{\beta}_0; s^\star, t) - \tilde{l}(\boldsymbol{\beta}; s^\star, t)] &\text{(A5)} \\ &= \frac{1}{n}E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})}\left[\log\left(\frac{L(\boldsymbol{\beta}_0; s^\star, t)}{\tilde{L}(\boldsymbol{\beta}; s^\star, t)}\right)\right]. &\text{(A6)} \end{aligned}$$

Also, define the deterministic function

$$d(\boldsymbol{\beta}; s^{\star}, t) = \int_0^{t^{\star}} \left[(\tilde{\boldsymbol{\beta}}_0 - \boldsymbol{\beta})' \mathbf{x}(\varphi^{-1}(w)) - \ln \left[\frac{s^{(0)}(s^{\star}, w|\boldsymbol{\beta})}{s^{(0)}(s^{\star}, w|\boldsymbol{\beta}_0)} \right] \right] \lambda_0(w) dw. \tag{A7}$$

Both of these functions are twice-differentiable with respect to $\boldsymbol{\beta} \in \mathcal{B}$. Observe that $\boldsymbol{\beta}_n^{\star}$ is the natural estimator of the working likelihood. Furthermore, $\tilde{\boldsymbol{\beta}}$ is the parameter value that minimizes KLI$(\boldsymbol{\mathcal{O}}; \boldsymbol{\beta})$ since the numerator in KLI is a function of the true likelihood and the denominator is the working likelihood, with the expectation taken with respect to the true measurement errors model distribution. Assume that, for $m = 0, 1$, and 2,

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \| S_i^{(m)}(\boldsymbol{\beta}; s, t) - E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})} \left(S_i^{(m)}(\boldsymbol{\beta}; s, t) \right) \| \xrightarrow{p} 0,$$

with $E_{\boldsymbol{\kappa}(\mathbf{x}, \boldsymbol{\epsilon})} \left(S_i^{(m)}(\boldsymbol{\beta}; s, t) \right) = s^{\star(m)}(\boldsymbol{\beta}; s, t)$. Furthermore, observe that $\tilde{\boldsymbol{\beta}}_n$ is minimizer of $\boldsymbol{\beta} \mapsto D(s^{\star}, \boldsymbol{\beta})$. We also observe that

$$\sup_{(\boldsymbol{\beta}, t) \in \mathcal{B} \times [0, \tau]} \| \mathcal{D}(\boldsymbol{\beta}; s^{\star}, t) - d(\boldsymbol{\beta}; s^{\star}, t) \| \xrightarrow{p} 0,$$

as $n \to \infty$. Easy calculations show that the first derivative of $\mathcal{D}(\boldsymbol{\beta}; s^{\star}, t)$ is the score process, whose derivative is the Fisher information matrix with limit coinciding with the Hessian of $d(\boldsymbol{\beta}; s^{\star}, t)$. The Hessian of $d(\boldsymbol{\beta}; s^{\star}, t)$ turns out to be positive definite, which means that $\boldsymbol{\beta} \mapsto D(\boldsymbol{\beta}; s^{\star}, t)$ is strictly convex for $\boldsymbol{\beta} \in \mathcal{B}$. We take $\boldsymbol{\beta}_n^{\star}$ to be the minimizer of $D(s^{\star}, \boldsymbol{\beta})$ inside $\mathcal{B}$, and $\boldsymbol{\beta}^{\star}$ to be the unique minimizer of $d(\boldsymbol{\beta}; s^{\star}, t)$ in $\mathcal{B}$. Therefore, by Theorem 2.2 of [34], we obtain the convergence in probability of the sequence $\{\boldsymbol{\beta}_n^{\star} : n = 1, 2, 3, ...\}$ to a value $\boldsymbol{\beta}^{\star}$ minimizing the KLI$(\boldsymbol{\mathcal{O}}; \boldsymbol{\beta})$. The proof of Part (b) follows along the lines of the convergence to multivariate normal distribution of the properly standardized corrected maximum likelihood when the errors terms are correctly specified. □

References

1. Prentice, R.L. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* **1982**, *69*, 331–342. [CrossRef]
2. Stefanski, L.A. The effects of measurement error on parameter estimation. *Biometrika* **1985**, *72*, 583–592. [CrossRef]
3. Armstrong, B.G. The Effects of Measurement Errors on Relatwe Risk Regressions. *Am. J. Epidemiol.* **1990**, *132*, 1176–1184. [CrossRef]
4. Nakamura, T. Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **1990**, *77*, 127–137. [CrossRef]
5. Nakamura, T. Proportional hazards model with covariates subject to measurement error. *Biometrics* **1992**, *48*, 829–838. [CrossRef]
6. Hu, P.; Tsiatis, A.A.; Davidian, M. Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* **1998**, *54*, 1407–1419. [CrossRef] [PubMed]
7. Hu, C.; Lin, D.Y. Cox regression with covariate measurement error. *Scand. J. Statist.* **2002**, *29*, 637–655. [CrossRef]
8. Song, X.; Huang, Y. On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* **2005**, *61*, 702–714. [CrossRef]
9. Yan, Y.; Yi, G.Y. A corrected profile likelihood method for survival data with covariate measurement error under the Cox model. *Canad. J. Statist.* **2015**, *43*, 454–480. [CrossRef]
10. Huque, M.H.; Bondell, H.D.; Carroll, R.J.; Ryan, L.M. Spatial regression with covariate measurement error: A semiparametric approach. *Biometrics* **2016**, *72*, 678–686. [CrossRef]
11. Alexeeff, S.E.; Carroll, R.J.; Coull, B. Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. *Biostatistics* **2016**, *17*, 377–389. [CrossRef] [PubMed]
12. Carroll, R.J.; Ruppert, D.; Stefanski, L.A. *Measurement Error in Nonlinear Models*; Monographs on Statistics and Applied Probability; Chapman & Hall: London, UK, 1995; Volume 63, p. xxiv+305.
13. Fuller, W.A. *Measurement Error Models*; Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1987; p. xxiv+440.

14. Yi, G.Y. *Statistical Analysis with Measurement Error or Misclassification*; Strategy, method and application, With a foreword by Raymond J. Carroll.; Springer Series in Statistics; Springer: New York, NY, USA, 2017; p. xxvii+479.
15. Turnbull, B.W.; Jiang, W.; Clark, L.C. Regression models for recurrent event data: Parametric random effects models with measurement error. *Stat. Med.* **1997**, *16*, 853–864. [CrossRef]
16. Hu, C.; Lin, D.Y. Semiparametric failure time regression with replicates of mismeasured covariates. *J. Am. Statist. Assoc.* **2004**, *99*, 105–118. [CrossRef]
17. Yi, G.Y.; Lawless, J.F. Likelihood-based and marginal inference methods for recurrent event data with covariate measurement error. *Canad. J. Statist.* **2012**, *40*, 530–549. [CrossRef]
18. Stefanski, L.A.; Cook, J.R. Simulation-extrapolation: The measurement error jackknife. *J. Am. Statist. Assoc.* **1995**, *90*, 1247–1256. [CrossRef]
19. Yu, H.; Cheng, Y.J.; Wang, C.Y. Semiparametric regression estimation for recurrent event data with errors in covariates under informative censoring. *Int. J. Biostat.* **2016**, *12*, 20160001. [CrossRef] [PubMed]
20. Yu, H.; Cheng, Y.J.; Wang, C.Y. Methods for multivariate recurrent event data with measurement error and informative censoring. *Biometrics* **2018**, *74*, 966–976. [CrossRef]
21. Chen, L.P.; Yi, G.Y. Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error. *Ann. Inst. Statist. Math.* **2021**, *73*, 481–517. [CrossRef]
22. Augustin, T. An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scand. J. Statist.* **2004**, *31*, 43–50. [CrossRef]
23. Kong, F.H.; Gu, M. Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statist. Sinica* **1999**, *9*, 953–969.
24. Wang, C.Y.; Hsu, L.; Feng, Z.D.; Prentice, R.L. Regression calibration in failure time regression. *Biometrics* **1997**, *53*, 131–145. [CrossRef]
25. Yu, M.; Nan, B. Regression calibration in semiparametric accelerated failure time models. *Biometrics* **2010**, *66*, 405–414. [CrossRef]
26. Cox, D. Regression models and life tables (with discussion). *J. R. Stat. Soc. B* **1972**, *34*, 187–220. [CrossRef]
27. Peña, E.A.; Strawderman, R.L.; Hollander, M. Nonparametric estimation with recurrent event data. *J. Am. Statist. Assoc.* **2001**, *96*, 1299–1315. [CrossRef]
28. Adekpedjou, A.; Stocker, R. A general class of semiparametric models for recurrent event data. *J. Stat. Plann. Inference* **2015**, *156*, 48–63. [CrossRef]
29. Hjort, N.L. *Bootstrapping Cox's Regression Model*; Report Number: LCS-21; Laboratory for Computational Statistics: Genova, Italy, 1985.
30. Efron, B. Censored data and the bootstrap. *J. Am. Stat. Assoc.* **1981**, *76*, 312–319. [CrossRef]
31. Efron, B.; Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.* **1986**, *1*, 54–77. [CrossRef]
32. Cook, R.J.; Lawless, J. *The Statistical Analysis of Recurrent Events (Statistics for Biology and Health)*; Statistics for Biology and Health; Springer: Berlin/Heidelberg, Germany, 2007.
33. Peña, E.A.; Slate, E.H.; Gonzalez, J.R. Semiparametric inference for a General Class of Models for Recurrent Events. *J. Stat. Plan. Inference* **2007**, *137*, 1727–1747. [CrossRef] [PubMed]
34. White, H.; Domowitz, I. Nonlinear Regression with Dependent Observations. *Econometrica* **1984**, *52*, pp. 143–162. [CrossRef]
35. Jennrich, R.I. Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **1969**, *40*, 633–643. [CrossRef]
36. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]
37. Domowitz, I.; White, H. Misspecified models with dependent observations. *J. Econom.* **1982**, *20*, 35–58. [CrossRef]
38. van der Vaart, A.W. *Asymptotic Statistics*; Cambridge Series in Statistical and Probabilistic Mathematics; Cambridge University Press: Cambridge, UK, 1998; Volume 3, p. xvi+443.
39. Kosorok, M.R. *Introduction to Empirical Processes and Semiparametric Inference*; Springer Series in Statistics; Springer: New York, NY, USA, 2008; p. xiv+483. [CrossRef]
40. Bickel, P.J.; Freedman, D.A. Some asymptotic theory for the bootstrap. *Ann. Stat.* **1981**, *9*, 1196–1217. [CrossRef]
41. Pollard, D. *Empirical Processes: Theory and Applications*; NSF-CBMS Regional Conference Series in Probability and Statistics; Institute of Mathematical Statistics: Hayward, CA, USA; American Statistical Association: Alexandria, VA, USA, 1990; Volume 2, p. viii+86.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.