

# Estimation in Cox proportional hazards model with heteroscedastic errors in covariates

Oksana Chernova<sup>a,b,\*</sup>

<sup>a</sup>*Taras Shevchenko National University of Kyiv, Ukraine*

<sup>b</sup>*Technical University of Munich, Germany*

[oksanachernova@knu.ua](mailto:oksanachernova@knu.ua) (O. Chernova)

Received: 1 March 2024, Revised: 13 May 2024, Accepted: 14 May 2024,  
Published online: 30 May 2024

**Abstract** Consistent estimators of the baseline hazard rate and the regression parameter are constructed in the Cox proportional hazards model with heteroscedastic measurement errors, assuming that the baseline hazard function belongs to a certain class of functions with bounded Lipschitz constants.

**Keywords** Cox proportional hazards model, right censoring, heteroscedastic measurement errors

**2010 MSC** 62H12, 62N02

## 1 Introduction

Survival analysis is a set of statistical methods for analysis of data representing times to the occurrence of some specified event. It is an important part of mathematical statistics due to a wide range of applications: medicine, reliability theory, etc.

The Cox proportional hazards (CPH) model is a semi-parametric regression model that is used to study the association between the survival time of subjects (the so-called lifetime) and one or more predictor variables (the so-called covariates). One of the main features of survival analysis is the presence of incomplete observations or censoring. In such cases only partial information about the true lifetime is available,

---

\*Corresponding author.

e.g., that it exceeds some value in case of right censoring. The primary quantities of interest to estimate are the hazard function and the regression parameter. The hazard function represents the instantaneous rate at which events occur at a particular time, given that the individual has survived up to that time. The vector of regression parameters represents effects of covariates on the hazard.

D. R. Cox [4] estimates the regression parameter using partial likelihood without additional assumptions on the baseline hazard rate. Estimates of cumulative hazard are proposed by [9] and [3]. In these papers, it is assumed that the baseline hazard rate belongs to some parametric family (piece-wise constant on certain intervals). Additionally, models without errors in variables are considered.

Nonparametric inference under shape constraints has been an actively researched field in recent decades. It is a framework where the estimated parameters or functions are constrained to satisfy certain shape properties such as monotonicity, convexity or log-concavity. The development of nonparametric methods for estimation of monotone density started with pioneering paper of Grenander [7]. The 2018 special issue of *Statistical Science* was devoted to inference under shape constraints. A review of recent progress in log-concave density estimation is given in [16]. A review of methods for shape constrained baseline hazard function in case of censored data is presented in [8]. A monotone baseline hazard rate in Cox model is considered in [14] and [5]. A different variation of shape constrained Cox model with application to breast cancer patients' survival is considered in [15].

One should be cautious to make inference using regression when covariates could be measured with errors. It is known that applying of naive methodology may lead to inconsistent estimation, see Wallace [17]. CPH model with measurement errors is studied, among others, by Kong and Gu [10] and Augustin [1]. Typically one estimates at first the vector of regression parameters, and then an estimator of the *cumulative* baseline hazard rate is constructed.

In Kukush et al. [11] the baseline hazard rate belongs to a bounded set of nonnegative Lipschitz functions, and is estimated simultaneously with the vector of regression parameters. This approach is further developed in Kukush and Chernova [12], where the baseline hazard rate belongs to an unbounded set of nonnegative Lipschitz functions.

In all aforementioned papers, measurement errors are assumed to be independent and identically distributed. In practice, measurement errors can be expected to vary considerably among different subjects. E.g., Augustin et al. [2] propose the regression calibration estimation method in CPH model under heteroscedastic measurement errors for nutritional data.

In the present paper we consider a CPH model similar to [11] and [12], but with *heteroscedastic* measurement errors. The paper is organized as follows. Section 2 describes the observation model and gives main assumptions. In Section 3 we construct a simultaneous consistent estimator  $(\hat{\lambda}_n, \hat{\beta}_n)$  of the baseline hazard rate and the regression parameter in CPH model with heteroscedastic errors in covariates under *bounded* parameter set. In Section 4 we do the same for *unbounded* parameter set, and Section 5 concludes.

## 2 Model description

The Cox proportional hazards model introduced in [4] assumes that the lifetime  $T$  has a hazard rate at moment  $t$  for a subject with random vector of covariates  $\mathbf{X}$  as follows:

$$\lambda(t|\mathbf{X}; \lambda_0, \beta_0) = \lambda_0(t) \exp(\beta_0^\top \mathbf{X}), \quad t \geq 0.$$

Here,  $\beta_0$  is a regression parameter belonging to  $\Theta_\beta \subset \mathbb{R}^k$ , and  $\lambda_0(\cdot) \in \Theta_\lambda \subset C[0, \tau]$  is a baseline hazard function. The unbounded parameter set  $\Theta_\lambda$  consists of all non-negative functions with bounded Lipschitz constant. Instead of the lifetime  $T$ , right-censored data  $Y := \min\{T, C\}$  and  $\Delta := I_{\{T \leq C\}}$  are available. The censor  $C$  has unknown distribution concentrated on a given interval  $[0, \tau]$ . A pair  $(\mathbf{X}, T)$  and random variable  $C$  are independent.

Assume that instead of true covariates, one can only observe surrogate vector variables

$$\mathbf{W}_i = \mathbf{X}_i + \mathbf{U}_i,$$

where (not necessarily identically distributed) measurement errors  $\mathbf{U}_i, i = 1, 2, \dots, n$ , are mutually independent *centered* random vectors that are also independent of the random sequence  $(\mathbf{X}_i, T_i, C_i, Y_i, \Delta_i), i = 1, 2, \dots, n$ . The moment generating functions  $M_{\mathbf{U}_i}(z) := \mathbf{E} e^{z^\top \mathbf{U}_i}$  of the random measurement errors  $\mathbf{U}_i$  are assumed known. The goal is to estimate  $\beta$  and  $\lambda$  based on observations  $(Y_i, \Delta_i, \mathbf{W}_i), i = 1, \dots, n$ .

Introduce the following assumptions.

- (i)  $\Theta_\lambda := \{ f : [0, \tau] \rightarrow \mathbb{R} \mid f(t) \geq a, \forall t \in [0, \tau], f(0) \leq A, \text{ and } |f(t) - f(s)| \leq L|t - s|, \forall t, s \in [0, \tau] \}$ , where  $a, A$  and  $L$  are fixed positive constants, with  $a < A$ .
- (i')  $\Theta_\lambda := \{ f : [0, \tau] \rightarrow \mathbb{R} \mid f(t) \geq 0, \forall t \in [0, \tau], \text{ and } |f(t) - f(s)| \leq L|t - s|, \forall t, s \in [0, \tau] \}$ , where  $L$  is a fixed positive constant.
- (ii)  $\Theta_\beta$  is a compact set in  $\mathbb{R}^k$ .
- (iii) There exist positive  $K$  and  $\epsilon$  such that for all  $n \geq 1$ ,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{E} e^{2D\|\mathbf{U}_i\|} \leq K, \text{ with } D := \max_{\beta \in \Theta_\beta} \|\beta\| + \epsilon.$$

- (iv)  $\mathbf{E} e^{2D\|\mathbf{X}\|} < \infty$ , with  $D$  defined in (iii).
- (v)  $\tau > 0$  is the right endpoint of censor's distribution, i.e.  $\mathbf{P}(C > \tau) = 0$  and for all  $\epsilon > 0, \mathbf{P}(C > \tau - \epsilon) > 0$ .
- (vi) The matrix  $\mathbf{E} \mathbf{X}\mathbf{X}^\top$  of second moments is positive definite.

Likelihood construction in presence of censored data is described in [13]. In case where covariates are observed without errors, the log-likelihood function is given by

$$Q_n(\lambda, \boldsymbol{\beta}) := \frac{1}{n} \sum_{i=1}^n q(Y_i, \Delta_i, \mathbf{X}_i; \lambda, \boldsymbol{\beta}), \quad \text{with}$$

$$q(Y_i, \Delta_i, \mathbf{X}_i; \lambda, \boldsymbol{\beta}) := \Delta_i \cdot (\log \lambda(Y_i) + \boldsymbol{\beta}^\top \mathbf{X}_i) - \exp(\boldsymbol{\beta}^\top \mathbf{X}_i) \int_0^{Y_i} \lambda(u) du.$$

T. Augustin [1] proposed the following objective function to adjust for homoscedastic measurement errors

$$Q_n^{cor}(\lambda, \boldsymbol{\beta}) := \frac{1}{n} \sum_{i=1}^n q^{cor}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \boldsymbol{\beta}),$$

with

$$q^{cor}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \boldsymbol{\beta}) := \Delta_i \cdot (\log \lambda(Y_i) + \boldsymbol{\beta}^\top \mathbf{W}_i) - \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du.$$

We have

$$\mathbb{E}[q^{cor}(Y_i, \Delta_i, \mathbf{W}_i) \mid Y_i, \Delta_i, \mathbf{X}_i] = q(Y_i, \Delta_i, \mathbf{X}_i).$$

Therefore,

$$\mathbb{E} q^{cor}(Y_i, \Delta_i, \mathbf{W}_i) = \mathbb{E} q(Y_i, \Delta_i, \mathbf{X}_i) = \mathbb{E} q(Y_1, \Delta_1, \mathbf{X}_1).$$

The latter equality means that  $\mathbb{E}[q^{cor}(Y_i, \Delta_i, \mathbf{W}_i)]$  does not depend on  $i$ .

Denote

$$q_\infty(\lambda, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} q^{cor}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \boldsymbol{\beta}) = \mathbb{E} q^{cor}(Y_1, \Delta_1, \mathbf{W}_1).$$

We will make use of both forms of  $q_\infty$ .

We define a simultaneous estimator of the baseline hazard rate and regression parameter under bounded and unbounded parameter sets as follows.

**Definition 1** (under bounded parameter set). A Borel function  $(\hat{\lambda}, \hat{\boldsymbol{\beta}}) = (\hat{\lambda}_n, \hat{\boldsymbol{\beta}}_n)$  of observations  $(Y_i, \Delta_i, \mathbf{W}_i), i = 1, \dots, n$ , with values in  $\Theta = \Theta_\beta \times \Theta_\lambda$ , where  $\Theta_\lambda$  is bounded, and such that

$$(\hat{\lambda}, \hat{\boldsymbol{\beta}}) = \arg \max_{(\lambda, \boldsymbol{\beta}) \in \Theta} Q_n^{cor}(\lambda, \boldsymbol{\beta}), \tag{1}$$

is called a simultaneous estimator of the baseline hazard rate and the regression parameter under the bounded parameter set  $\Theta$ .

**Definition 2** (under unbounded parameter set). Let  $\{\varepsilon_n\}$  be a fixed sequence of positive numbers such that  $\varepsilon_n \downarrow 0$  as  $n \rightarrow \infty$ . A Borel function  $(\hat{\lambda}, \hat{\beta}) = (\hat{\lambda}_n, \hat{\beta}_n)$  of observations  $(Y_i, \Delta_i, \mathbf{W}_i)$ ,  $i = 1, \dots, n$ , with values in  $\Theta = \Theta_\beta \times \Theta_\lambda$ , where  $\Theta_\lambda$  is unbounded, and such that

$$Q_n^{cor}(\hat{\lambda}, \hat{\beta}) \geq \sup_{(\lambda, \beta) \in \Theta} Q_n^{cor}(\lambda, \beta) - \varepsilon_n, \tag{2}$$

is called a simultaneous estimator of the baseline hazard rate and the regression parameter over the unbounded parameter set  $\Theta$ .

### 3 Simultaneous estimation under bounded parameter set

We extend the result about consistency of a simultaneous estimator of regression parameter and baseline hazard under bounded parameter set from Kukush et al. [11] to the case of heterogeneous measurement errors. In the next section we proceed with a similar result for unbounded parameter set. The main result of this section is the following theorem.

**Theorem 1.** Under (i)–(vi),  $(\hat{\lambda}, \hat{\beta})$  defined in (1) is a strongly consistent estimator of true parameters  $(\lambda_0, \beta)$ .

In what follows we rely on the following version of the Strong Law of Large Numbers for not necessary identically distributed random variables and the next easy to prove Statement 1.

**Theorem** (Kolmogorov’s Strong Law of Large Numbers, section 10.7 [6]). Let  $\{\xi_n\}_{n \geq 1}$  be a sequence of independent random variables (not necessary identically distributed) and such that

$$\sum_{n=1}^{\infty} \frac{\text{Var}(\xi_n)}{n^2} < \infty.$$

Then

$$\frac{1}{n} \sum_{i=1}^n \xi_i - \frac{1}{n} \sum_{i=1}^n \mathbb{E}\xi_i \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

**Statement 1.** Let  $\{s_n\}_{n \geq 1}$  be a real valued sequence such that  $\frac{1}{n} \sum_{i=1}^n s_i$  is bounded. Then  $\sum_{n=1}^{\infty} s_n/n^2$  converges.

**Remark 1.** By  $K$  we will denote any positive deterministic constant the exact value of which is not important. Note that  $K$  may change from line to line (or even within one line).

**Remark 2.** Condition (iii) implies that there exists positive  $K$ , such that for all  $n \geq 1$ ,

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \|\mathbf{U}_i\|^2 \leq K. \tag{3}$$

**Proof of Theorem 1.** Similarly to the proof of Theorem 1 from [11], one can show the strong consistency of the estimators if

(a)  $\sup_{(\lambda, \beta) \in \Theta} |Q_n^{cor}(\lambda, \beta) - q_\infty(\lambda, \beta)| \rightarrow 0$  a.s. as  $n \rightarrow \infty$ ;

(b)  $q_\infty(\lambda, \beta) \leq q_\infty(\lambda_0, \beta_0)$ , and equality holds if and only if  $\lambda = \lambda_0, \beta = \beta_0$ .

Denote by  $\frac{\partial q^{cor}}{\partial \beta}$  the derivative of  $q^{cor}$  with respect to vector  $\beta$ . For a fixed value of  $\beta$ , consider  $q^{cor}$  as a function of  $\lambda$ , i.e.  $q^{cor}(\cdot, \beta) : C[0, \tau] \rightarrow \mathbb{R}$ . Denote by  $\frac{\partial q^{cor}}{\partial \lambda}$  the Fréchet derivative of  $q^{cor}$  with respect to the function  $\lambda$ . Then  $\frac{\partial q^{cor}}{\partial \lambda} : C[0, \tau] \rightarrow \mathcal{L}(C[0, \tau], \mathbb{R})$  is a linear continuous functional. For  $h \in C[0, \tau]$ , let  $\left\langle \frac{\partial q^{cor}}{\partial \lambda}, h \right\rangle$  denote the action of the functional  $\frac{\partial q^{cor}}{\partial \lambda}$  on  $h$ . We have

$$\left\langle \frac{\partial q^{cor}}{\partial \lambda}(Y, \Delta, W; \lambda, \beta), h \right\rangle = \frac{\Delta h(Y)}{\lambda(Y)} - \frac{e^{\beta^\top W}}{M_U(\beta)} \int_0^Y h(u) du,$$

$$\left\| \frac{\partial q^{cor}}{\partial \lambda}(Y, \Delta, W; \lambda, \beta) \right\| = \sup_{\|h\|=1} \left\langle \frac{\partial q^{cor}}{\partial \lambda}(Y, \Delta, W; \lambda, \beta), h \right\rangle,$$

$$\frac{\partial q^{cor}}{\partial \beta}(Y_i, \Delta_i, W_i; \lambda, \beta) = \Delta_i \cdot W_i - \frac{M_{U_i}(\beta) W_i - E(U_i e^{\beta^\top U_i})}{M_{U_i}^2(\beta)} e^{\beta^\top U_i} \int_0^Y \lambda(u) du.$$

According to [11] in order to verify (a) it suffices to show:

(a1)  $Q_n^{cor}(\lambda, \beta) - q_\infty(\lambda, \beta) \rightarrow 0$  a.s. as  $n \rightarrow \infty$  for all  $(\lambda, \beta) \in \Theta$ ;

(a2) there exists a positive constant  $K$  such that

$$\frac{1}{n} \sum_{i=1}^n E \sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \beta}(Y_i, \Delta_i, W_i; \lambda, \beta) \right\| \leq K, \tag{4}$$

$$\frac{1}{n} \sum_{i=1}^n E \sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \lambda}(Y_i, \Delta_i, W_i; \lambda, \beta) \right\| \leq K, \tag{5}$$

for all  $n \geq 1$ , where supremum is taken over  $\Theta_\lambda \times \text{conv}(\Theta_\beta)$ .

(a3)  $q_\infty(\lambda, \beta)$  is continuous in  $(\lambda, \beta)$ .

To investigate when the condition (a1) holds, rewrite

$$\begin{aligned} Q_n^{cor}(\lambda, \beta) := & \frac{1}{n} \sum_{i=1}^n \Delta_i \cdot \log \lambda(Y_i) + \frac{1}{n} \sum_{i=1}^n \Delta_i \beta^\top X_i + \frac{1}{n} \sum_{i=1}^n \Delta_i \beta^\top U_i - \\ & - \frac{1}{n} \sum_{i=1}^n \frac{\exp(\beta^\top W_i)}{M_{U_i}(\beta)} \int_0^{Y_i} \lambda(u) du. \end{aligned}$$

First and second summand converge to their expectations due to SLLN. Consider third summand. It holds  $\text{Var}(\Delta_i \boldsymbol{\beta}^\top \mathbf{U}_i) \leq \mathbf{E}(\boldsymbol{\beta}^\top \mathbf{U}_i)^2 \leq K \cdot \mathbf{E}\|\mathbf{U}_i\|^2$ . Then

$$\frac{1}{n} \sum_{i=1}^n \text{Var}(\Delta_i \boldsymbol{\beta}^\top \mathbf{U}_i) \leq K \cdot \frac{1}{n} \sum_{i=1}^n \mathbf{E}\|\mathbf{U}_i\|^2$$

is bounded due to Remark 2. Therefore by Statement 1

$$\sum_{i=1}^{\infty} \frac{\text{Var}(\Delta_i \boldsymbol{\beta}^\top \mathbf{U}_i)}{i^2} < \infty .$$

SLLN yields

$$\frac{1}{n} \sum_{i=1}^n \Delta_i \boldsymbol{\beta}^\top \mathbf{U}_i - \frac{1}{n} \sum_{i=1}^n \mathbf{E}[\Delta_i \boldsymbol{\beta}^\top \mathbf{U}_i] \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty .$$

Consider forth summand. We have

$$\exp(\boldsymbol{\beta}^\top \mathbf{W}_i) \int_0^{Y_i} \lambda(u) du \leq K \cdot e^{(D-\varepsilon)\|\mathbf{W}_i\|} .$$

Due to Jensen’s inequality  $M_{\mathbf{U}_i}(\boldsymbol{\beta}) \geq e^{\mathbf{E}\boldsymbol{\beta}^\top \mathbf{U}_i} = 1$ . For all  $i \geq 1$  using (iii)–(iv) we obtain

$$\begin{aligned} \text{Var} \left( \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du \right) &\leq \mathbf{E} \left( \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du \right)^2 \leq \\ &\leq \frac{K}{\min_{\boldsymbol{\beta}} M_{\mathbf{U}_i}^2(\boldsymbol{\beta})} \mathbf{E} \left[ e^{2(D-\varepsilon)\|\mathbf{X}_i\|} \right] \cdot \mathbf{E} \left[ e^{2(D-\varepsilon)\|\mathbf{U}_i\|} \right] \leq K \cdot \mathbf{E} \left[ e^{2(D-\varepsilon)\|\mathbf{U}_i\|} \right] . \end{aligned}$$

Then

$$\frac{1}{n} \sum_{i=1}^n \text{Var} \left( \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du \right) \leq K \cdot \frac{1}{n} \mathbf{E} \left[ e^{2(D-\varepsilon)\|\mathbf{U}_i\|} \right] .$$

Using Statement 1

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \text{Var} \left( \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du \right) < \infty , \tag{6}$$

and therefore

$$\frac{1}{n} \sum_{i=1}^n \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du - \frac{1}{n} \sum_{i=1}^n \mathbf{E} \left[ \frac{\exp(\boldsymbol{\beta}^\top \mathbf{W}_i)}{M_{\mathbf{U}_i}(\boldsymbol{\beta})} \int_0^{Y_i} \lambda(u) du \right] \rightarrow 0$$

a.s. as  $n \rightarrow \infty$ . Thus, under conditions (i)–(iv) (a1) holds.

Next, we will verify (a2). We have

$$\begin{aligned} \sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \beta}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta) \right\| &\leq \|\mathbf{W}_i\| + \frac{K}{\min_{\beta} M_{U_i}(\beta)} \sup_{\beta} \|\mathbf{W}_i\| e^{\beta^{\top} \|\mathbf{W}_i\|} + \\ &+ \frac{K}{\min_{\beta} M_{U_i}^2(\beta)} \sup_{\beta} \mathbb{E} \left[ \|\mathbf{U}_i\| e^{\beta^{\top} \|\mathbf{U}_i\|} \right] e^{\beta^{\top} \|\mathbf{W}_i\|}. \end{aligned}$$

Then

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \beta}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta) \right\| &\leq \mathbb{E} \|X_1\| + \frac{1}{n} \sum_{i=1}^n \mathbb{E} \|U_i\| + \\ &+ \mathbb{E} \sup_{\beta} \left( \|X\| e^{\beta^{\top} X} \right) \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{E} e^{D\|U_i\|} + e^{D\|X\|} \cdot \frac{1}{n} \sum_{i=1}^n \mathbb{E} \sup_{\beta} \left( \|U_i\| e^{\beta^{\top} U_i} \right). \end{aligned}$$

Conditions (iii)–(iv) imply that  $\mathbb{E} \|X_i\| e^{(D-\varepsilon)\|X_i\|}$  and  $\mathbb{E} \|U_i\| e^{(D-\varepsilon)\|U_i\|}$  are bounded for all  $i \geq 1$ . So there exists a positive constant, such that (4) holds.

One can show that

$$\sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \lambda}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta) \right\| \leq \sup_{\lambda} \frac{1}{\|\lambda\|} + \frac{\tau \cdot e^{D(\|X_i\| + \|U_i\|)}}{\min_{\beta} M_{U_i}(\beta)}.$$

Therefore

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \sup_{(\lambda, \beta) \in \Theta} \left\| \frac{\partial q^{cor}}{\partial \lambda}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta) \right\| \leq K \left( 1 + \mathbb{E} e^{D\|X\|} \frac{1}{n} \sum_{i=1}^n \mathbb{E} e^{D\|U_i\|} \right).$$

Conditions (i)–(iv) imply that (5) holds.

(a3) It is clear that  $q^{cor}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta)$  is continuous in  $(\lambda, \beta) \in \Theta$ , and under conditions (i)–(iv)  $\mathbb{E}|q^{cor}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta)|$  is bounded for all  $1 \leq i \leq n$ . Thus, the dominated convergence theorem implies that  $q_{\infty}(\lambda, \beta)$  is continuous in  $(\lambda, \beta) \in \Theta$ .

Proof that (b) holds is essentially the same as in [11].

To sum up, under (i)–(iv), the pair  $(\hat{\lambda}, \hat{\beta})$  is a strongly consistent estimator of true parameter  $(\lambda_0, \beta)$ . □

#### 4 Simultaneous estimator under unbounded parameter set

Let us now consider baseline hazards function form unbounded parameter set defined in (i'). The goal of this section is to extend Theorem 3 from [12] in the case of heteroscedastic measurement errors. The main result of this section follows.

**Theorem 2.** Under (i')–(vi),  $(\hat{\lambda}, \hat{\beta})$  defined in (2) is a strongly consistent estimator of true parameters  $(\lambda_0, \beta)$ .

**Proof.** We follow the line of the proof of Theorem 3 from [12]. A relation holds eventually if it is valid for all sample sizes  $n$  starting from some random number, almost surely.



We firstly show that

$$\sup_{(\lambda, \beta) \in \Theta^R} Q_n^{cor}(\lambda, \beta) > \sup_{(\lambda, \beta) \in \Theta \setminus \Theta^R} Q_n^{cor}(\lambda, \beta) \tag{7}$$

eventually for sufficiently large nonrandom numbers  $R > \|\lambda_0\|$ , where  $\Theta_\lambda^R = \Theta_\lambda \cap \bar{B}(0, R)$ ,  $\Theta^R = \Theta_\lambda^R \times \Theta_\beta$ .

Denote  $D_1 = \max_\beta \|\beta\|$ . We have

$$\sup_{(\lambda, \beta) \in \Theta \setminus \Theta^R} Q_n^{cor}(\lambda, \beta) \leq I_1 + \sup_{\substack{\lambda \in \Theta_\lambda; \\ \lambda(0) > R}} I_2 + I_3,$$

where

$$I_1 = -(R - L\tau) \frac{1}{n} \sum_{i=1}^n \frac{\exp(-D_1 \|W_i\|) Y_i \cdot I(\Delta = 0)}{\max_{\beta \in \Theta_\beta} M_U(\beta)},$$

$$I_2 = \ln(\lambda(0) + L\tau) \frac{1}{n} \sum_{i:\Delta_i=1} \Delta_i - (\lambda(0) + L\tau) \frac{1}{n} \sum_{i=1}^n \frac{\exp(-D_1 \|W_i\|) Y_i \cdot I(\Delta = 1)}{\max_{\beta \in \Theta_\beta} M_U(\beta)},$$

$$I_3 = \frac{1}{n} \sum_{i:\Delta_i=1} D_1 \|W_i\| + 2L\tau \frac{1}{n} \sum_{i=1}^n \frac{\exp(-D_1 \|W_i\|) Y_i \cdot I(\Delta = 1)}{\max_{\beta \in \Theta_\beta} M_U(\beta)}.$$

We have

$$\text{Var} \left( Y_i e^{-D_1 \|W_i\|} I(\Delta = 0) \right) \leq \mathbb{E} \left( Y_i^2 e^{-2D_1 \|W_i\|} \right) \leq K \cdot \mathbb{E} e^{-2D_1 \|X_1\|} \mathbb{E} e^{-2D_1 \|U_i\|},$$

$$\frac{1}{n} \sum_{i=1}^n \text{Var} \left( Y_i e^{-D_1 \|W_i\|} I(\Delta = 0) \right) \leq K \cdot \mathbb{E} e^{-2D_1 \|X_1\|} \frac{1}{n} \sum_{i=1}^n \mathbb{E} e^{-2D_1 \|U_i\|} \leq K.$$

SLLN yields

$$I_1 + (R - L\tau) \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{E} [ C \cdot I(\Delta = 0) \exp(-D_1 \|W_i\|) ]}{\max_{\beta \in \Theta_\beta} M_U(\beta)} \rightarrow 0$$

almost surely as  $n \rightarrow \infty$ . This means that eventually

$$I_1 \leq -(R - L\tau) D_2,$$

where  $D_2 > 0$ .

Let

$$A_n = \frac{1}{n} \sum_{i=1}^n \Delta_i, \quad B_n = \frac{1}{n} \sum_{i=1}^n \frac{\exp(-D_1 \|W_i\|) Y_i}{\max_{\beta \in \Theta_\beta} M_U(\beta)} 1_{\{\Delta_i=1\}}.$$

Since  $A_n > 0$  and  $B_n > 0$  eventually, we obtain

$$I_2 \leq \max_{z>0} (A_n \ln z - z B_n) = A_n \left( \ln \left( \frac{A_n}{B_n} \right) - 1 \right).$$

Analogously,

$$\begin{aligned} \text{Var} \left( Y_i e^{-D_1 \|W_i\|} I(\Delta = 1) \right) &\leq \mathbb{E} \left( Y_i^2 e^{-2D_1 \|W_i\|} \right) \leq K \cdot \mathbb{E} e^{-2D_1 \|X_i\|} \mathbb{E} e^{-2D_1 \|U_i\|}, \\ \frac{1}{n} \sum_{i=1}^n \text{Var} \left( Y_i e^{-D_1 \|W_i\|} I(\Delta = 1) \right) &\leq K. \end{aligned}$$

By SLLN,

$$A_n \rightarrow \mathbb{P}(\Delta = 1) > 0, \quad B_n - \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{E}[T \cdot I(\Delta = 1) \exp(-D_1 \|W_i\|)]}{\max_{\beta \in \Theta_\beta} M_U(\beta)} \rightarrow 0$$

respectively, almost surely as  $n \rightarrow \infty$ . Hence  $I_2$  is *eventually* bounded from above by some positive constant  $D_3$ .

Further, it follows from the strong law of large numbers that  $I_3$  is *eventually* bounded from above by some positive constant  $D_4$ . Hence

$$\overline{\lim}_{n \rightarrow \infty} \sup_{(\lambda, \beta) \in \Theta \setminus \Theta^R} Q_n^{cor}(\lambda, \beta) \leq -(R - L\tau)D_2 + D_3 + D_4.$$

Note that the constants  $D_2, D_3$  and  $D_4$  introduced above do not depend on  $\beta \in \Theta_\beta$ . Letting  $R \rightarrow +\infty$ , we get

$$\overline{\lim}_{n \rightarrow \infty} \sup_{(\lambda, \beta) \in \Theta \setminus \Theta^R} Q_n^{cor}(\lambda, \beta) \rightarrow -\infty, \quad R \rightarrow +\infty.$$

This proves that inequality (7) holds *eventually* for sufficiently large  $R$ .

Further, one can repeat reasoning from [12] to show that

$$(\hat{\lambda}_n(\omega), \hat{\beta}_n(\omega)) \rightarrow (\lambda_0, \beta), \quad n \rightarrow \infty,$$

for all  $\omega \in A, P(A) = 1$ .

The theorem is proved. □

### 5 Conclusions

We have shown consistency of a simultaneous consistent estimator of the baseline hazard rate and the regression parameter in the Cox proportional hazards model assuming baseline hazard function is Lipschitz continuous with fixed constant in the case of heteroscedastic measurement errors for both bounded and unbounded parameter set, respectively.

### Acknowledgement

I would like to express my sincere gratitude to my mentor, Professor Oleksandr Kukush, for his invaluable guidance and support throughout my journey in survival analysis. His expertise, encouragement, and dedication have been instrumental in shaping my understanding of this field. I am deeply thankful for his mentorship, which has greatly enriched my research experience and academic pursuits.

## References

- [1] Augustin, T.: An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scand. J. Stat.* **31**(1), 43–50 (2004). [MR2042597](#). <https://doi.org/10.1111/j.1467-9469.2004.00371.x>
- [2] Augustin, T., Döring, A., Rummel, D.: Regression calibration for Cox regression under heteroscedastic measurement error—determining risk factors of cardiovascular diseases from error-prone nutritional replication data. In: *Recent Advances in Linear Models and Related Areas*, pp. 253–278. Springer (2008). [MR2523854](#). [https://doi.org/10.1007/978-3-7908-2064-5\\_13](https://doi.org/10.1007/978-3-7908-2064-5_13)
- [3] Breslow, N.: Covariance analysis of censored survival data. *Biometrics*, 89–99 (1974)
- [4] Cox, D.R.: Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* **34**, 187–220 (1972). [MR0341758](#)
- [5] Durot, C., Lopuhaä, H.P.: Limit theory in monotone function estimation. *Statist. Sci.* **33**(4), 547–567 (2018). [MR3881208](#). <https://doi.org/10.1214/18-STS664>
- [6] Feller, W.: *An Introduction to Probability Theory and Its Applications*. Vol. I, 509 (1968). [MR0228020](#)
- [7] Grenander, U.: On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* **39**, 125–153 (1956). [MR0093415](#). <https://doi.org/10.1080/03461238.1956.10414944>
- [8] Groeneboom, P., Jongbloed, G.: Some developments in the theory of shape constrained inference. *Statist. Sci.* **33**(4), 473–492 (2018). [MR3881204](#). <https://doi.org/10.1214/18-STS657>
- [9] Kalbfleisch, J.D., Prentice, R.L.: *The Statistical Analysis of Failure Time Data* vol. 360. John Wiley & Sons (2011) [MR0570114](#)
- [10] Kong, F.H., Gu, M.: Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statist. Sinica* **9**(4), 953–969 (1999). [MR1744820](#)
- [11] Kukush, A., Baran, S., Fazekas, I., Usoltseva, E.: Simultaneous estimation of baseline hazard rate and regression parameters in Cox proportional hazards model with measurement error. *J. Statist. Res.* **45**(2), 77–94 (2011). [MR2934363](#)
- [12] Kukush, O.G., Chernova, O.O.: Consistent estimation in the Cox proportional hazards model with measurement errors under an unboundedness condition for the parameter set. *Teor. Imovir. Mat. Stat.* **96**, 100–109 (2017). [MR3666874](#). <https://doi.org/10.1090/tpms/1036>
- [13] Lawless, J.F.: *Statistical Models and Methods for Lifetime Data* vol. 362. John Wiley & Sons (2011) [MR0640866](#)
- [14] Lopuhaä, H.P., Nane, G.F.: Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. *Scand. J. Stat.* **40**(3), 619–646 (2013). [MR3091700](#). <https://doi.org/10.1002/sjos.12008>
- [15] Qin, J., Deng, G., Ning, J., Yuan, A., Shen, Y.: Estrogen receptor expression on breast cancer patients' survival under shape-restricted Cox regression model. *Ann. Appl. Stat.* **15**(3), 1291–1307 (2021). [MR4316649](#). <https://doi.org/10.1214/21-aos1446>
- [16] Samworth, R.J.: Recent progress in log-concave density estimation. *Statist. Sci.* **33**(4), 493–509 (2018). [MR3881205](#). <https://doi.org/10.1214/18-STS666>
- [17] Wallace, M.: Analysis in an imperfect world. *Significance* **17**(1), 14–19 (2020). [MR4446481](#)