

**CONSISTENT ESTIMATION IN COX PROPORTIONAL HAZARDS  
MODEL WITH MEASUREMENT ERRORS  
AND UNBOUNDED PARAMETER SET**

ALEXANDER KUKUSH AND OKSANA CHERNOVA

ABSTRACT. Cox proportional hazards model with measurement error is investigated. In Kukush et al. (2011) [Journal of Statistical Research **45**, 77–94] and Chimisov and Kukush (2014) [Modern Stochastics: Theory and Applications **1**, 13–32] asymptotic properties of simultaneous estimator  $\lambda_n(\cdot)$ ,  $\beta_n$  were studied for baseline hazard rate  $\lambda(\cdot)$  and regression parameter  $\beta$ , at that the parameter set  $\Theta = \Theta_\lambda \times \Theta_\beta$  was assumed bounded. In the present paper, the set  $\Theta_\lambda$  is unbounded from above and not separated away from 0. We construct the estimator in two steps: first we derive a strongly consistent estimator and then modify it to provide its asymptotic normality.

1. INTRODUCTION

Consider the Cox proportional hazards model (Cox, 1972), where a lifetime  $T$  has the following intensity function

$$\lambda(t|X; \lambda, \beta) = \lambda(t) \exp(\beta^T X), \quad t \geq 0. \quad (1)$$

A covariate  $X$  is a given random vector distributed in  $\mathbb{R}^m$ ,  $\beta$  is a parameter belonging to  $\Theta_\beta \subset \mathbb{R}^m$ , and  $\lambda(\cdot) \in \Theta_\lambda \subset C[0, \tau]$  is a baseline hazard function.

We use common random censorship model: instead of  $T$ , only a couple  $(Y, \Delta)$  is available, where  $Y := \min\{T, C\}$  and  $\Delta := I_{\{T \leq C\}}$  is the censorship indicator. The censor  $C$  is distributed on  $[0, \tau]$ . Its survival function  $G_C(u) = 1 - F_C(u)$  is unknown, while we know  $\tau$ . The conditional pdf of  $T$  given  $X$  is

$$f_T(t|X, \lambda, \beta) = \lambda(t|X; \lambda, \beta) \exp\left(-\int_0^t \lambda(s|X; \lambda, \beta) ds\right).$$

Throughout this paper an additive error model is considered, i.e., instead of  $X$  a surrogate variable

$$W = X + U$$

is observed, where a random error  $U$  has known moment generating function  $M_U(z) := \mathbb{E}e^{z^T U}$ . A couple  $(T, X)$ , censor  $C$ , and measurement error  $U$  are stochastically independent.

Consider independent copies of the model  $(X_i, T_i, C_i, Y_i, \Delta_i, U_i, W_i)$ ,  $i = 1, \dots, n$ . Based on triples  $(Y_i, \Delta_i, W_i)$ ,  $i = 1, \dots, n$ , we estimate true parameters  $\beta_0$  and  $\lambda_0(t)$ ,  $t \in [0, \tau]$ . Due to the suggestion of Augustin (2004) we use the following objective function

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

---

Received 01/03/2017.

2000 *Mathematics Subject Classification*. Primary 62N02; Secondary 62N01.

*Key words and phrases*. asymptotically normal estimator, consistent estimator, Cox proportional hazards model, simultaneous estimation of baseline hazard rate and regression parameter.

where

$$q(Y, \Delta, W; \lambda, \beta) := \Delta(\log \lambda(Y) + \beta^T W) - \frac{\exp(\beta^T W)}{M_U(\beta)} \int_0^Y \lambda(u) du.$$

The corrected estimator is defined as

$$(\hat{\lambda}_n, \hat{\beta}_n) = \arg \max_{(\lambda, \beta) \in \Theta} Q_n^{cor}(\lambda, \beta), \quad (2)$$

where  $\Theta := \Theta_\lambda \times \Theta_\beta$ . If the parameter sets  $\Theta_\lambda$  and  $\Theta_\beta$  are compact, then  $\Theta$  is compact as well and the maximum in (2) is attained.

The issue of estimating  $\beta_0$  and cumulative hazard  $\Lambda(t) = \int_0^t \lambda_0(s) ds$  has been extensively studied in the literature in past decades: in Andersen and Gill (1982) general ideas are presented based on partial likelihood; model with measurement errors is considered in Gu and Kong (1999), where, based on Corrected Score method, consistent and asymptotically normal estimators are constructed for regression parameter and cumulative hazard function; Royston (2011) discusses some problems where the behavior of baseline hazard function  $\lambda_0(\cdot)$  itself, rather than cumulative hazard, is needed.

Our model is presented in Augustin (2004), where the baseline hazard function is assumed to belong to a parameter space, while we consider  $\lambda_0(\cdot)$  from a compact set of  $C[0, \tau]$ .

In [6] the consistency of estimator (2) is proven for a bounded parameter set. In [3] its asymptotic normality is presented. We remark that in [6] the authors write  $\Theta_\lambda$  without a formal requirement that  $\lambda(0)$  is bounded, though actually this assumption was used throughout the paper. We prove that this condition and the separation of  $\lambda(\cdot)$  away from zero are too restrictive and can be omitted.

The paper is organized as follows. In Section 2, we define an estimator under unbounded parameter set and prove its consistency. Additionally, we describe a numerical scheme for calculation of the estimator. In Section 3, we modify the estimator constructed in Section 2 to produce the asymptotically normal estimator, and Section 4 concludes.

## 2. CONSISTENT ESTIMATION ON THE FIRST STAGE

Impose conditions on the parameter sets.

- (i)  $K_\lambda \subset C[0, \tau]$  is the following closed convex set of nonnegative functions

$$K_\lambda := \{ f : [0, \tau] \rightarrow \mathbb{R} \mid f(t) \geq 0, \forall t \in [0, \tau] \text{ and } |f(t) - f(s)| \leq L|t - s|, \forall t, s \in [0, \tau] \},$$

where  $L > 0$  is a fixed constant.

- (ii)  $\Theta_\beta \subset \mathbb{R}^m$  is a compact set.

The following conditions (iii) – (vi) are borrowed from [6].

- (iii)  $EU = 0$  and for some constant  $\epsilon > 0$ ,

$$Ee^{D\|U\|} < \infty, \text{ where } D := \max_{\beta \in \Theta_\beta} \|\beta\| + \epsilon.$$

- (iv)  $Ee^{D\|X\|} < \infty$ , where  $D > 0$  is defined in (iii).

- (v)  $\tau$  is the right endpoint of the distribution of  $C$ , that is  $P(C > \tau) = 0$  and for all  $\epsilon > 0$ ,  $P(C > \tau - \epsilon) > 0$ .

- (vi) The covariance matrix of random vector  $X$  is positive definite.

Denote

$$K = K_\lambda \times \Theta_\beta. \quad (3)$$

If  $\lambda(Y) = 0$  then we put

$$\Delta \cdot \log \lambda(Y) = \begin{cases} 0 & \text{if } \Delta = 0 \\ -\infty & \text{if } \Delta = 1. \end{cases}$$

**Definition 1.** Fix a sequence  $\{\varepsilon_n\}$  of positive numbers, with  $\varepsilon_n \downarrow 0$ , as  $n \rightarrow \infty$ . The corrected estimator  $(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)})$  of  $(\lambda, \beta)$  is a Borel measurable function of observations  $(Y_i, \Delta_i, W_i)$ ,  $i = 1, \dots, n$ , with values in  $K$  and such that

$$Q_n^{cor}(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}) \geq \sup_{(\lambda, \beta) \in K} Q_n^{cor}(\lambda, \beta) - \varepsilon_n. \quad (4)$$

The corrected estimator exists due to Pfanzagl (1969) (it is essential here that the supremum in (4) is finite). Additionally, assume the following.

- (vii) True parameters  $(\lambda_0, \beta_0)$  belong to  $K$ , which is given in (3), and moreover  $\lambda_0(t) > 0$ ,  $t \in [0, \tau]$ .

**Definition 2.** Let  $A_n = A_n(\omega)$ ,  $n = 1, 2, \dots$ , be a sequence of statements depending on an elementary event  $\omega \in \Omega$ . We say that  $A_n$  holds eventually if for almost all  $\omega$  there exists  $n_0 = n_0(\omega)$  such that for all  $n \geq n_0(\omega)$ ,  $A_n$  holds true.

**Theorem 3.** Under conditions (i) to (vii) the estimator  $(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)})$  is a strongly consistent estimator of the true parameters  $(\lambda_0, \beta_0)$ , that is

$$\max_{t \in [0, \tau]} |\hat{\lambda}_n^{(1)}(t) - \lambda_0(t)| \rightarrow 0 \quad \text{and} \quad \hat{\beta}_n^{(1)} \rightarrow \beta_0$$

almost surely, as  $n \rightarrow \infty$ .

*Proof.* For  $R > 0$  denote

$$K_\lambda^R = K_\lambda \cap \bar{B}(0, R) \quad \text{and} \quad K^R = K_\lambda^R \times \Theta_\beta,$$

where  $\bar{B}(0, R)$  is closed ball in  $C[0, \tau]$  with center in the origin and radius  $R$ .

1. In the first part of the proof, we show that for large enough nonrandom  $R > \|\lambda_0\|$ , it holds eventually

$$\sup_{(\lambda, \beta) \in K^R} Q_n^{cor}(\lambda, \beta) > \sup_{(\lambda, \beta) \in K \setminus K^R} Q_n^{cor}(\lambda, \beta). \quad (5)$$

For  $\lambda \in K_\lambda$  the Lipschitz condition implies

$$\lambda(0) - L\tau \leq \lambda(t) \leq \lambda(0) + L\tau, \quad (6)$$

therefore,

$$q(Y_i, \Delta_i, W_i; \lambda, \beta) \leq \Delta_i (\ln(\lambda(0) + L\tau) + \beta^T W_i) - \frac{\exp(\beta^T W_i) Y_i}{M_U(\beta)} (\lambda(0) - L\tau).$$

Using the Lipschitz condition for  $\lambda \in K_\lambda$ , one can show that if  $\lambda(t_1) > R$  for some  $t_1 \in [0, \tau]$ , then  $\lambda(t) > R - L\tau$  for all  $t \in [0, \tau]$ . On the other side,  $\lambda(0) > R$  yields  $\lambda(t) > R - L\tau$ ,  $t \in [0, \tau]$ . Thus, supremum on the right hand side of (5) can be taken over the set  $\{\lambda \in K_\lambda : \lambda(0) > R\} \times \Theta_\beta$ .

Denote

$$D_1 = \max_{\beta \in \Theta_\beta} \|\beta\|.$$

We have

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

with

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

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

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

By the strong law of large numbers (SLLN),

$$I_1 \rightarrow -(R - L\tau) \frac{\mathbb{E}[C \cdot I(\Delta = 0) \exp(-D_1 \|W\|)]}{\max_{\beta \in \Theta_\beta} M_U(\beta)}$$

a.s., as  $n \rightarrow \infty$ . This means that *eventually*

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

where  $D_2 > 0$ .

Denote

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

Since  $A_n > 0$  and  $B_n > 0$  *eventually*, for  $\lambda(0) > R$  we get

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

By the SLLN

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

a.s., as  $n \rightarrow \infty$ . Therefore,  $I_2$  *eventually* is bounded from above by some positive constant  $D_3$ .

Further, using the SLLN it can be shown that *eventually*  $I_3$  is also bounded from above by some positive constant  $D_4$ . Thus,

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

Here  $D_2$ ,  $D_3$ , and  $D_4$  do not depend on  $\beta \in \Theta_\beta$ . Tend  $R \rightarrow +\infty$  and obtain

$$\limsup_{n \rightarrow \infty} \sup_{(\lambda, \beta) \in K \setminus K^R} Q_n^{cor}(\lambda, \beta) \rightarrow -\infty, \text{ as } R \rightarrow +\infty.$$

This proves that the inequality (5) holds *eventually* for large enough  $R$ .

Therefore, we may and do replace  $K$  for  $K^R$  in Definition 1. Thus, we assume that for all  $n \geq 1$ ,

$$Q_n^{cor}(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}) \geq \sup_{(\lambda, \beta) \in K^R} Q_n^{cor}(\lambda, \beta) - \varepsilon_n \quad (7)$$

and  $(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}) \in K^R$ . Notice that  $K^R$  is a compact set in  $C[0, \tau]$ .

2. Since  $R > \|\lambda_0\|$ , we have  $(\lambda_0, \beta_0) \in K^R$ . Then (7) implies the inequality

$$Q_n^{cor}(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}) \geq Q_n^{cor}(\lambda_0, \beta_0) - \varepsilon_n. \quad (8)$$

We fix  $\omega \in A \subset \Omega$ , with  $\mathbb{P}(A) = 1$ . Further, we will impose additional conditions on  $A$ .

We want to show that at point  $\omega$ ,  $(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}) \rightarrow (\lambda_0, \beta_0)$ . We have

$$Q_n^{cor}(\lambda_0, \beta_0) \rightarrow q_\infty(\lambda_0, \beta_0) := \mathbb{E}[q(Y, \Delta, W; \lambda_0, \beta_0)]. \quad (9)$$

This holds almost surely, so we can assume (9) for the fixed  $\omega$ . Therefore, the first condition is

$$Q_n^{cor}(\lambda_0, \beta_0; \omega) \rightarrow q_\infty(\lambda_0, \beta_0), \quad \omega \in A.$$

The sequence  $\{(\hat{\lambda}_n^{(1)}(\omega), \hat{\beta}_n^{(1)}(\omega)), n \geq 1\}$  belongs to the compact set  $K^R$ . Consider an arbitrary convergent subsequence

$$(\hat{\lambda}_{n'}^{(1)}(\omega), \hat{\beta}_{n'}^{(1)}(\omega)) \rightarrow (\lambda_*, \beta_*) \in K^R. \quad (10)$$

Then (8), (9) imply that

$$\begin{aligned} q_\infty(\lambda_0, \beta_0) &\leq \liminf_{n' \rightarrow \infty} Q_{n'}^{cor}(\hat{\lambda}_{n'}^{(1)}, \hat{\beta}_{n'}^{(1)}) = \\ &= \liminf_{n' \rightarrow \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \log \hat{\lambda}_{n'}^{(1)}(Y_i) + \\ &+ \lim_{n' \rightarrow \infty} \frac{1}{n'} \sum_{i=1}^{n'} \left( \Delta_i \hat{\beta}_{n'}^{(1)T} W_i - \frac{\exp(\hat{\beta}_{n'}^{(1)T} W_i)}{M_U(\hat{\beta}_{n'}^{(1)})} \int_0^{Y_i} \hat{\lambda}_{n'}^{(1)}(u) du \right). \end{aligned}$$

The next assumption on  $A$  is as follows: for all  $\omega \in A$ , a sequence of random functions

$$\frac{1}{n} \sum_{i=1}^n \left( \Delta_i \beta^T W_i - \frac{\exp(\beta^T W_i)}{M_U(\beta)} \int_0^{Y_i} \lambda(u) du \right)$$

converges uniformly in  $(\lambda, \beta) \in K^R$  to

$$\mathbb{E} \left[ \Delta \beta^T W - \frac{\exp(\beta^T W)}{M_U(\beta)} \int_0^Y \lambda(u) du \right] =: q_\infty^2(\lambda, \beta).$$

Such condition can be imposed because for any fixed  $(\lambda, \beta) \in K^R$  the latter sequence converges to  $q_\infty^2$  a.s., the sequence is equicontinuous a.s. on the compact set  $K^R$ , and the limit function is continuous on  $K^R$ . These three statements ensure that the sequence converges to  $q_\infty^2$  uniformly on  $K^R$ .

The function  $q_\infty^2$  is continuous in  $(\lambda, \beta) \in K^R$ , thus,

$$q_\infty(\lambda_0, \beta_0) \leq \liminf_{n' \rightarrow \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \ln \hat{\lambda}_{n'}^{(1)}(Y_i) + q_\infty^2(\lambda_*, \beta_*).$$

For large  $n'$ , it holds

$$\hat{\lambda}_{n'}^{(1)}(t) \leq \lambda_*(t) + \varepsilon, \quad t \in [0, \tau],$$

with fixed  $\varepsilon > 0$ . We demand also that

$$\frac{1}{n} \sum_{i=1}^n \Delta_i \ln \lambda(Y_i) \rightarrow \mathbb{E}[\Delta \lambda(Y)]$$

uniformly in  $(\lambda, \beta) \in \left( K_\lambda^{R+\delta_k} \cap \{\lambda : \lambda(t) \geq \delta_k\} \right) \times \Theta_\beta$ , for each  $k \geq 1$  and  $\omega \in A$ , where  $\delta_k \downarrow 0$ ,  $\{\delta_k\}$  is a fixed sequence of positive numbers. Then by the SLLN

$$\begin{aligned} \liminf_{n' \rightarrow \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \log \hat{\lambda}_{n'}^{(1)}(Y_i) &\leq \liminf_{n' \rightarrow \infty} \frac{1}{n'} \sum_{i=1}^{n'} \Delta_i \log(\lambda_*(Y_i) + \varepsilon) = \\ &= \mathbb{E}[\Delta \cdot \log(\lambda_*(Y) + \varepsilon)] =: q_\infty^{1,\varepsilon}(\lambda_*). \end{aligned}$$

Hence, for each  $\varepsilon > 0$ ,

$$q_\infty(\lambda_0, \beta_0) \leq q_\infty^{1,\varepsilon}(\lambda_*) + q_\infty^2(\lambda_*, \beta_*).$$

Now, tend  $\varepsilon \rightarrow 0$ . We have

$$q_\infty^{1,\varepsilon}(\lambda_*) = \mathbb{E} \left[ \Delta \cdot \log(\lambda_*(Y) + \varepsilon) I(\lambda_*(Y) > \frac{1}{2}) \right] + \mathbb{E} \left[ \Delta \cdot \log(\lambda_*(Y) + \varepsilon) I(\lambda_*(Y) \leq \frac{1}{2}) \right].$$

The first expectation tends to

$$\mathbb{E} \left[ \Delta \cdot \log(\lambda_*(Y)) I(\lambda_*(Y) > \frac{1}{2}) \right]$$

by the Lebesgue dominance convergence theorem, and the second expectation tends to

$$\mathbb{E} \left[ \Delta \cdot \log(\lambda_*(Y)) I(\lambda_*(Y) \leq \frac{1}{2}) \right]$$

by the Lebesgue monotone convergence theorem. Then

$$q_{\infty}^{1,\varepsilon}(\lambda_*, \beta_*) \rightarrow q_{\infty}^1(\lambda_*, \beta_*) := \mathbf{E}[\Delta \cdot \log \lambda_*(Y)],$$

as  $\varepsilon \rightarrow 0$ . Therefore,

$$q_{\infty}(\lambda_0, \beta_0) \leq q_{\infty}^1(\lambda_*) + q_{\infty}^2(\lambda_*, \beta_*) = q_{\infty}(\lambda_*, \beta_*).$$

But according to [6] the inequality

$$q_{\infty}(\lambda_0, \beta_0) \geq q_{\infty}(\lambda_*, \beta_*)$$

holds true, moreover the equality is attained if, and only if,  $\lambda_* = \lambda_0$  and  $\beta_* = \beta_0$ . Therefore, a convergent subsequence (10) converges exactly to  $(\lambda_0, \beta_0)$ . Since the whole sequence belongs to a compact set, this implies the convergence

$$(\hat{\lambda}_n^{(1)}(\omega), \hat{\beta}_n^{(1)}(\omega)) \rightarrow (\lambda_0, \beta_0), \quad \text{as } n \rightarrow \infty.$$

This holds for almost every  $\omega \in \Omega$ , and the strong consistency is proven.  $\square$

Now, we explain how the estimator can be computed. Like in [3] we prove that for a fixed  $\beta \in \Theta_{\beta}$ , the function  $\hat{\lambda}_n^{(1)}$  that maximizes  $Q_n^{cor}$  is a linear spline.

**Theorem 4.** *Under conditions (i) and (ii) the function  $\hat{\lambda}_n^{(1)}$ , which maximizes  $Q_n^{cor}$ , is a linear spline.*

*Proof.* Let  $(Y_{i_1}, \dots, Y_{i_n})$  be a variational series of  $Y_1, \dots, Y_n$ . Fix  $\beta \in \Theta_{\beta}$ . Suppose that we are given  $\hat{\lambda}_n^{(1)} \in \Theta_{\lambda}$  that maximizes  $Q_n^{cor}(\cdot, \beta)$ . Together with  $(\hat{\lambda}_n^{(1)}, \beta)$  consider  $(\bar{\lambda}_n, \beta)$ , where  $\bar{\lambda}_n$  is the following function. We set  $\bar{\lambda}_n(Y_{i_k}) = \hat{\lambda}_n^{(1)}(Y_{i_k})$ ,  $k = 1, \dots, n$ . At each interval  $[Y_{i_k}, Y_{i_{k+1}}]$ ,  $k = 1, \dots, n-1$ , draw straight lines

$$L_{i_k}^1(t) = \hat{\lambda}_n^{(1)}(Y_{i_k}) + L(Y_{i_k} - t),$$

$$L_{i_k}^2(t) = \hat{\lambda}_n^{(1)}(Y_{i_{k+1}}) + L(t - Y_{i_{k+1}}),$$

with  $L$  defined in (i). Denote by  $B_{i_k}$  the intersection of  $L_{i_k}^1(t)$  and  $L_{i_k}^2(t)$ . Additionally,  $B_{i_0} := 0$ ,  $B_{i_n} := \tau$ ,  $Y_{i_0} := 0$ ,  $Y_{i_{n+1}} := \tau$ . Finally, define the function  $\bar{\lambda}_n(t)$  as

$$\bar{\lambda}_n(t) = \begin{cases} L_{i_0}^2(t) & \text{if } t \in [0, Y_{i_1}], \\ \max\{L_{i_k}^1(t), 0\} & \text{if } t \in [Y_{i_k}, B_{i_k}], k = 1, \dots, n-1, \\ \max\{L_{i_k}^2(t), 0\} & \text{if } t \in [B_{i_k}, Y_{i_{k+1}}], k = 1, \dots, n-1, \\ L_{i_n}^1(t) & \text{if } t \in [Y_{i_n}, \tau]. \end{cases} \quad (11)$$

It is easily seen that  $\bar{\lambda}_n \in \Theta_{\lambda}$ . By construction,  $\hat{\lambda}_n^{(1)} \geq \bar{\lambda}_n$ . Thus,

$$Q_n^{cor}(\hat{\lambda}_n^{(1)}, \beta) \leq Q_n^{cor}(\bar{\lambda}_n, \beta).$$

This yields  $\hat{\lambda}_n^{(1)} = \bar{\lambda}_n$  and completes the proof.  $\square$

Notice that *eventually*  $\bar{\lambda}_n(B_{i_k}) > 0$ , and then one can omit maximum in (11).

As soon as we constructed a linear spline

$$\bar{\lambda}_n(\beta) = \arg \max_{\lambda: (\lambda, \beta) \in \Theta} Q_n^{cor},$$

we maximize  $Q(\beta) := Q_n^{cor}(\bar{\lambda}_n(\beta), \beta)$  in  $\beta \in \Theta_{\beta}$ , i.e., we search for a  $\hat{\beta} \in \Theta_{\beta}$  such that

$$Q(\hat{\beta}) \geq \sup_{\beta \in \Theta_{\beta}} Q(\beta) - \varepsilon_n.$$

Since  $Q(\beta)$  is bounded,  $\hat{\beta}$  does exist.

We have

$$Q_n^{cor}(\bar{\lambda}_n(\hat{\beta}), \hat{\beta}) \geq \sup_{\beta \in \Theta_\beta} \max_{\lambda \in \Theta_\lambda} Q(\beta) - \varepsilon_n = \sup_{(\lambda, \beta) \in \Theta} Q_n^{cor}(\lambda, \beta) - \varepsilon_n.$$

Therefore, the estimator  $(\bar{\lambda}_n(\hat{\beta}), \hat{\beta})$  satisfies Definition 1 and its evaluation is just a parametric problem.

### 3. CONSTRUCTION OF ASYMPTOTICALLY NORMAL ESTIMATOR ON THE SECOND STAGE

In this section, we modify the estimator  $(\hat{\lambda}_n^{(1)}(\omega), \hat{\beta}_n^{(1)}(\omega))$  from Definition 1 in order to produce the asymptotically normal estimator.

**Definition 5.** *The modified corrected estimator  $(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$  of  $(\lambda, \beta)$  is a Borel measurable function of observations  $(Y_i, \Delta_i, W_i)$ ,  $i = 1, \dots, n$ , with values in  $K$  and such that*

$$(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)}) = \begin{cases} \arg \max\{Q_n^{cor}(\lambda, \beta) \mid (\lambda, \beta) \in K, \mu_\lambda \geq \frac{1}{2}\mu_{\hat{\lambda}_n^{(1)}}\}, & \text{if } \mu_{\hat{\lambda}_n^{(1)}} > 0, \\ (\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)}), & \text{otherwise,} \end{cases}$$

with  $\mu_\lambda := \min_{t \in [0, \tau]} \lambda(t)$ .

Such estimator exists due to results of Pfanzagl [7]. Notice that by Theorem 3  $\mu_{\hat{\lambda}_n^{(1)}} \rightarrow \mu_{\lambda_0} > 0$  a.s., and *eventually* it holds

$$\begin{aligned} K_1 &:= \{(\lambda, \beta) \in K \mid \mu_\lambda \geq \frac{3}{4}\mu_{\lambda_0}\} \subset \{(\lambda, \beta) \in K \mid \mu_\lambda \geq \frac{1}{2}\mu_{\hat{\lambda}_n^{(1)}}\} \subset \\ &\subset \{(\lambda, \beta) \in K \mid \mu_\lambda \geq \frac{1}{4}\mu_{\lambda_0}\} =: K_2. \end{aligned}$$

The estimator

$$(\hat{\lambda}_n^{(3)}, \hat{\beta}_n^{(3)}) = \arg \max_{(\lambda, \beta) \in K_2} Q_n^{cor}(\lambda, \beta) \quad (12)$$

is strongly consistent under the conditions (i) – (vii), because according to Theorem 3 *eventually* it can be taken as an estimator  $(\hat{\lambda}_n^{(1)}, \hat{\beta}_n^{(1)})$ . Therefore  $(\hat{\lambda}_n^{(3)}, \hat{\beta}_n^{(3)}) \in K_1$  *eventually*, and *eventually*  $(\hat{\lambda}_n^{(3)}, \hat{\beta}_n^{(3)})$  can be taken as an estimator  $(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$ . This implies the strong consistency of  $(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$ .

Introduce additional assumptions, under which the estimator  $(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$  is asymptotically normal.

- (viii)  $\beta_0$  is an interior point of  $\Theta_\beta$ .
- (ix)  $\lambda_0 \in \Theta_\lambda^\epsilon$  for some  $\epsilon > 0$ , where

$$\begin{aligned} \Theta_\lambda^\epsilon &:= \{f : [0, \tau] \rightarrow \mathbb{R} \mid f(t) \geq \epsilon, \forall t \in [0, \tau], \\ &|f(t) - f(s)| \leq (L - \epsilon)|t - s|, \forall t, s \in [0, \tau]\}. \end{aligned}$$

- (x)  $P(C > 0) = 1$ .
- (xi)  $EU = 0$  and for some  $\epsilon > 0$ ,

$$\mathbb{E}e^{2D\|U\|} < \infty, \quad D := \max_{\beta \in \Theta_\beta} \|\beta\| + \epsilon.$$

- (xii)  $\mathbb{E}e^{2D\|X\|} < \infty$ , with  $D$  defined in (xi).

Further, we use notations from [3]. Let

$$\begin{aligned} a(t) &= \mathbb{E}[Xe^{\beta_0^T X} G_T(t|X)], \quad b(t) = \mathbb{E}[e^{\beta_0^T X} G_T(t|X)], \\ p(t) &= \mathbb{E}[XX^T e^{\beta_0^T X} G_T(t|X)], \quad T(t) = p(t)b(t) - a(t)a^T(t), \quad K(t) = \frac{\lambda_0(t)}{b(t)}, \end{aligned}$$

$$A = \mathbb{E} \left[ XX^T e^{\beta_0^T X} \int_0^Y \lambda_0(u) du \right], \quad M = \int_0^\tau T(u)K(u)G_c(u)du.$$

For  $i = 1, 2, \dots$ , introduce random variables

$$\zeta_i = -\frac{\Delta_i a(Y_i)}{b(Y_i)} + \frac{\exp(\beta_0^T W_i)}{M_U(\beta_0)} \int_0^{Y_i} a(u)K(u)du + \frac{\partial q}{\partial \beta}(Y_i, \Delta_i, W_i, \beta_0, \lambda_0),$$

with

$$\frac{\partial q}{\partial \beta}(Y, \Delta, W; \lambda, \beta) = \Delta \cdot W - \frac{M_U(\beta)W - E(Ue^{\beta^T U})}{M_U(\beta)^2} \exp(\beta^T W) \int_0^Y \lambda(u)du.$$

Let

$$\begin{aligned} \Sigma_\beta &= 4 \cdot \text{Cov}(\zeta_1), \quad m(\varphi_\lambda) = \int_0^\tau \varphi_\lambda(u)a(u)G_C(u)du, \\ \sigma_\varphi^2 &= 4 \cdot \text{Var} \langle q'(Y, \Delta, W, \lambda_0, \beta_0), \varphi \rangle = \\ &= 4 \cdot \text{Var} \left[ \frac{\Delta \cdot \varphi_\lambda(Y)}{\lambda_0(Y)} - \frac{\exp(\beta_0^T W)}{M_U(\beta_0)} \int_0^Y \varphi_\lambda(u)du + \Delta \cdot \varphi_\beta^T W + \right. \\ &\quad \left. + \varphi_\beta^T \frac{M_U(\beta_0)W - E(Ue^{\beta_0^T U})}{M_U(\beta_0)^2} \exp(\beta_0^T W) \int_0^Y \lambda_0(u)du \right] \end{aligned}$$

with  $\varphi = (\varphi_\lambda, \varphi_\beta) \in C[0, \tau] \times R^m$ , where  $q'$  denotes the Fréchet derivative.

Now, we can apply Theorem 1 from [3] to state the asymptotic normality for  $\hat{\beta}_n^{(2)}$  and  $\hat{\lambda}_n^{(2)}$  (it follows from the asymptotic normality of consistent estimators  $\hat{\beta}_n^{(3)}$  and  $\hat{\lambda}_n^{(3)}$ ).

**Theorem 6.** *Assume conditions (i), (ii), (v) – (xii). Then  $M$  is nonsingular and*

$$\sqrt{n}(\hat{\beta}_n^{(2)} - \beta_0) \xrightarrow{d} N_m(0, M^{-1}\sigma_\beta M^{-1}).$$

Moreover, for any Lipschitz continuous function  $f$  on  $[0, \tau]$ ,

$$\sqrt{n} \int_0^\tau (\hat{\lambda}_n^{(2)} - \lambda_0)(u)f(u)G_C(u)du \xrightarrow{d} N(0, \sigma_\varphi^2(f)),$$

where  $\sigma_\varphi^2(f) = \sigma_\varphi^2$  with  $\varphi = (\varphi_\lambda, \varphi_\beta)$ ,  $\varphi_\beta = -A^{-1}m(\varphi_\lambda)$ , and  $\varphi_\lambda$  is a unique solution to the Fredholm's integral equation

$$\frac{\varphi_\lambda}{K(u)} - a^T(u)A^{-1}m(\varphi_\lambda) = f(u).$$

For computation of estimator  $(\hat{\lambda}_n^{(2)}, \hat{\beta}_n^{(2)})$  we refer to [3].

#### 4. CONCLUSION

Under quite mild assumptions, we construct the estimator for the function  $\lambda(\cdot)$  and parameter  $\beta$  in Cox proportional hazards model with measurement errors. Contrary to Kukush et al. (2011) and Chimisov and Kukush (2014), we consider an unbounded parameter set. The obtained estimator is consistent and can be modified to be asymptotically normal. Also, we describe a numerical scheme for calculation of the estimator. In future we intend to construct confidence regions based on the estimator.

#### REFERENCES

- [1] Andersen, P. K., and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* **10**, 4, 1100-1120.
- [2] Augustin, T. (2004). An exact corrected log-likelihood function for Cox's proportional hazards model under measurement error and some extensions. *Scandinavian Journal of Statistics* **31**, 1, 43-50.
- [3] Chimisov, C., and Kukush, A. (2014). Asymptotic normality of corrected estimator in Cox proportional hazards model with measurement error. *Modern Stochastics: Theory and Applications* **1**, 1, 13-32.
- [4] Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B* **34**, 187-220.



- [5] Gu, M., and Kong, F. H. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* **32**, 9, 953-969.
- [6] Kukush, A., Baran, S., Fazekas, I., and Usoltseva, E. (2011). Simultaneous estimation of baseline hazard rate and regression parameters in Cox proportional hazards model with measurement error. *Journal of Statistical Research* **45**, 2, 77-94.
- [7] Pfanzagl G. (1969). On the measurability and consistency of minimum contrast estimates. *Metrika* **14**, 249-273.
- [8] Royston, P. (2011). Estimating a smooth baseline hazard function for the Cox model. Research report N 314, Department of Statistical Science, University College London.

TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV, VOLODYMYRSKA ST. 64, 01601, KYIV, UKRAINE.  
*E-mail address:* alexander\_kukush@univ.kiev.ua

TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV, VOLODYMYRSKA ST. 64, 01601, KYIV, UKRAINE.  
*E-mail address:* chernovaoksan@gmail.com