

TESTING LINEAR AND NONLINEAR HYPOTHESES IN A COX PROPORTIONAL HAZARDS MODEL WITH ERRORS IN COVARIATES

Oksana Chernova, Alexander Kukush

Taras Shevchenko National University of Kyiv, Mechanics and Mathematics Faculty.
Address: 64 Volodymyrska Street, Kyiv, 01033, Ukraine.

E-mail: chernovaoksan@gmail.com, alexander_kukush@univ.kiev.ua

Received: October 2019 Revised: November 2019 Published: December 2019

Abstract. We investigate linear and nonlinear hypotheses testing in a Cox proportional hazards model for right-censored survival data when the covariates are subject to measurement errors. In Kukush and Chernova (2018) [Theor. Probability and Math. Statist. 96, 101–110], a consistent simultaneous estimator is introduced for the baseline hazard rate and the vector of regression parameters. Therein the baseline hazard rate belongs to an unbounded set of nonnegative Lipschitz functions, with fixed constant, and the vector of regression parameters belongs to a compact parameter set. Based on the estimator, we develop two procedures to test nonlinear and linear hypotheses about the vector of regression parameters: Wald-type and score-type tests. The latter is based on an unbiased estimating equation. The consistency of the tests is shown.

Key words : Cox proportional hazards model, hypothesis testing, right censoring, simultaneous estimator of baseline hazard rate and regression parameter.

1. Introduction

Survival analysis (or failure time data analysis) means statistical inference for data, where the response of interest is time T from the time origin to the occurrence of a certain event. It is widely used, for example, in medicine, insurance and reliability. Key elements of survival analysis are nonnegative response and censoring. The aim of survival analysis is to estimate some aspects of the unavailable complete data from those observed, which are incomplete due to censoring. In medicine, the objective could be to compare different treatment effects on the survival time, possibly correcting for information available on each patient. This leads to the following regression analysis problem.

Let nonnegative continuous random variable T denote the survival time (the so-called lifetime). Consider the Cox proportional hazard model [8], where the lifetime T for a subject with covariate $\mathbf{X} \in \mathbb{R}^k$ has the following hazard (intensity) function

$$\lambda^*(t|\mathbf{X}; \lambda, \beta) = \lambda(t) \exp(\beta^\top \mathbf{X}), \quad t \geq 0. \quad (1)$$

An unknown parameter β belongs to the given parameter set $\Theta_\beta \subset \mathbb{R}^k$, and $\lambda(\cdot)$ is an unknown nonnegative baseline hazard function.

Let nonnegative random variable C denote the censoring time. The observed survival data comprise the censored lifetime $Y := \min\{T, C\}$ and the non-censoring indicator $\Delta := I_{\{T \leq C\}}$. Assume that the censor C is distributed on a given interval $[0, \tau]$ and its survival function $G_C(u) = 1 - F_C(u)$ is unknown. Since censored lifetimes belong to $[0, \tau]$, we will estimate the baseline hazard only on this interval. Moreover, we assume that $\lambda(\cdot) \in \Theta_\lambda \subset C[0, \tau]$. The conditional pdf of T given \mathbf{X} is

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

Consider an additive error model, i.e., a surrogate variable

$$\mathbf{W} = \mathbf{X} + \mathbf{U}, \quad (2)$$

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

Consider independent copies of the model $(\mathbf{X}_i, T_i, C_i, Y_i, \Delta_i, \mathbf{U}_i, \mathbf{W}_i)$, $i = 1, 2, \dots, n$. Based on triples $(Y_i, \Delta_i, \mathbf{W}_i)$, $i = 1, \dots, n$, we estimate the parameters $\boldsymbol{\beta}$ and $\lambda(t)$, $t \in [0, \tau]$.

In the seminal paper of Cox [8], baseline hazard $\lambda(\cdot)$ is unspecified, covariates are observed without measurement errors, and the parameter $\boldsymbol{\beta}$ is estimated by maximization of a partial likelihood function that does not involve $\lambda(\cdot)$. This paper engendered a lively discussion. Breslow [4] proposed an estimator of cumulative baseline hazard $\Lambda(t) = \int_0^t \lambda(s) ds$, and the estimator relies on the estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. Ramlau-Hansen [13] derived the estimator for the baseline hazard itself from the Breslow estimator using kernel smoothing and studied its asymptotic properties. In all of these papers the dimension of $\boldsymbol{\beta}$ is small compared to n . Andersen and Gill [1] studied the large sample properties of a counting processes model with intensity given in (1).

The Cox proportional hazards model with measurement errors has been studied in papers, including the following. Nakamura (1992) [12], based on an approximately corrected score function (only using the first and second terms of its Taylor expansion), introduced heuristically an estimator of the regression parameter in the Cox proportional hazards model where measurement errors are additive and normally distributed. The results of numerical simulations were presented as well. Gu & Kong (1999) [10] showed that the latter estimator is consistent and asymptotically normal. They also proved that the corrected cumulative baseline hazard estimator is consistent and converges to a Gaussian process. Augustin (2004) [2] noticed that the estimator in Gu & Kong (1999) [10] is an exact corrected log-likelihood estimator. Wang (2006) [14] considered Cox proportional hazards regression with longitudinal covariates. In those papers the baseline hazard rate was either unspecified or assumed piece-wise constant.

Kukush and Chernova [11] constructed a simultaneous estimator $(\hat{\lambda}(\cdot), \hat{\boldsymbol{\beta}})$ of $(\lambda(\cdot), \boldsymbol{\beta})$ for the case where the baseline hazard rate $\lambda(\cdot)$ belongs to an unbounded set of nonnegative Lipschitz functions, with fixed Lipschitz constant, and the regression parameter $\boldsymbol{\beta}$ is from a compact parameter set. Right-censored lifetimes are observed, and covariates are corrupted by additive measurement errors. The estimator is strongly consistent and asymptotically normal. In Chernova [6], procedures for testing simple hypotheses are presented for the regression parameter and integral functionals of the baseline hazard rate. More precisely, hypotheses have the following form: (a) “ $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\lambda \in \Theta_\lambda$ ” vs. “ $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$, $\lambda \in \Theta_\lambda$ ”, or (b) “ $\int_0^{\tau-\varepsilon} \lambda(u) \vec{f}(u) du = \vec{c}_0$, $\boldsymbol{\beta} \in \Theta_\beta$ ” vs. “ $\int_0^{\tau-\varepsilon} \lambda(u) \vec{f}(u) du \neq \vec{c}_0$, $\boldsymbol{\beta} \in \Theta_\beta$ ”, where $0 < \varepsilon < \tau$, \vec{f} is some vector function, and \vec{c}_0 is a fixed vector.

In the present paper, we discuss a Wald-type test for the general nonlinear hypothesis “ $h(\boldsymbol{\beta}) = 0$, $\lambda \in \Theta_\lambda$ ” vs. “ $h(\boldsymbol{\beta}) \neq 0$, $\lambda \in \Theta_\lambda$ ”, where a Borel function $h : \Theta_\beta \rightarrow \mathbb{R}^r$, $r \leq k$, is continuously differentiable in a neighborhood of $\boldsymbol{\beta}$, whose Jacobian matrix has full rank, and a score-type test for the linear hypothesis “ $V\boldsymbol{\beta} = v$, $\lambda \in \Theta_\lambda$ ” vs. “ $V\boldsymbol{\beta} \neq v$, $\lambda \in \Theta_\lambda$ ”, where V is an $r \times k$ matrix of full rank. We refer to Chapter 8 of Boos and Stefanski [3] for discussion of Wald-type and score-type tests; the latter is based on an unbiased estimation equation.

This paper is organized as follows. In Section 2 we define the simultaneous estimator for the model parameters $(\lambda, \boldsymbol{\beta})$ and describe properties of the estimator. In Section 3 we discuss a Wald-type test for the nonlinear hypothesis and prove its consistency. In Section 4 a score-type test for linear hypothesis is presented. We show that testing a linear hypothesis is equivalent to testing a hypothesis about a sub-vector of the regression parameter, and the consistency of the test is shown. Conclusions are given in Section 5.

2. Simultaneous estimator of the baseline hazard function and regression parameter

As in [2], we use the objective function corrected for measurement errors

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

with

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

We impose the following conditions:

- (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 $L > 0$ is a fixed constant.
- (ii) $\Theta_\beta \subset \mathbb{R}^k$ is a compact set.
- (iii) $E \mathbf{U} = 0$ and for some constant $\varepsilon > 0$,

$$E e^{D\|\mathbf{U}\|} < \infty, \text{ with } D := \max_{\beta \in \Theta_\beta} \|\beta\| + \varepsilon.$$

- (iv) $E e^{D\|\mathbf{X}\|} < \infty$, with D defined in (iii).
- (v) τ is the right endpoint of the censor's distribution, i.e., $P(C > \tau) = 0$ and for all $\varepsilon > 0$, $P(C > \tau - \varepsilon) > 0$.
- (vi) Matrix $E \mathbf{X}\mathbf{X}^\top$ is positive definite.
- (vii) True parameter β is an inner point of Θ_β .
- (viii) True parameter $\lambda \in \Theta_\lambda^\varepsilon$ for some unknown $\varepsilon > 0$, where

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

- (ix) Censor C has a continuous distribution function.
- (ix') $P(C > 0) = 1$.

$$(x) \text{ For all } \beta \in \Theta_\beta \text{ and all } R > 0, \sum_{k=0}^{\infty} \frac{a_{k+1}(\beta)}{k!} R^k < \infty \text{ with } a_{k+1}(\beta) := \frac{E\|\mathbf{U}\|^2 e^{(k+1)\beta^\top \mathbf{U}}}{M_{\mathbf{U}}((k+1)\beta)}.$$

In [7], it is shown that condition (x) holds true if, for example, the centered measurement error \mathbf{U} is either bounded, or normal, or has a shifted Poisson distribution.

Definition 1. Fix a sequence $\{\varepsilon_n\}$ of positive numbers, with $\varepsilon_n \downarrow 0$, as $n \rightarrow \infty$. Any 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 Θ and such that

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

is called a corrected estimator of (λ, β) .

According to [11], under conditions (i)–(viii) the simultaneous estimator is strongly consistent, i.e.,

$$\max_{t \in [0, \tau]} |\hat{\lambda}(t) - \lambda(t)| \rightarrow 0 \text{ and } \hat{\beta} \rightarrow \beta$$

a.s. as $n \rightarrow \infty$.

Denote by $G_T(t|\mathbf{X})$ the conditional survival function of T given \mathbf{X} , let also

$$a(t) = E[\mathbf{X} e^{\beta^\top \mathbf{X}} G_T(t|\mathbf{X})], \quad b(t) = E[e^{\beta^\top \mathbf{X}} G_T(t|\mathbf{X})],$$

$$p(t) = E[\mathbf{X}\mathbf{X}^\top e^{\beta^\top \mathbf{X}} G_T(t|\mathbf{X})], \quad T(t) = p(t)b(t) - a(t)a^\top(t), \quad K(t) = \frac{\lambda(t)}{b(t)},$$

$$A = \mathbb{E} \left[\mathbf{X} \mathbf{X}^\top e^{\beta^\top \mathbf{X}} \int_0^Y \lambda(u) du \right], \quad M = \int_0^\tau T(u) K(u) G_C(u) du.$$

For $i \geq 1$, introduce random vectors

$$\zeta_i = -\frac{\Delta_i a(Y_i)}{b(Y_i)} + \frac{\exp(\beta^\top \mathbf{W}_i)}{M_{\mathbf{U}}(\beta)} \int_0^{Y_i} a(u) K(u) du + \frac{\partial q}{\partial \beta}(Y_i, \Delta_i, \mathbf{W}_i; \lambda, \beta),$$

with the column vector of partial derivatives

$$\frac{\partial q}{\partial \beta}(Y, \Delta, \mathbf{W}; \lambda, \beta) = \Delta \cdot \mathbf{W} - \frac{M_{\mathbf{U}}(\beta) \mathbf{W} - \mathbb{E}(\mathbf{U} e^{\beta^\top \mathbf{U}})}{M_{\mathbf{U}}(\beta)^2} \exp(\beta^\top \mathbf{W}) \int_0^Y \lambda(u) du.$$

Denote also

$$\Sigma_\beta = 4 \cdot \text{Cov}(\zeta_1), \quad \Sigma = \Sigma(\lambda, \beta) = M^{-1} \Sigma_\beta M^{-1}, \quad m(\varphi_\lambda) = \int_0^\tau \varphi_\lambda(u) a(u) G_C(u) du,$$

$$\sigma_\varphi^2 = 4 \cdot \text{Var} \langle q'(Y, \Delta, \mathbf{W}; \lambda, \beta), \varphi \rangle = 4 \cdot \text{Var} \xi(Y, \Delta, \mathbf{W}),$$

$$\begin{aligned} \text{where } \xi(Y, \Delta, \mathbf{W}) &= \frac{\Delta \cdot \varphi_\lambda(Y)}{\lambda(Y)} - \frac{\exp(\beta^\top \mathbf{W})}{M_{\mathbf{U}}(\beta)} \int_0^Y \varphi_\lambda(u) du + \Delta \cdot \varphi_\beta^\top \mathbf{W} \\ &\quad - \varphi_\beta^\top \frac{M_{\mathbf{U}}(\beta) \mathbf{W} - \mathbb{E}[\mathbf{U} e^{\beta^\top \mathbf{U}}]}{M_{\mathbf{U}}(\beta)^2} \exp(\beta^\top \mathbf{W}) \int_0^Y \lambda(u) du. \end{aligned}$$

Here $\varphi = (\varphi_\lambda, \varphi_\beta) \in C[0, \tau] \times \mathbb{R}^k$ and q' is the Fréchet derivative of q with respect to the pair (λ, β) .

Theorem 2 ([11]). *Assume assumptions (i)–(viii) and (ix'). Then the matrix Σ introduced above is nonsingular and*

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N_k(0, \Sigma).$$

Moreover,

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

for all Lipschitz functions f defined on the interval $[0, \tau]$, where $\sigma_\varphi^2(f) = \sigma_\varphi^2$, $\varphi = (\varphi_\lambda, \varphi_\beta)$, $\varphi_\beta = -A^{-1}m(\varphi_\lambda)$ and φ_λ is a unique solution in $C[0, \tau]$ to the Fredholm integral equation

$$\frac{\varphi_\lambda(u)}{K(u)} - a^\top(u) A^{-1} m(\varphi_\lambda) = f(u), \quad u \in [0, \tau].$$

3. Wald-type test for a nonlinear hypothesis

In this section, we explain how to test the composite null hypothesis about β in the model (1)–(2)

$$\mathbf{H}_0: \quad “h(\beta) = 0, \quad \lambda \in \Theta_\lambda” \quad \text{vs.} \quad \mathbf{H}_1: \quad “h(\beta) \neq 0, \quad \lambda \in \Theta_\lambda”,$$

where $h: \Theta_\beta \rightarrow \mathbb{R}^r$, $r \leq k$, is continuously differentiable in a neighborhood of β Borel $r \times k$ vector function with $r \times k$ Jacobian matrix $H(\beta) = \frac{\partial h(\beta)}{\partial \beta^\top}$ of full rank in the neighborhood. We assume that β is an unknown inner point of Θ_β .

This form of hypothesis contains the following important particular cases: if $h(\beta) = \beta - \beta_0$, then the simple hypothesis is tested, and if $h(\beta) = V\beta - v$ with a fixed $r \times k$ matrix V and a fixed vector v , then we test a linear hypothesis about vector β .

Asymptotic properties of the estimator $(\hat{\lambda}, \hat{\beta})$ were presented in Section 2. Moreover, it is shown in [7] that under conditions (i)–(x), a strongly consistent estimator $\hat{\Sigma}$ (being a symmetric random

matrix itself) for the asymptotic covariance matrix Σ of the estimator $\hat{\beta}$ can be constructed. Let $\hat{H} = H(\hat{\beta})$. Define the test statistic

$$T_{n1} = n \cdot \left\| \left(\hat{H} \hat{\Sigma} \hat{H}^\top \right)^{-1/2} h(\hat{\beta}) \right\|^2,$$

for all $n \geq 1$ and ω from the underlying probability space Ω such that $\hat{\Sigma}$ is positive definite; otherwise we put $T_{n1} = 0$.

Lemma 3. *Assume conditions (i)–(x). Then under \mathbf{H}_0 it holds that $T_{n1} \xrightarrow{d} \chi_r^2$ as $n \rightarrow \infty$.*

Proof. Using the asymptotic normality of $\hat{\beta}$ stated in Theorem 2 and the Delta Theorem (p.14 in [3]) one can obtain the following:

$$\sqrt{n} (h(\hat{\beta}) - h(\beta)) \xrightarrow{d} N_r(0, H(\beta) \Sigma(\lambda, \beta) H^\top(\beta)).$$

Under \mathbf{H}_0 , $h(\beta) = 0$. The symmetric matrix $\hat{\Sigma}$ converges a.s. to the positive definite matrix $\Sigma(\lambda, \beta)$. Thus, $\hat{\Sigma}$ is positive definite *eventually*, i.e., with probability 1 for all n starting from some random integer. Since the function H is continuous in a neighborhood of β , it holds that $H(\hat{\beta}) \rightarrow H(\beta)$ almost surely. Then Slutsky's lemma yields

$$\sqrt{n} \left(\hat{H} \hat{\Sigma} \hat{H}^\top \right)^{-1/2} h(\hat{\beta}) \xrightarrow{d} N_r(0, I),$$

which completes the proof. \square

Given $\alpha \in (0, 1)$, denote the upper α -quantile of χ_r^2 distribution by $\chi_{r\alpha}^2$. Lemma 3 implies the next statement.

Theorem 4. *Under conditions (i)–(x), the test that fails to reject \mathbf{H}_0 if $T_{n1} \leq \chi_{r\alpha}^2$ and rejects the null hypothesis if $T_{n1} > \chi_{r\alpha}^2$, has asymptotic significance level α .*

Now, we prove that the test is consistent. Assume that the alternative \mathbf{H}_1 holds true. Then $h(\beta) \neq 0$. Let λ be the true value of baseline hazard function. We get *eventually*:

$$\begin{aligned} T_{n1} &= \left\| \left(\hat{H} \hat{\Sigma} \hat{H}^\top \right)^{-1/2} \sqrt{n} (h(\hat{\beta}) - h(\beta)) + \left(\hat{H} \hat{\Sigma} \hat{H}^\top \right)^{-1/2} \sqrt{n} h(\beta) \right\|^2 = \\ &= n \cdot \left\| \left(H(\beta) \Sigma(\lambda, \beta) H^\top(\beta) \right)^{-1/2} h(\beta) \right\|^2 + n \cdot o_p(1), \end{aligned}$$

since $(\hat{H} \hat{\Sigma} \hat{H}^\top)^{-1/2}$ converges in probability to $(H(\beta) \Sigma(\lambda, \beta) H^\top(\beta))^{-1/2}$ and the sequence $\sqrt{n} (h(\hat{\beta}) - h(\beta))$ converges in distribution and hence is bounded in probability. Thus, under \mathbf{H}_1 the test statistic converges to infinity in probability. The rate of convergence is *const* · n with

$$\text{const} = \left\| \left(H(\beta) \Sigma(\lambda, \beta) H^\top(\beta) \right)^{-1/2} h(\beta) \right\|^2, \quad \text{const} > 0.$$

Therefore, the test is consistent, i.e. the probability of a Type II error tends to 0 as $n \rightarrow \infty$.

4. Score-type test for a linear hypothesis

As an alternative to the Wald test one can use a score-type test based on the unbiased score function

$$s(Y, \Delta, \mathbf{W}; \lambda, \beta) := \frac{\partial q}{\partial \beta^\top}(Y, \Delta, \mathbf{W}; \lambda, \beta) = \Delta \cdot \mathbf{W} - g(\mathbf{W}, \beta) \int_0^Y \lambda(u) du,$$

where

$$g(\mathbf{W}, \beta) := \frac{\mathbf{W} \cdot M_{\mathbf{U}}(\beta) - E[\mathbf{U} e^{\beta^\top \mathbf{U}}]}{M_{\mathbf{U}}^2(\beta)} e^{\beta^\top \mathbf{W}}.$$

We start with a hypothesis about a sub-vector of the regression parameter. Let $\beta^\top = (\theta_1^\top, \theta_2^\top)$, where θ_1 and θ_2 are column vectors of size $(k-r)$ and r , respectively, $1 \leq r \leq k-1$. We partition the score function as follows

$$s(Y, \Delta, W; \lambda, \beta) = \begin{pmatrix} s_1(Y, \Delta, W; \lambda, \beta) \\ s_2(Y, \Delta, W; \lambda, \beta) \end{pmatrix}.$$

Any $r \times r$ matrix can be partitioned accordingly, in particular

$$\Sigma = \begin{pmatrix} (k-r) & (r) \\ \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} (k-r) \\ (r) \end{pmatrix}, \quad A = \begin{pmatrix} (k-r) & (r) \\ A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} (k-r) \\ (r) \end{pmatrix}.$$

4.1. Hypothesis about sub-vector of regression parameter

We test the null hypothesis about the true value of the partitioned regression parameter β in the model (1)–(2)

$$\mathbf{H}_0: \quad “\theta_2 = \theta_{20}, \quad \lambda \in \Theta_\lambda” \quad \text{vs.} \quad \mathbf{H}_1: \quad “\theta_2 \neq \theta_{20}, \quad \lambda \in \Theta_\lambda”,$$

where θ_{20} is an inner point of the projection of Θ_β on the last r coordinates and $\beta = (\theta_1^\top, \theta_2^\top)^\top$ is an inner point of Θ_β with unknown first component θ_1 .

Let $Z = (Y, \Delta, \mathbf{W})$, $Z_i = (Y_i, \Delta_i, \mathbf{W}_i)$, $\hat{\beta}^\top = (\hat{\theta}_{1n}^\top, \hat{\theta}_{2n}^\top)$. In [7], a strongly consistent estimator $\hat{A} = \hat{A}(\hat{\lambda}, \hat{\beta}) = \hat{A}(\hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n})$ of the positive definite matrix A from Section 2 is constructed; the estimator is a symmetric random matrix. Define the test statistic

$$T_{n2} = \left\| \left(\hat{A}_{22} \hat{\Sigma}_{22} \hat{A}_{22} \right)^{-1/2} \frac{1}{\sqrt{n}} \sum_1^n s_2(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20}) \right\|^2,$$

for all $n \geq 1$ and $\omega \in \Omega$ such that $\hat{\Sigma}_{22}$ and \hat{A}_{22} are positive definite; otherwise we put $T_{n2} = 0$.

Lemma 5. Assume conditions (i)–(x). Then under \mathbf{H}_0 it holds that $T_{n2} \xrightarrow{d} \chi_r^2$ as $n \rightarrow \infty$.

Proof. The proof follows the line of the proof of Lemma 4 in [6]. Condition (vii) and the strong consistency of $\hat{\beta}$ ensure that

$$\sum_1^n s(Z_i; \hat{\lambda}, \hat{\beta}) = 0 \quad \text{eventually.}$$

Denote

$$G_2(\mathbf{W}) = \max_{\beta \in \text{conv}(\Theta_\beta)} \left\| g_{22}''(\mathbf{W}, \beta) \right\|,$$

where "conv" stands for the convex hull of a set. Expand $s_2(Z; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n})$ in a neighborhood of θ_{20} . Due to the Theorem about finite increments of vector valued functions [5]:

$$s_2(Z; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}) = s_2(Z; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20}) + \frac{\partial s_2}{\partial \theta_2^\top}(Z; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20})(\hat{\theta}_{2n} - \theta_{20}) + r(Z; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20}),$$

with

$$\left\| r(Z; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20}) \right\| \leq \max_{\beta \in \text{conv}(\Theta_\beta)} \left\| s_{22}''(Z; \hat{\lambda}, \beta) \right\| \cdot \|\hat{\theta}_{2n} - \theta_{20}\|^2 \leq G_2(\mathbf{W}) \cdot \|\hat{\lambda}\| \cdot \|\hat{\theta}_{2n} - \theta_{20}\|^2.$$

We have

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_1^n s_2(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20}) &= -\frac{1}{\sqrt{n}} \sum_1^n \frac{\partial s_2}{\partial \theta_2^\top}(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20})(\hat{\theta}_{2n} - \theta_{20}) - \\ &- \frac{1}{\sqrt{n}} \sum_1^n r(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20}) = \frac{1}{\sqrt{n}} \sum_1^n g'_{22}(W_i; \hat{\theta}_{1n}, \theta_{20})(\hat{\theta}_{2n} - \theta_{20}) \int_0^{Y_i} \lambda(u) du - \\ &- \frac{1}{\sqrt{n}} \sum_1^n g'_{22}(W_i; \hat{\theta}_{1n}, \theta_{20})(\hat{\theta}_{2n} - \theta_{20}) \int_0^{Y_i} (\hat{\lambda} - \lambda)(u) du - \frac{1}{\sqrt{n}} \sum_1^n r(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20}), \end{aligned}$$

with

$$\left\| \frac{1}{\sqrt{n}} \sum_1^n r \left(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20} \right) \right\| \leq \frac{1}{n} \sum_1^n G_2(\mathbf{W}_i) \cdot \|\hat{\lambda}\| \cdot \sqrt{n} \|\hat{\theta}_{2n} - \theta_{20}\|^2.$$

The random variables $\frac{1}{n} \sum_{i=1}^n G_2(\mathbf{W}_i)$ are bounded in probability because they converge to $E G_2(\mathbf{W})$ a.s. as $n \rightarrow \infty$. According to [11], $\|\hat{\lambda}\|$ is bounded. Applying the Delta Theorem one can obtain the following convergence

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \xrightarrow{d} N_r(0, \Sigma_{22}),$$

hence $\sqrt{n}(\hat{\theta}_{2n} - \theta_{20})$ is bounded in probability. Finally, the strong consistency of $\hat{\theta}_{2n}$ implies

$$\frac{1}{\sqrt{n}} \sum_1^n r \left(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \hat{\theta}_{2n}, \theta_{20} \right) = o_p(1).$$

Similarly,

$$\frac{1}{\sqrt{n}} \sum_1^n g'_{22}(\mathbf{W}_i; \hat{\theta}_{1n}, \theta_{20})(\hat{\theta}_{2n} - \theta_{20}) \int_0^{Y_i} (\hat{\lambda} - \lambda)(u) du = o_p(1).$$

For any fixed $z = (z_1^\top, z_2^\top)^\top \in \Theta_\beta$, we have the almost sure convergence

$$\frac{1}{n} \sum_1^n g'_{22}(\mathbf{W}_i; z) \int_0^{Y_i} \lambda(u) du \rightarrow E_\beta \left[\left(\mathbf{X}\mathbf{X}^\top \right)_{22} e^{z^\top \mathbf{X}} \int_0^Y \lambda(u) du \right] = A_{22}(\lambda, \beta; z_1, z_2).$$

By the Arzelà-Ascoli theorem, the random sequence on the left-hand side is equicontinuous in z almost surely on the compact set Θ_β , and the limit function is continuous. These facts imply the uniform convergence in z a.s. Thus,

$$\frac{1}{n} \sum_1^n g'_{22}(\mathbf{W}_i; \hat{\theta}_{1n}, \theta_{20}) \int_0^{Y_i} \lambda(u) du \rightarrow A_{22} := A_{22}(\lambda, \beta; \theta_1, \theta_{20}) \quad \text{a.s.}$$

By Slutsky's Lemma

$$\frac{1}{n} \sum_1^n g'_{22}(\mathbf{W}_i; \hat{\theta}_{1n}, \theta_{20}) \int_0^{Y_i} \lambda(u) du \sqrt{n}(\hat{\theta}_{2n} - \theta_{20}) \xrightarrow{d} N_r(0, A_{22}\Sigma_{22}A_{22}).$$

Under \mathbf{H}_0 ,

$$\hat{A}_{22} \rightarrow A_{22}(\lambda, \theta_1, \theta_{20}) \quad \text{a.s.} \quad \text{and} \quad \hat{\Sigma}_{22} \rightarrow \Sigma_{22}(\lambda, \theta_1, \theta_{20}) \quad \text{a.s.}$$

Matrices Σ and A are positive definite, therefore, $\hat{\Sigma}$ and \hat{A} are positive definite *eventually*. Sylvester's criterion guarantees that minors \hat{A}_{22} and $\hat{\Sigma}_{22}$ are positive definite *eventually* as well.

This yields that under \mathbf{H}_0 , $T_{n2} \xrightarrow{d} \chi_r^2$ as $n \rightarrow \infty$. □

Theorem 6. *Under conditions (i)–(x), the test that fails to reject \mathbf{H}_0 if $T_{n2} \leq \chi_{r\alpha}^2$ and rejects the null hypothesis if $T_{n2} > \chi_{r\alpha}^2$, has asymptotic significance level α .*

Under \mathbf{H}_1 , $\hat{\beta}$ is an estimator of the true parameter β , with $\theta_2 = \theta_{21} \neq \theta_{20}$, where θ_{20} and θ_{21} are inner points of the projection of Θ_β on the last r coordinates and $\lambda \in \Theta_\lambda$ is the baseline hazard function.

Similarly to Lemma 7 from [6] one can prove the following.

Lemma 7. *Assume conditions (i)–(x). Under \mathbf{H}_1 ,*

$$E_{H_1, s_2}(Z; \lambda, \theta_1, \theta_{20}) = E \left[X_2 \left(1 - e^{(\theta_{20} - \theta_{21})^\top X_2} \right) \int_0^\tau f(u|X; \lambda, \beta) G_C(u) du \right] =: K_2 = K_2(\lambda, \beta, \theta_{20}).$$

Notice that due to condition (iv), the function K_2 is well defined if $\|\theta_{20} - \theta_{21}\| < D$. Next, under \mathbf{H}_1 by Lemma 7 we have almost surely as $n \rightarrow \infty$:

$$\frac{1}{n} \sum_1^n s_2 \left(Z_i; \hat{\lambda}, \hat{\theta}_{1n}, \theta_{20} \right) \rightarrow K_2, \quad \hat{A}_{22} \rightarrow A_{22}(\lambda, \beta), \quad \hat{\Sigma}_{22} \rightarrow \Sigma_{22}(\lambda, \beta).$$

Thus,

$$\frac{1}{n} T_{n2} \rightarrow \left\| \left(A_{22}(\lambda, \beta) \Sigma_{22}(\lambda, \beta) A_{22}(\lambda, \beta) \right)^{-1/2} K_2 \right\|^2 \quad \text{a.s.}$$

Hence under \mathbf{H}_1 , it holds that

$$T_{n2} = n \cdot \left\| \left(A_{22}(\lambda, \beta) \Sigma_{22}(\lambda, \beta) A_{22}(\lambda, \beta) \right)^{-1/2} K_2 \right\|^2 + n \cdot o_P(1).$$

If $K_2 \neq 0$ then the test is consistent. We show that given the true values of β and λ , the set $\{\theta_{20} \in B(\theta_{21}, D) : K_2(\lambda, \beta, \theta_{20}) = 0\}$ is nowhere dense.

First, assume that $K_2 = K_2(\theta_{20})$ is identical zero at the open ball $B(\theta_{21}, D)$. Then its Jacobian matrix is singular. However, condition (vi) guarantees that

$$\left. \frac{\partial K_2}{\partial \theta_{20}^\top} \right|_{\theta_{20} = \theta_{21}} = -E \left[X_2 X_2^\top \int_0^\tau f(u|X; \lambda, \beta) G_C(u) du \right]$$

is a negative definite matrix. This leads to a contradiction. Therefore, K_2 is not identical zero.

Second, assume that K_2 equals zero at a certain open ball that belongs to $B(\theta_{21}, D)$. Notice that in a natural way the function $\|K_2\|^2$ can be extended to an analytic function in the complex vector variable $\theta_{20} \in B_D := \{z \in \mathbb{C}^r : \|z - \theta_{21}\| < D\}$. Theorem 9.4.4 from [9] implies that $\|K_2\|^2$ is identically zero at the domain B_D , which contradicts the statement proved above.

Third, the set of zeros of K_2 is closed in the ball $B(\theta_{21}, D)$. Then the statement proved at the 2nd step implies, that given λ and β , the set of zeros of K_2 is indeed nowhere dense. Hence the consistency of the test can be violated only for exceptional values of θ_{20} .

4.2. Testing a linear hypothesis

Now, consider a linear hypothesis

$$\mathbf{H}_0 : \quad "V\beta = v, \lambda \in \Theta_\lambda" \quad \text{versus} \quad \mathbf{H}_1 : \quad "V\beta \neq v, \lambda \in \Theta_\lambda",$$

where V is an $r \times k$ full-rank matrix and β is an unknown inner point of Θ_β .

Without loss of generality we may and do assume that the rows of the matrix V are orthonormal. Denote them as e_{k-r+1}, \dots, e_k . One can complement them to a basis $\{e_i\}_{i=1}^k$, and any $\beta \in \Theta_\beta$ can be decomposed as follows

$$\beta = \sum_{i=1}^{k-r} (\beta, e_i) e_i + \sum_{i=k-r+1}^k (\beta, e_i) e_i.$$

Here, $(\beta, e_{k-r+1}) = v_1, \dots, (\beta, e_k) = v_r$. Therefore, the problem of testing the linear hypothesis reduces to testing the hypothesis about a sub-vector

$$\mathbf{H}_0 : \quad "\theta_2 = v, \lambda \in \Theta_\lambda" \quad \text{vs.} \quad \mathbf{H}_1 : \quad "\theta_2 \neq v, \lambda \in \Theta_\lambda",$$

which was studied in the previous section.

5. Conclusions

We have extended the results of [6] where simple hypotheses were tested, and presented two procedures for testing composite hypotheses about the regression parameter β using the simultaneous consistent estimator $(\hat{\lambda}, \hat{\beta})$ from [7]. To test linear hypotheses, one can use either the score-type test or the Wald-type test, while for nonlinear hypotheses only the latter is applicable. Although the consistency of the tests is shown, the question concerning the comparison of their power functions in case of linear hypotheses remains open.

References

- [1] Andersen P. K., Gill R. D., 1982. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics* 10(4): p. 1100–1120.
- [2] Augustin T., 2004. An exact corrected log-likelihood function for Cox's proportional hazard model under measurement error and some extensions. *Scandinavian Journal of Statistics* 31(1): p. 43–50.
- [3] Boos D. D., Stefanski L. A., 2013. *Essential Statistical Inference Theory and methods*. Springer, New York.
- [4] Breslow N., 1972. Contribution to the discussion of the paper by D. R. Cox. *J. Roy. Statist. Soc., Series B* 34: p. 187–220.
- [5] Cartan H., 1971. *Differential Calculus*, Hermann/Houghton Mifflin Co., Paris/Boston, MA.
- [6] Chernova O., 2019. Hypothesis testing in Cox proportional hazard model with measurement error. *Teor. Imovir. Mat. Stat.* 100: p. 198–207 (in Ukrainian).
- [7] Chernova O., Kukush A., 2018. Confidence regions in Cox proportional hazard model with measurement errors and unbounded parameter set. *Mod. Stoch. Theory Appl.* 5(1): p. 37–52.
- [8] Cox D.R., 1972. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* 34: p. 187–220.
- [9] Dieudonné J., 1969. *Foundations of Modern Analysis*, enlarged and corrected printing. Academic Press, New York and London.
- [10] Gu M., Kong F. H., 1999. Consistent estimation in Cox proportional hazard model with covariate measurement errors. *Statistica Sinica* 32(9): p. 953–969.
- [11] Kukush A., Chernova O., 2018. Consistent estimation in Cox proportional hazard model with measurement errors and unbounded parameter set. *Theor. Probability and Math. Statist.* 96: p. 101–110.
- [12] Nakamura T., 1992. Proportional hazards models with covariates subject to measurement error. *Biometrics* 48: p. 829–838.
- [13] Ramlau-Hansen H., 1983. Smoothing counting process intensities by means of kernel functions. *Ann. Statist.* 11(2): p. 453–466.
- [14] Wang C. Y., 2006. Corrected score estimator for joint modeling of longitudinal and failure time data. *Statistica Sinica*, 16: p. 235–253.

TIESINIŲ IR NETIESINIŲ HIPOTEZIŲ TIKRINIMAS KOKSO PROPORCINGŲJŲ INTENSYVUMŲ MODELyje SU PAKLAIDOMIS KOVARIANTĖSE

Oksana Chernova, Aleksander Kukush

Santrauka. Straipsnyje sprendžiamas tiesinių ir netiesinių hipotezių tikrinimo uždavinys Kokso proporcingųjų intensyvumų modelyje cenzūruotiems iš dešinės duomenims, kai kovariantės matuojamos su paklaidomis. Kukush ir Chernova (2018) [Theor. Probability and Math. Statist. 96, 101–110] pasiūlė pagrįstą jungtinį pradinio intensyvumo ir regresijos parametrų vektoriaus įvertinį. Buvo tariama, kad pradinis intensyvumas priklauso neapibrėžtai aibei neneigiamų Lipschitzo funkcijų su fiksuota konstanta, o regresijos parametrų vektorius priklauso kompaktiškai parametrų aibei. Naudojantis minėtu įvertiniu sukonstruojami du testai netiesinėms ir tiesinėms hipotezėms apie regresijos parameterus tikrinti: Waldo tipo testas ir informantės tipo testas. Pastarasis remiasi nepaslinktąja įvertinimo lygtimi. Įrodomas testų pagrįstumas.

Reikšminiai žodžiai: Kokso proporcingųjų intensyvumų modelis, hipotezių tikrinimas, cenzūravimas iš dešinės, jungtinis pradinio intensyvumo ir regresijos parametrų įvertinys.