Testing proportional hazards for specified covariates

Vilijandas Bagdonavičius*, Rūta Levulienė

Faculty of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius LT-03225, Lithuania

vilijandas.bagdonavicius@mif.vu.lt (V. Bagdonavičius), ruta.levuliene@mif.vu.lt (R. Levulienė)

Abstract   Tests for proportional hazards assumption concerning specified covariates or groups of covariates are proposed. The class of alternatives is wide: log-hazard rates under different values of covariates may cross, approach, go away. The data may be right censored. The limit distribution of the test statistic is derived. Power of the test against approaching alternatives is given. Real data examples are considered.

1   Introduction

The most known semi-parametric model for analysis of failure time regression data is the proportional hazards (PH) model. There are many tests for the PH model from right censored data given by Cox [5], Moreau et al. [11], Lin [10], Nagelkerke et al. [12], Grambsch and Therneau [6], Quantin et al. [13], Bagdonavičius et al. [4]. Tests for specified covariates are given in Kvaloy and Neef [9], Kraus [8].

We consider tests for proportional hazards assumption concerning specified covariates or groups of covariates. The class of alternatives is wide: log-hazard rates under different values of covariates may cross, approach, go away. The data may be right censored. The limit distribution of the test statistics is derived. Power of the test under approaching alternatives is given. Real examples are considered.

---

*Corresponding author.

VTeX

2 Modeling non-proportional hazards for specified covariates

Let $S(t|z)$, $\lambda(t|z)$ and $\Lambda(t|z)$ be the survival, hazard rate and cumulative hazard functions under a m-dimensional explanatory variable (or covariate) $z = (z_1, \ldots, z_m)^T$.

Let us consider the proportional hazards (PH) model:

$$\lambda(t|z) = e^{\beta^T z} \lambda(t),$$

where $\lambda(t)$ is unknown baseline hazard function, the parameter $\beta = (\beta_1, \ldots, \beta_m)^T$ is m-dimensional.

Under the PH model the ratios of hazard functions under any two different explanatory variables are constant over time, i.e. they are proportional.

Suppose that proportionality of hazard functions with respect to a specified covariate z_j or a group of covariates $z_{j_1}, \ldots, z_{j_k}$ may be violated. Our purpose is to detect such violations. So we seek a wider model which includes not only possibility of constant hazard rates ratios but also time-varying ratios.

We propose a model of the form

$$\lambda(t|z) = g\big(z, \Lambda(t), \beta, \gamma_j\big)\lambda(t), \quad \Lambda(t) = \int_0^t \lambda(u)du, \tag{1}$$

where

$$g(z, \Lambda(t), \beta, \gamma_j) = \frac{e^{\beta^T z + \Lambda(t)e^{\gamma_j z_j}}}{1 + e^{\gamma_j z_j}[e^{\Lambda(t)e^{\gamma_j z_j}} - 1]}.$$

$\beta = (\beta_1, \ldots, \beta_m)^T$ and γ_j are unknown regression parameters, $\lambda(t)$ and $\Lambda(t)$ are unknown baseline hazard and cumulative hazard, respectively.

The PH model is a particular case of this model when $\gamma_j = 0$. The model (1) is very wide as compared to the PH model.

Indeed, suppose that two different values $z^{(1)}$ and $z^{(2)}$ of the covariate vector differ only by the value of the covariate z_j and denote by $c(t)$ the ratio of hazard rates. For the model (1)

$$c(0) = e^{\beta_j(z_j^{(2)} - z_j^{(1)})}, \qquad c(\infty) = e^{(\beta_j - \gamma_j)(z_j^{(2)} - z_j^{(1)})}.$$

So this model gives a large choice of alternatives: the ratio of hazard rates $c(t)$ can vary from any $a > 0$ to any $b > 0$, whereas $c(t)$ is constant under the PH model. The difference of logarithms of hazard rates under different values of the covariate z_j is constant under the PH model, so the logarithms of hazard rates as time functions are parallel, whereas in dependence on the values of its parameters the model (1) includes also the possibilities for these functions to approach, to go away, to intersect. So the model (1) may help to detect non-proportionality of hazard rates (or, equivalently, non-parallelism of log-hazard functions) in above mentioned directions. Other directions are very rare in real data.

We do not discuss application of the model for analysis of survival regression data (which is a subject of another article). Such analysis could be done if the PH model would be rejected. Here the model is used only as a generator of a wide class of alternatives to the PH model.

3 Test statistic

Let us consider right censored failure time regression data:

$$(X_1, \delta_1, z^{(1)}), \ldots, (X_n, \delta_n, z^{(n)}),$$

where

$$X_i = T_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{T_i \leq C_i\}},$$

T_i are failure times and C_i are censoring times. X_i is observation time of the ith unit, the event $\delta_i = 1$ indicates that X_i is the failure time T_i, and the event $\delta_i = 0$ indicates that X_i is the censoring time C_i.

Set

$$N_i(t) = \mathbf{1}_{\{X_i \leq t, \delta_i = 1\}}, \quad Y_i(t) = \mathbf{1}_{\{X_i \geq t\}},$$

$$N(t) = \sum_{i=1}^n N_i(t), \quad Y(t) = \sum_{i=1}^n Y_i(t);$$

here $\mathbf{1}_A$ denotes the indicator of the event A. The processes $N_i(t)$ and $N(t)$ are right-continuous counting process showing the numbers of observed failures in the time interval $[0, t]$ for the ith unit and for all n units, respectively. $Y_i(t)$ and $Y(t)$ are decreasing (not strongly) left-continuous stochastic processes showing the numbers of units which are still not-failed and not-censored just prior to t for the ith and for all units, respectively.

Suppose that the survival distribution of the i object given $z^{(i)}$ is absolutely continuous with the survival functions $S_i(t)$ and the hazard rates $\lambda_i(t)$.

Suppose that the multiplicative intensities model is verified: the compensators of the counting processes N_i with respect to the history of the observed processes are $\int Y_i \lambda_i du$. It is equivalent to the assumption that for any i and t: $P(X_i > t) > 0$ with almost all $s \in [0, t]$

$$\lim_{h \downarrow 0} \frac{1}{h} P\{T_i \in [s, s+h] | X_i \geq s, x^{(i)}\}$$

$$= \lim_{h \downarrow 0} \frac{1}{h} P\{T_i \in [s, s+h] | T_i \geq s, x^{(i)}\} = \lambda_i(s).$$

It means that for almost all $s \in [0, t]$ the risk to fail just after the time s given that units were non-censored and did not fail to time s is equal to the risk to fail just after the time s given that units did not fail to time s when censoring does not exist. So censoring has no influence to the risk of failure.

Information about time-to-failure distribution contains the points, where the counting processes N_i have jumps. A jump at the point t is possible if $Y_i(t) = 1$. Partial information give the points where the counting processes N_i have not jumps but the predictable processes have jumps.

The multiplicative intensities model is verified in the case of type I, type II, independent random censoring.

In the parametric case with known λ the unknown finite-dimensional parameter consists of the parameters β and γ_j. The component of the parametric score function U corresponding to γ_j has the form

$$U_j(\beta, \gamma_j) = \sum_{i=1}^{n} \int_0^\infty \left\{ w_j^{(i)}(u, \Lambda, \beta, \gamma_j) - \tilde{E}_j(u, \Lambda, \beta, \gamma_j) \right\} dN_i(u), \qquad (2)$$

where

$$w_j^{(i)}(v, \Lambda, \beta, \gamma_j) = \frac{\partial}{\partial \gamma_j} \log\{ g(z^{(i)}, \Lambda(v), \beta, \gamma_j) \},$$

$$\tilde{E}(v, \Lambda, \beta, \gamma_j) = \frac{\tilde{S}^{(1)}(v, \Lambda, \beta, \gamma_j)}{\tilde{S}^{(0)}(v, \Lambda, \beta, \gamma_j)},$$

$$\tilde{S}^{(0)}(v, \Lambda, \beta, \gamma_j) = \sum_{i=1}^{n} Y_i(v) g\{ z^{(i)}, \Lambda(v), \beta, \gamma_j \},$$

$$\tilde{S}^{(1)}(v, \Lambda, \beta, \gamma_j) = \sum_{i=1}^{n} Y_i(v) \frac{\partial}{\partial \gamma_j} g\{ z^{(i)}, \Lambda(v), \beta, \gamma_j \}.$$

We consider semiparametric case when the baseline hazard rate λ is unknown. The test statistic is constructed in the following way. In the expression of U_j the parameter β is replaced by its partial likelihood estimator $\hat{\beta}$ under the Cox model, the parameter γ_j is replaced by 0, and the baseline cumulative intensity Λ is replaced by the Breslow estimator (see Andersen et al. [2])

$$\hat{\Lambda}(t) = \int_0^t \frac{dN(u)}{S^{(0)}(u, \hat{\beta})}.$$

The estimator $\hat{\beta}$ verifies the equation

$$\sum_{i=1}^{n} \int_0^\infty \left\{ z^{(i)} - E(t, \hat{\beta}) \right\} dN_i(t) = 0, \qquad (3)$$

where

$$E(t, \beta) = \frac{S^{(1)}(t, \beta)}{S^{(0)}(t, \beta)},$$

$$S^{(0)}(t, \beta) = \sum_{i=1}^{n} Y_i(t) e^{\beta^T z^{(i)}}, \qquad S^{(1)}(t, \beta) = \sum_{i=1}^{n} z^{(i)} Y_i(t) e^{\beta^T z^{(i)}}.$$

Set $\hat{F}(t) = 1 - e^{-\hat{\Lambda}(t)}$. It is an estimator of the baseline cumulative distribution function. After the above mentioned replacement the following statistic is obtained:

$$\hat{U}_{\gamma_j} = -\sum_{i=1}^{n} \int_0^\infty \hat{F}(t) \left\{ z_j^{(i)} - E_j(t, \hat{\beta}) \right\} dN_i(t), \qquad (4)$$

where E_j is the component of E corresponding to the covariate z_j.

Let us consider asymptotic distribution of the statistic (4) under the PH model. Set

$$S^{(2)}(u, \beta) = \sum_{i=1}^{n} (z^{(i)})^{\otimes 2} Y_i(u) e^{\beta^T z^{(i)}},$$

$$V(u, \beta) = S^{(2)}(u, \beta)/S^{(0)}(u, \beta) - E^{\otimes 2}(u, \beta),$$

where $A^{\otimes 2} = AA^T$ for any matrix A. Denote by $V_j(u, \beta)$ the jth column and by $V_{jj}(u, \beta)$ the jth diagonal element of the matrix $V(u, \beta)$. Set

$$\hat{\Sigma}_{jj}(t) = n^{-1} \int_0^t \hat{F}(u) V_{jj}(u, \hat{\beta}) dN(u), \qquad \hat{\Sigma}_j(t) = n^{-1} \int_0^t \hat{F}(u) V_j(u, \hat{\beta}) dN(u),$$

$$\hat{\Sigma}(t) = n^{-1} \int_0^t V(u, \hat{\beta}) dN(u).$$

Assumptions A.   a) $(X_i, \delta_i, z^{(i)})$ are i.i.d.,

b) $\sup\{t : Y_i(t) > 0\} \xrightarrow{P} \tau > 0$,

c) the covariates $z^{(i)}$ are bounded: $\|z^{(i)}\| \le C, C > 0$,

d) $\Lambda(\tau) < \infty$,

e) the matrix

$$\Sigma(\beta) = \int_0^\tau v(u, \beta) s^{(0)}(u, \beta) d\Lambda(u)$$

is positively definite; here

$$s^{(i)}(t, \beta) = E S^{(i)}(t, \beta)/n, \qquad e(u, \beta) = s^{(1)}(u, \beta)/s^{(0)}(u, \beta) \quad i = 0, 1, 2,$$

$$v(u, \beta) = s^{(2)}(u, \beta)/s^{(0)}(u, \beta) - e(u, \beta)(e(u, \beta))^T.$$

Assumption a) can be weakened considerably but we avoid writing complicated formulas for easier reading. Assumption b) simply means that at some finite time moment (perhaps very remote) observations are stopped. This is a usual assumption for asymptotic results to hold in survival analysis. Assumption c) also can be weakened. Assumption d) means that if censoring would be absent then units might survive after the moment τ with positive probability. Assumption e) is needed to have non-degenerated asymptotic distribution of the test statistic. This assumption is the usual assumption needed for asymptotic normality of the maximum partial likelihood estimator $\hat{\beta}$ of the regression parameter β.

Under Assumptions A [3] there exists a neighborhood Θ of β such that

$$\sup_{b \in \Theta, u \in [0, \tau]} \left\| \frac{1}{n} S^{(i)}(u, b) - s^{(i)}(u, b) \right\| \xrightarrow{P} 0,$$

$$\sup_{b \in \Theta, u \in [0, \tau]} \| E(u, b) - e(u, b) \| \xrightarrow{P} 0,$$

$$\sup_{b\in\Theta,u\in[0,\tau]}\left\|V(u,b)-v(u,b)\right\|\overset{P}{\rightarrow}0, \tag{5}$$

as $n\rightarrow\infty$.

Assumption A a) may be weakened assuming that there exist non-random $s^{(i)}$ such that the convergences (5) hold.

Convergences (5) and Assumption A d) imply that

$$\sup_{b\in\Theta,t\in[0,\tau]}\left\|\int_0^t(\frac{1}{n}S^{(i)}(u,b)-s^{(i)}(u,b))d\Lambda(u)\right\|=o_P(1),$$

$$\sup_{b\in\Theta,t\in[0,\tau]}\left\|\int_0^t(E(u,b)-e(u,b))d\Lambda(u)\right\|=o_P(1),$$

$$\sup_{b\in\Theta,t\in[0,\tau]}\left\|\int_0^t(V(u,b)-v(u,b))d\Lambda(u)\right\|=o_P(1).$$

Theorem 1. *Under Assumptions A the following weak convergence holds:*

$$T=n^{-1/2}\hat{U}_{\gamma_j}/\sqrt{\hat{D}_j}\overset{d}{\rightarrow}N(0,1),\quad as\ n\rightarrow\infty, \tag{6}$$

where

$$\hat{D}_j=\hat{\Sigma}_{jj}(\tau)-\hat{\Sigma}_j^T(\tau)\hat{\Sigma}^{-1}(\tau)\hat{\Sigma}_j(\tau). \tag{7}$$

Proof. Under the Cox model,

$$N_i(t)=\int_0^t Y_i(u)e^{\beta^T z^{(i)}}d\Lambda(u)+M_i(t),$$

where M_i are martingales with respect to the history generated by the data. So using it and taking into account that

$$\sum_{i=1}^n z_j^{(i)}Y_i(u)e^{\hat{\beta}^T z^{(i)}}-E_j(u,\hat{\beta})\sum_{i=1}^n Y_i(u)e^{\hat{\beta}^T z^{(i)}}=0,$$

the statistic $\hat{U}_{\gamma_j}$ can be written in the form

$$\hat{U}_{\gamma_j}=\sum_{i=1}^n\int_0^t\hat{F}(u)\{z_j^{(i)}-E_j(u,\hat{\beta})\}Y_i(u)\left(e^{\hat{\beta}^T z^{(i)}}-e^{\beta^T z^{(i)}}\right)d\Lambda(u)$$

$$-\sum_{i=1}^n\int_0^t\hat{F}(u)\{z_j^{(i)}-E_j(u,\hat{\beta})\}dM_i(u).$$

Assumptions A a)–d) and consistence of the partial likelihood estimator $\hat{\beta}$ imply that

$$n^{-1/2}\hat{U}_{\gamma_j}(t)=n^{-1}\sum_{i=1}^n\int_0^t\hat{F}(u)\{z_j^{(i)}$$

$$-E_j(u,\hat{\beta})\}\left(z^{(i)}\right)^T Y_i(u)e^{\hat{\beta}^T z^{(i)}}d\Lambda(u)\,n^{1/2}(\hat{\beta}-\beta)$$

$$- n^{-1/2} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\} dM_i(u) + o_P(1)$$

uniformly on $[0, \tau]$. Using convergences (5) and consistence of the estimator $\hat{\beta}$ we write the first integral in the form

$$n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}(z^{(i)})^T Y_i(u) e^{\hat{\beta}^T z^{(i)}} d\hat{\Lambda}(u) + o_P(1)$$

$$= n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(t, \hat{\beta})\}(z^{(i)})^T Y_i(u) e^{\hat{\beta}^T z^{(i)}} \frac{dN(u)}{S^{(0)}(u, \hat{\beta})} + o_P(1)$$

$$= \hat{\Sigma}_j^T(\tau) + o_P(1).$$

Under Assumptions A (now Assumption A e) is crucial) the following expression holds:

$$n^{1/2}(\hat{\beta} - \beta) = \hat{\Sigma}^{-1}(\tau) n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \{z^{(i)} - E(u, \hat{\beta})\} dM_i(u) + o_p(1)$$

uniformly on $[0, \tau]$ (see Andersen et al. [2], Theorem VII.2.2, where even weaker assumptions instead of a)–d) are used). It implies

$$n^{-1/2} \hat{U}_{\gamma_j}(t) = \hat{\Sigma}_j^T(t) \hat{\Sigma}^{-1}(\tau) n^{-1/2} \sum_{i=1}^{n} \int_0^\tau \{z^{(i)} - E(u, \hat{\beta})\} dM_i(u)$$

$$- n^{-1/2} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(t, \hat{\beta})\} dM_i(u) + o_p(1)$$

$$=: M^*(t) + o_p(1)$$

uniformly on $[0, \tau]$.

The predictable variation of the local martingale $M^*(t)$ is

$$\langle M^* \rangle(t)$$

$$= \hat{\Sigma}_j^T(t) \hat{\Sigma}^{-1}(\tau) n^{-1} \sum_{i=1}^{n} \int_0^t \{z^{(i)} - E(u, \hat{\beta})\}^{\otimes 2} Y_i(u) e^{\hat{\beta}^T z^{(i)}} d\Lambda(u) \hat{\Sigma}^{-1}(\tau) \hat{\Sigma}_j(t)$$

$$- \hat{\Sigma}_j^T(t) \hat{\Sigma}^{-1}(\tau) n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}\{z^{(i)} - E(u, \hat{\beta})\}$$

$$\times Y_i(u) e^{\hat{\beta}^T z^{(i)}} d\Lambda(u)$$

$$- n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}\{z^{(i)} - E(u, \hat{\beta})\}^T$$

$$\times Y_i(u) e^{\hat{\beta}^T z^{(i)}} d\Lambda(u) \hat{\Sigma}^{-1}(\tau) \hat{\Sigma}_j(t)$$

$$+ n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}^2(u) \{z_j^{(i)} - E_j(u, \hat{\beta})\}^2 Y_i(u) e^{\beta^T z^{(i)}} d\Lambda(u).$$

Note that

$$n^{-1} \sum_{i=1}^{n} \int_0^t \{z^{(i)} - E(u, \hat{\beta})\}^{\otimes 2} Y_i(u) e^{\beta^T z^{(i)}} d\Lambda(u) = \hat{\Sigma}(t) + o_P(1),$$

$$n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}(u) \{z_j^{(i)} - E_j(u, \hat{\beta})\} \{z^{(i)} - E(u, \hat{\beta})\} Y_i(u) e^{\beta^T z^{(i)}} d\Lambda(u)$$
$$= \hat{\Sigma}_j(t) + o_P(1),$$

$$n^{-1} \sum_{i=1}^{n} \int_0^t \hat{F}^2(u) \{z_j^{(i)} - E_j(u, \hat{\beta})\}^2 Y_i(u) e^{\beta^T z^{(i)}} d\Lambda(u) = \hat{\Sigma}_{jj}(t) + o_P(1).$$

So the predictable variation

$$\langle M^* \rangle(t) = \hat{\Sigma}_{jj}(t) - \hat{\Sigma}_j^T(t) \hat{\Sigma}^{-1}(\tau) \hat{\Sigma}_j(t) + o_P(1) \overset{P}{\to} \Sigma_{jj}(t) - \Sigma_j^T(t) \Sigma^{-1}(\tau) \Sigma_j(t)$$

uniformly on $[0, \tau]$; here $\Sigma_{jj}(t)$, $\Sigma_j(t)$, $\Sigma(\tau)$ are the limits in probability of the random matrices $\hat{\Sigma}_{jj}(t)$, $\hat{\Sigma}_j(t)$, $\hat{\Sigma}(\tau)$.

Under Assumptions A for any $\varepsilon > 0$ the predictable variation (see Andersen et al. [2]), Theorem VII.2.2, where even weaker assumptions are used)

$$\left\langle n^{-1/2} \sum_{i=1}^{n} \int_0^t \{z_j^{(i)} - E_j(t, \beta)\} \mathbf{1}_{\{|z_j^{(i)} - E_j(u,\beta)| > \sqrt{n}\varepsilon\}} dM_i(u) \right\rangle \overset{P}{\to} 0$$

and β can be replaced by $\hat{\beta}$ in the expression of the left side because $\hat{\beta}$ is consistent. Similarly

$$\left\langle n^{-1/2} \sum_{i=1}^{n} \int_0^t \hat{F}(u) \{z_j^{(i)} - E_j(t, \hat{\beta})\} \mathbf{1}_{\{\hat{F}(u)|z_j^{(i)} - E_j(u,\hat{\beta})| > \sqrt{n}\varepsilon\}} dM_i(u) \right\rangle \overset{P}{\to} 0.$$

Hence, the Lindeberg conditions of the central limit theorem for martingales (see Andersen et al. [2]) are verified for M^*. We saw that the predictable variations $\langle M^* \rangle(t)$ converge in probability to non-random non-degenerated matrices. So the stochastic process $n^{-1/2} \hat{U}_{\gamma_j}(\cdot)$ converges in distribution to a zero mean Gaussian process on $[0, \tau]$, in particular

$$T = n^{-1/2} \hat{U}_{\gamma_j} / \sqrt{\hat{D}} \overset{\mathcal{D}}{\to} N(0, 1) \quad \text{as } n \to \infty,$$

here

$$\hat{D} = \hat{\Sigma}_{jj}(\tau) - \hat{\Sigma}_j^T(\tau) \hat{\Sigma}^{-1}(\tau) \hat{\Sigma}_j(\tau). \qquad \square$$

The test: the null hypothesis $H_0 : \gamma_j = 0$ is rejected with approximate significance level α if $|T| > z_{\alpha/2}$; here $z_{\alpha/2}$ is the $\alpha/2$ critical value of the standard normal distribution.

4 Power under approaching alternatives

Let us consider the alternative model (1) and suppose that alternatives are approaching: $\gamma_j = \frac{c}{\sqrt{n}}$, where $c \neq 0$ is a fixed constant.

So the model

$$\lambda(t|z) = \frac{\exp\{\beta^T z + \Lambda(t)\exp\{\frac{c}{\sqrt{n}}z_j\}\}}{1 + \exp\{\frac{c}{\sqrt{n}}z_j\}[\exp\{\Lambda(t)\exp\{\frac{c}{\sqrt{n}}z_j\}\} - 1]}\lambda(t), \quad \Lambda(t) = \int_0^t \lambda(u)du, \tag{8}$$

is considered. Denote by β_0 the true value of β under the model (8). The counting process $N_i(t)$ has the form

$$N_i(t) = \int_0^t g\{z^{(i)}, \Lambda(u), \beta_0, c/\sqrt{n}\}d\Lambda(u) + M_i(t),$$

where M_i are local martingales with respect to the history generated by the data. Note that

$$g\{z^{(i)}, \Lambda(u), \beta_0, c/\sqrt{n}\} = e^{\beta^T z^{(i)}}\left(1 - z_j^{(i)}F(u)\frac{c}{\sqrt{n}} + o_P\left(\frac{1}{\sqrt{n}}\right)\right).$$

Let us consider the stochastic process

$$Q_n(t, \beta) = \frac{1}{n}[\ell(t, \beta) - \ell(t, \beta_0)],$$

where

$$\ell(t, \beta) = \sum_{i=1}^n \int_0^t \{\beta^T z^{(i)} - \ln S^{(0)}(u, \beta)\}\, dN_i(u).$$

Note that the derivative with respect to β is

$$\dot{Q}_n(t, \beta) = \frac{1}{n}\dot{\ell}(t, \beta) = \frac{1}{n}\sum_{i=1}^n \int_0^t \{z^{(i)} - E(u, \beta)\}dN_i(u).$$

Constructing the test statistic we defined $\hat{\beta}$ as a random vector satisfying the estimating equation $\dot{\ell}(\tau, \beta) = 0$. Let us show that $\hat{\beta}$ is a consistent estimator of β_0.

Theorem 2. *If Assumptions A are satisfied then the probability that the equation $\dot{\ell}(\beta) = 0$ has a unique solution, converges to 1 and $\hat{\beta} \xrightarrow{P} \beta_0$.*

Proof. Fix $\beta \in \mathbf{R}^m$. Using the Doob–Meier decomposition of $N_i(t)$ write the stochastic process Q_n in the form

$$Q_n(t, \beta) = \frac{1}{n}\sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^T z^{(i)} - \ln \frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right]dN_i(u)$$

$$= \frac{1}{n}\sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^T z^{(i)} - \ln \frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right]$$

$$\times Y_i(u)e^{\beta_0^T z^{(i)}}\left(1 - \frac{c}{\sqrt{n}}F(u)z_j^{(i)} + o_P\left(\frac{1}{\sqrt{n}}\right)\right)d\Lambda(u)$$

$$+ \frac{1}{n}\sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^T z^{(i)} - \ln\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right]dM_i(u)$$

$$= \int_0^t \left[(\beta - \beta_0)^T S^{(1)}(u, \beta_0) - \ln\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}S^{(0)}(u, \beta_0)\right]d\Lambda(u)$$

$$+ o_P(1) + \frac{1}{n}\sum_{i=1}^n \int_0^t \left[(\beta - \beta_0)^T z^{(i)} - \ln\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right]dM_i(u)$$

$$= K_n(t, \beta) + \tilde{M}_n(t) + o_P(1).$$

The predictable variation of the martingale $\tilde{M}_n$ is

$$\langle \tilde{M}_n\rangle(t) = \frac{1}{n^2}\sum_{i=1}^n \int_0^t \left\{(\beta - \beta_0)^T z^{(i)} - \ln\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right\}^2 Y_i(u)e^{\beta_0^T z^{(i)}}$$

$$\times \left(1 - \frac{c}{\sqrt{n}}F(u)z_j^{(i)} + o_P\left(\frac{1}{\sqrt{n}}\right)\right)d\Lambda(u)$$

$$= \frac{1}{n^2}\int_0^t \left\{(\beta - \beta_0)^T S^{(2)}(u, \beta_0)(\beta - \beta_0)\right.$$

$$- 2(\beta - \beta_0)^T S^{(1)}(u, \beta_0)\ln\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}$$

$$\left.+ S^{(0)}(u, \beta_0)\ln^2\frac{S^{(0)}(u, \beta)}{S^{(0)}(u, \beta_0)}\right\}d\Lambda(u) + o_p\left(\frac{1}{n}\right).$$

So $\langle \tilde{M}_n\rangle(\tau) \xrightarrow{P} 0$. $K_n(\tau, \beta) \xrightarrow{P} k(\beta)$; here

$$k(\beta) = \int_0^\tau \left[(\beta - \beta_0)^T s^{(1)}(u, \beta_0) - \ln\frac{s^{(0)}(u, \beta)}{s^{(0)}(u, \beta_0)}s^{(0)}(u, \beta_0)\right]d\Lambda(u),$$

which implies

$$Q_n(\beta) = Q_n(\tau, \beta) \xrightarrow{P} k(\beta).$$

Note that

$$\dot{k}(\beta) = \int_0^\tau \left[s^{(1)}(u, \beta_0) - e(u, \beta)s^{(0)}(u, \beta_0)\right]d\Lambda(u), \quad \dot{k}(\beta_0) = 0.$$

$$\ddot{k}(\beta) = -\int_0^\tau v(u, \beta)s^{(0)}(u, \beta_0)\,d\Lambda(u), \quad \ddot{k}(\beta_0) = -\Sigma(\beta_0).$$

So the matrix $\ddot{k}(\beta_0)$ is negatively definite. The remaining part of the proof coincides with the proof of analogous theorem for the Cox model (see Theorem VII.2.1 of Andersen et al.), i.e. Andersen's and Gill's theorem [3] is applied. This theorem says that if the sequence of concave differentiable stochastic processes $Q_n(\beta)$ pointwise converges in probability to a real function $k(\beta)$ on a convex open set $E \subset \mathbf{R}^m$, then:

a) the function $k(\beta)$ is concave on E; b) the convergence is uniform in probability on compact subsets of the set E; c) if the function $k(\beta)$ has a unique maximum at the point β_0 then the probability that the equation $\dot{Q}(\beta) = 0$ has a unique root $\hat{\beta}$ in the set E tends to 1 and $\hat{\beta} \xrightarrow{P} \beta_0$.

Assumption A e) was crucial for application of Andersen's and Gill's theorem. $\square$

Note that $\hat{\Lambda}$ is uniformly consistent estimator of Λ on $[0, \tau]$:

$$
\hat{\Lambda}(t) = \int_0^t \frac{dN(u)}{S^{(0)}(u, \hat{\beta})} = \Lambda(t) + \int_0^t \left[\frac{S^{(0)}(u, \beta_0)}{S^{(0)}(u, \hat{\beta})} - 1 \right] d\Lambda(u)
$$
$$
- \frac{c}{\sqrt{n}} \int_0^t \frac{S_j^{(1)}(u, \beta_0)}{S^{(0)}(u, \hat{\beta})} d\Lambda(u) + \int_0^t \frac{dM(u)}{S^{(0)}(u, \hat{\beta})} + o_p(1) = \Lambda(t) + o_p(1)
$$

uniformly on $[0, \tau]$ because $\hat{\beta}$ is consistent and

$$
\left\langle \int_0^\tau \frac{dM(u)}{S^{(0)}(u, \hat{\beta})} \right\rangle = \int_0^\tau \frac{S^{(0)}(u, \beta_0)}{(S^{(0)}(u, \hat{\beta}))^2} d\Lambda(u) = o_p(1).
$$

Set

$$
\mu_j(t) = -c \int_0^t v_j(u, \beta_0) s^{(0)}(u, \beta_0) F(u) d\Lambda(u).
$$

Theorem 3. *If Assumptions A are satisfied then*

1)
$$
\frac{1}{\sqrt{n}} \dot{\ell}(\cdot, \beta_0) \xrightarrow{D} \mu_j(t) + Z(\cdot, \beta_0) \text{ on } (D[0, \tau])^m,
$$

where Z is an m-dimensional Gaussian process with components having independent increments, $Z_j(0) = 0$ a.s. and for all $0 \le s \le t \le \tau$:

$$
\mathbf{cov}(Z_j(s), Z_{j'}(t)) = \sigma_{jj'}(s);
$$

here $\sigma_{jj'(t)}$ are the elements of the matrix $\Sigma(t)$. In particular,

$$
\frac{1}{\sqrt{n}} \dot{\ell}(\beta_0) \xrightarrow{D} \mu_j(\tau) + Z(\tau, \beta_0) \sim N(\mu_j(\tau), \Sigma(\tau)), \quad \text{as } n \to \infty.
$$

2) $\hat{\Sigma}(\tau) = -\frac{1}{n} \ddot{\ell}(\hat{\beta}) \xrightarrow{P} \Sigma(\tau)$,

3) $\sqrt{n}(\hat{\beta} - \beta_0) = (-\frac{1}{n} \ddot{\ell}(\hat{\beta}))^{-1} \frac{1}{\sqrt{n}} \dot{\ell}(\beta_0) + o_P(1)$.

Proof. Set

$$
M^*(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \{z^{(i)} - E(u, \beta_0)\} dM_i(u).
$$

Using the Doob–Meier decomposition we have

$$
\frac{1}{\sqrt{n}}\dot{\ell}(t,\beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \{z^{(i)} - E(u,\beta_0)\}dN_i(u)
$$

$$
= M^*(t) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \{z^{(i)} - E(u,\beta_0)\}Y_i(u)e^{\beta_0^T z^{(i)}}
$$

$$
\times \left(1 - \frac{c}{\sqrt{n}}F(u)z_j^{(i)} + o_P\left(\frac{1}{\sqrt{n}}\right)\right)d\Lambda(u)
$$

$$
= M^*(t) - \frac{c}{n}\sum_{i=1}^{n}\int_0^t V_j(u,\beta_0)S^{(0)}(u,\beta_0)F(u)d\Lambda(u) + o_P(1)
$$

$$
= M^*(t) - c\int_0^t v_j(u,\beta_0)s^{(0)}(u,\beta_0)F(u)d\Lambda(u) + o_P(1)
$$

$$
= M^*(t) + \mu_j(t) + o_P(1),
$$

where V_j is the jth column of the matrix V.

The first term converges weakly to $Z(\cdot,\beta_0)$ on $(D[0,\tau])^m$ because the limit in probability of its predictable covariation matrix has the same expression as that of analogous term under the Cox model:

$$
\langle M^*\rangle(t) = \frac{1}{n}\int_0^t V(u,\beta_0)S^{(0)}(u,\beta_0)\,d\Lambda(u) - \frac{c}{n^{3/2}}\sum_{i=1}^{n}\int_0^t \{z^{(i)} - E(u,\beta_0)\}
$$

$$
\times \{z^{(i)} - E(u,\beta_0)\}^T Y_i(u)\,e^{\beta_0^T z^{(i)}}\left(F(u)z_j^{(i)} + o_P(1)\right)d\Lambda(u)
$$

$$
= \frac{1}{n}\int_0^t V(u,\beta_0)S^{(0)}(u,\beta_0)\,d\Lambda(u) + o_P(1) \xrightarrow{P} \Sigma(t).
$$

Analogously, verification of the Lindeberg's condition is done by the same way as in the case of the Cox model.

Let us consider the norm of the difference:

$$
\left\| -\frac{1}{n}\ddot{\ell}(\hat{\beta}) - \Sigma(\tau) \right\| \leq \sup_{u\in[0,\tau]}\left|V(u,\hat{\beta}) - v(u,\hat{\beta})\right| + \sup_{u\in[0,\tau]}\left|v(u,\hat{\beta}) - v(u,\beta_0)\right|
$$

$$
+ \left\| \frac{1}{n}\int_0^\tau v(u,\beta_0)dM(u) \right\|
$$

$$
+ \left\| \int_0^\tau v(u,\beta_0)\left(\frac{1}{n}S^{(0)}(u,\beta_0) - s^{(0)}(u,\beta_0)\right)d\Lambda(u) \right\|
$$

$$
+ \left\| \frac{c}{n\sqrt{n}}\int_0^\tau F(u)\sum_{i=1}^{n}z_j^{(i)}e^{\beta_0^T z^{(i)}}(1 + o_p(1))d\Lambda(u) \right\|.
$$

Using the fact that the estimator $\hat{\beta}$ is consistent and that the first four terms have the same structure as analogous terms in the proof of Theorem VII.2.2 of Andersen et al., we have that the first four terms converge in probability to zero. Such convergence of

the last term is obvious because the last term multiplied by $\sqrt{n}$ converges to a finite limit in probability.

The mean value theorem and consistency of the estimator $\hat{\beta}$ imply

$$\dot{\ell}_j(\hat{\beta}) - \dot{\ell}_j(\beta_0) = \ddot{\ell}_j(\beta_j^*)(\hat{\beta} - \beta_0) = n\left(\frac{1}{n}\ddot{\ell}_j(\hat{\beta}) + o_P(1)\right)(\hat{\beta} - \beta_0);$$

here β_j^* is a point on the line segment joining the points $\hat{\beta}$ and β_0. Since $\dot{\ell}_j(\hat{\beta}) = 0$, we obtain

$$\sqrt{n}(\hat{\beta} - \beta_0) = \left(-\frac{1}{n}\ddot{\ell}(\hat{\beta}) + o_P(1)\right)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\beta_0) = \left(-\frac{1}{n}\ddot{\ell}(\hat{\beta})\right)^{-1}\frac{1}{\sqrt{n}}\dot{\ell}(\beta_0) + o_p(1).$$

$$\square$$

Set

$$d = c\left[\int_0^\tau F^2(u)v_{jj}(u, \beta_0)s^{(0)}(u, \beta_0)d\Lambda(u)\right.$$

$$\left. - \int_0^\tau F(u)v_j^T(u, \beta_0)d\Lambda(u)\Sigma^{-1}(\tau)\int_0^\tau F(u)v_j(u, \beta_0)s^{(0)}(u, \beta_0)d\Lambda(u)\right].$$

Note that for $m = 1$

$$d/c = \frac{\int F^2 v s^{(0)} d\Lambda \int v s^{(0)} d\Lambda - (\int F v s^{(0)} d\Lambda)^2}{\int_v s^{(0)} d\Lambda} > 0,$$

because F is not equal to 1 a.s. on $[0, \tau]$.

Theorem 4. *Under Assumptions A*

$$T \xrightarrow{d} \mu + N(0, 1) = N(\mu, 1),$$

where $\mu = d/D_j$ and D_j is the limit in probability of the random variable $\hat{D}_j$.

Proof. Set

$$\bar{M}(t) = n^{-1/2}\sum_{i=1}^n \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}dM_i(u).$$

Similarly as in the proof of Theorem 1 using the equality $S_j^{(1)} - E_j S^{(0)} = 0$, consistency of $\hat{\beta}$, Assumptions A, uniform consistency of $\hat{F}$ on $[0, \tau]$, we write the following expression:

$$n^{-1/2}\hat{U}_{c/\sqrt{n}}(t) = -n^{-1/2}\sum_{i=1}^n \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}dN_i(u)$$

$$= -\bar{M}(t) - n^{-1/2}\sum_{i=1}^n \int_0^t \hat{F}(u)\{z_j^{(i)} - E_j(u, \hat{\beta})\}e^{\beta_0^T z^{(i)}}$$

$$\times \left\{1 - z_j^{(i)} F(u) \frac{c}{\sqrt{n}} (1 + o_p(1))\right\} Y_i(u) d\Lambda(u)$$

$$= -\bar{M}(t) + \frac{c}{n} \int_0^t \hat{F}(u) F(u) V_{jj}(u, \beta_0) S^{(0)}(u, \beta_0) d\Lambda(u)$$

$$+ n^{-1} \int_0^t \hat{F}(u) V_j^T(u, \beta_0) S^{(0)}(u, \beta_0) d\Lambda(u) n^{1/2}(\hat{\beta} - \beta_0) + o_P(1)$$

uniformly on $[0, \tau]$. Applying Theorem 3 write the right side in the form

$$n^{-1/2} \hat{U}_{c/\sqrt{n}}(t) = c\left[\int_0^t F^2(u) v_{jj}(u, \beta_0) s^{(0)}(u, \beta_0) d\Lambda(u)\right.$$

$$- \int_0^t F(u) v_j^T(u, \beta_0) d\Lambda(u) \Sigma^{-1}(\tau)$$

$$\left.\times \int_0^\tau F(u) v_j(u, \beta_0) s^{(0)}(u, \beta_0) d\Lambda(u)\right]$$

$$- n^{-1/2} \sum_{i=1}^n \int_0^t \hat{F}(u) \{z_j^{(i)} - E_j(u, \hat{\beta})\} dM_i(u)$$

$$- \hat{\Sigma}_j^T(t) \hat{\Sigma}^{-1}(\tau) n^{-1/2} \sum_{i=1}^n \int_0^\tau \{z^{(i)} - E(u, \hat{\beta})\} dM_i(u) + o_p(1)$$

uniformly on $[0, \tau]$.

Note that the non-martingale part is d and the martingale part of the expression is exactly of the same form as in the case of the PH model and has the same limit distribution as in latter case. □

The power function of the test against approaching alternatives is

$$\beta = 2 - \Phi(z_{\frac{\alpha}{2}} - \mu) - \Phi(z_{\frac{\alpha}{2}} + \mu),$$

where Φ and $z_{\frac{\alpha}{2}}$ are the c.d.f. and the upper $\alpha/2$ critical value of the standard normal law, respectively. If $d \neq 0$ and c is large then μ is large and the power is near to 1.

5 Case of several covariates

The proportional hazards hypothesis for several covariates $z_{j_1}, \ldots, z_{j_k}$ (for all covariates $z_1, \ldots, z_m$, in particular) is tested similarly.

Set $\bar{z} = (z_{j_1}, \ldots, z_{j_k})^T$, $\gamma = (\gamma_{j_1}, \ldots, \gamma_{j_k})^T$. The term $\gamma_j z_j$ is replaced by $\gamma^T \bar{z}$ in the model (1).

Replacing $z_j^{(i)}$ by $\bar{z}^{(i)}$, $E_j(t, \hat{\beta})$ by $(E_{j_1}(t, \hat{\beta}), \ldots, E_{j_k}(t, \hat{\beta}))^T$ in the statistic $\hat{U}_{\gamma_j}$ we obtain a statistic denoted by $\hat{U}_\gamma$. The statistic T has the form (6), where the element $V_{jj}(u, \hat{\beta})$ is replaced by the $k \times k$ matrix $(V_{j_l, j_s}(u, \hat{\beta}))_{k \times k}$, the vector $V_j(u, \hat{\beta})$ is replaced by the $m \times k$ matrix $(V_{i, j_s}(u, \hat{\beta}))$, $i = 1, \ldots, m$, $s = 1, \ldots, k$ in the definitions of Σ_{jj}, Σ_j, and $\hat{U}_{\gamma_j}$ is replaced by $\hat{U}_\gamma$.

So $\hat{\Sigma}_{jj}(t)$, $\hat{\Sigma}_j(t)$, and $\hat{D}$ become $k \times k$, $m \times k$, and $k \times k$ matrices, respectively, and U_γ becomes a $k \times 1$ vector.

The test statistic

$$T = n^{-1} \hat{U}_\gamma^T \hat{D}^{-1} \hat{U}_\gamma \xrightarrow{\mathcal{D}} \chi^2(k) \quad \text{as } n \to \infty.$$

The hypothesis $H_0 : \gamma = 0$ is rejected with approximate significance level α if $T > \chi_\alpha^2(k)$; here $\chi_\alpha^2(k)$ is the α critical value of the standard chi-squared distribution with k degrees of freedom.

6 Real data analysis

Example 1 (Chemo-radio data, one-dimensional dichotomous covariate). Stablein and Koutrouvelis [14] studied the well-known two-sample data of the Gastrointestinal Tumor Study Group concerning effects of chemotherapy ($z = 0$) and chemotherapy plus radiotherapy ($z = 1$) on the survival times of gastric cancer patients.

The number of patients $n = 90$. The data are right-censored. The value of the test statistic T is 3.651, the p-value is 0.0003. The Cox model is rejected. The result is natural because the Kaplan–Meier estimators of the survival functions of two patient groups intersect.

Example 2 (Prisoners data, 7 covariates). The data are given in [1]. They consist of 432 male inmates who were released from Maryland state prisons in the early 1970s. These men were followed for 1 year after their release, and the dates of any arrests were recorded. Time is measured by the week of the first arrest after release. There were seven covariates: financial aid after release (FIN; 0 – no, 1 – yes), age in years at the time of release (AGE), race (RACE; 1 – black, 0 – otherwise), full-time work experience before incarceration (WEX; 1 – yes, 0 – no), marital status (MAR; 1 – was married at the time of release, 0 – otherwise), released on parole (PAR; 1 – released on parole, 0 – otherwise), convictions prior to incarceration (PRI).

The results of testing hypothesis for all covariates: the value of the test statistic T is 17.58, the p-value is 0.014. The assumption of the PH assumption is rejected.

The results for each covariate are given in Table 1. The PH assumption is rejected for AGE and WEXP covariates.

Table 1. Prisoners data. The values of test statistics and p-values

Covar.	FIN	AGE	RACE	WEX	MAR	PAR	PRI	Glob. test
Stat.	0.162	2.464	1.423	−2.033	−1.017	−0.222	0.672	17.58
p-val.	0.872	0.014	0.155	0.042	0.309	0.824	0.502	0.014

Example 3 (UIS dataset, 10 covariates). Let us consider right censored UIS data set given in [7].

UIS was a 5-year research project comprised of two concurrent randomized trials of residential treatment for drug abuse. The purpose of the study was to compare treatment programs of different planned durations designed to reduce drug abuse and to prevent high-risk HIV behavior. The UIS sought to determine whether alternative residential treatment approaches are variable in effectiveness and whether

efficacy depends on planned program duration. The time variable is time to return to drug use (measured from admission). The individuals who did not returned to drug use are right censored. We use the model with 10 covariates (which support PH assumption) given by Hosmer, Lemeshow and May (2008). The covariates are: AGE (years); Beck depression score (beckt; $0 - 54$); NDR1 $= ((NDR + 1)/10)^{(-1)}$; NDR2 $= ((NDR + 1)/10)^{(-1)} \log((NDR + 1)/10)$; drug use history at admission (IVHX_3; 1 – recent, 0 – never or previous); RACE (0 – white, 1 – non-white); treatment randomization assigment (TREAT; 0 – short, 1 – long); treatment site (SITE; 0 – A, 1 – B); interaction of age and treatment site (AGEXS); interaction of race and treatment site (RACEXS). The NDR denotes number of prior drug treatments $(0 - 40)$. Due to missing data in covariates, the model is based on 575 of the 628 observations.

The results of testing hypothesis for all covariates: the value of the test statistic T is 6.781, the p-value is 0.7460. The assumption of the PH assumption is not rejected.

The results for each covariate are given in Table 2.

Table 2. UIS dataset. The values of test statistics and p-values

Covariate	AGE	beckt	NDR1	NDR2	IVHX_3	RACE
Stat.	-0.061	1.085	0.182	0.118	0.912	-1.278
p-value	0.952	0.278	0.856	0.906	0.362	0.201

Covariate	TREAT	SITE	AGEXS	RACEXS	Global test
Stat.	0.792	1.016	-0.378	6.781	-0.107
p-value	0.429	0.309	0.705	0.746	0.915

The assumption of the PH assumption for individual covariates is not rejected.

Acknowledgement

References

[1] Allison, P.D.: Survival Analysis Using SAS: A Practical Guide. SAS Institute Inc., Cary, NC (2010)

[2] Andersen, P., Borgan, O., Gill, R., Keiding, N.: Statistical Models Based on Counting Processes. Springer, New York (1993) MR1198884. https://doi.org/10.1007/978-1-4612-4348-9

[3] Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: A large sample study. Ann. Stat. **10**, 1100–1120 (1982) MR0673646

[4] Bagdonavičius, V., Hafdi, M., Nikulin, M.: Analysis of survival data with cross-effects of survival functions. Biostatistics **5**, 415–425 (2004) MR2127114. https://doi.org/10.1016/j.crma.2005.01.019

[5] Cox, D.R.: Regression models and life tables (with discussion). J. R. Stat. Soc. B **34**, 187–220 (1972) MR0341758

[6] Grambsch, P., Therneau, T.M.: Proportional hazards tests and diagnostics based on weighted residuals. Biometrika **81**, 515–526 (1994) MR1311094. https://doi.org/10.1093/biomet/81.3.515

[7] Hosmer, D., Lemeshow, S., May, S.: Applied Survival Analysis: Regression Modeling of Time-to-Event Data. Wiley, Stowe (2008) MR2383788. https://doi.org/10.1002/9780470258019

[8] Kraus, D.: Identifying nonproportional covariates in the cox model. Commun. Stat., Theory Methods **37**, 617–625 (2008) MR2432302. https://doi.org/10.1080/03610920701669744

[9] Kvaløy, J.T., Neef, L.R.: Tests for the proportional intensity assumption based on the score process. Lifetime Data Anal. **10**, 139–157 (2004) MR2081718. https://doi.org/10.1023/B:LIDA.0000030200.61020.85

[10] Lin, D.Y.: Goodness of fit analysis for the cox regression model based on a class of parameter estimators. J. Am. Stat. Assoc. **86**, 725–728 (1991) MR1147097

[11] Moreau, T., O'Quigley, J., Mesbah, M.: A global goodness-of-fit statistic for the proportional hazards model. Biometrics **34**, 212–218 (1985) MR0827669. https://doi.org/10.2307/2347465

[12] Nagelkerke, N.J., Oosting, J., A., H.A.: A simple test for goodness of fit of cox's proportional hazards model. Biometrics **40**, 483–486 (1984)

[13] Quantin, C., Moreau, T., Asselain, B., Lellouch, J.: A regression model for testing the proportional hazards hypothesis. Biometrics **52**, 874–885 (1996) MR1411737. https://doi.org/10.2307/2533049

[14] Stablein, D.M., Koutrouvelis, I.A.: A two sample test sensitive to crossing hazards in uncensored and singly censored data. Biometrics **41**, 643–652 (1985) MR0816008. https://doi.org/10.2307/2531284