Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

**DIPLOMOVÁ PRÁCE**

Jakub Marčiny

# Ověřování předpokladů modelu proporcionálního rizika

Katedra pravděpodobnosti a matematické statistiky

| | |
|---|---|
| Vedoucí diplomové práce: | doc. Mgr. Michal Kulich, Ph.D. |
| Studijní program: | Matematika |
| Studijní obor: | Pravděpodobnost, matematická statistika a ekonometrie |

Praha 2014

Rád bych poděkoval svému vedoucímu doc. Mgr. Michalu Kulichovi, Ph.D. za jeho užitečné postřehy a vstřícný přístup. Dále bych rád poděkoval svým rodičům Hedvice a Zdeňkovi za jejich podporu během studia.

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Praze dne 9. 4. 2014 Jakub Marčiny

Charles University in Prague
Faculty of Mathematics and Physics

**MASTER THESIS**

Jakub Marčiny

# Verifying the Assumptions of the Proportional Hazards Model

Department of Probability and Mathematical Statistics

| | |
|---|---|
| Supervisor of the master thesis: | doc. Mgr. Michal Kulich, Ph.D. |
| Study programme: | Mathematics |
| Specialization: | Probability, Mathematical Statistics and Econometrics |

Prague 2014

I would like to thank to my supervisor doc. Mgr. Michal Kulich, Ph.D. for his useful remarks and kind approach. I would also like to thank to my parents Hedvika and Zdeněk for their support throughout my studies.

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Coll., the Copyright Act, as amended, in particular the fact that the Charles University in Prague has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 paragraph 1 of the Copyright Act.

In Prague on 9. 4. 2014 Jakub Marčiny

Název práce: Ověřování předpokladů modelu proporcionálního rizika

Autor: Jakub Marčiny

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D

Abstrakt: Coxův model proporcionálního rizika je standardním nástrojem pro modelování vlivu regresorů na dobu do události za přítomnosti cenzorování. Vhodnost použití tohoto modelu je však podmíněna platností předpokladu proporcionálního rizika. Tento předpoklad je v textu vysvětlen a jsou podrobně popsány metody pro jeho testování. Testy jsou implementovány v R, včetně vlastnoručně napsaného testu Lin-Zhang-Davidianové. Testy jsou dále ilustrovány na lékařských datech. Jejich schopnost odhalit porušení předpokladu proporcionálního rizika je podrobena zkoumání v simulační studii. Její výsledky naznačují, že nejvyšší síly dosahuje nejčastěji nově implementovaný test Lin-Zhang-Davidianové. Naopak bylo zjištěno, že vážená verze Lin-Wei-Yingova testu nedodržuje hladinu pro malé rozsahy výběru.

Klíčová slova: Coxův model proporcionálního rizika, testy předpokladu proporcionálního rizika, simulační studie

Title: Verifying the Assumptions of the Proportional Hazards Model

Author: Jakub Marčiny

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D

Abstract: The Cox proportional hazards model is a standard tool for modelling the effect of covariates on time to event in the presence of censoring. The appropriateness of this model is conditioned by the validity of the proportional hazards assumption. The assumption is explained in the thesis and methods for its testing are described in detail. The tests are implemented in R, including self-written version of the Lin-Zhang-Davidian test. Their application is illustrated on medical data. The ability of the tests to reveal the violation of the proportional hazards assumption is investigated in a simulation study. The results suggest that the highest power is attained by the newly implemented Lin-Zhang-Davidian test in most cases. In contrast, the weighted version of the Lin-Wei-Ying test was found to have inadequate size for low sample sizes.

Keywords: Cox proportional hazards model, tests of the proportional hazards assumption, simulation study

# Contents

# Introduction

The Cox proportional hazards model is the most widely used tool for modelling the effect of covariates on time to event in the presence of censoring. It is a very flexible model since no distribution has to be specified for the event times. However, the model is not universal. The proportional hazards assumption – the central topic of the thesis – has to be fulfilled for the model to be applicable. Whether the assumption is plausible should be subject to careful inspection.

The thesis aims to make a contribution in several aspects. First, an overview of the tests of the proportional hazards is given. Such overviews are available in the literature (Ng'andu, 1994; Therneau and Grambsch, 2000) but more recent developments including those of Lin et al. (2006) are included in the thesis. Second, we provide a solid introduction to the field of survival analysis in Chapter 1 establishing unified terminology and notation which allows to present the derivations of the tests by Grambsch and Therneau (1994) and Lin et al. (1993) in a more compact way than is available in the literature. Third, the above mentioned Lin-Zhang-Davidian test is newly implemented in R as it is not part of any standard library yet. Fourth, simulation study investigating the power and the size of the tests is conducted complementing the very recent results of Grant et al. (2013). Furthermore, the tests are applied to medical data which enables a more detailed comparison. The analysis is performed with R 2.14.2. Probability theory background is provided in the Appendix.

Strong emphasis is placed on the continuity of presentation. Some of the methods such as the Lin-Zhang-Davidian test are only described in scientific papers so far which focus on bringing new ideas rather than thorough explanation. Great deal of effort is made in preserving the consistency of notation which varies spectacularly in the literature. In attempt to make matters as clear as possible, each subscript is reserved for a single purpose: $i = 1, \dots, n$ for individuals, $j = 1, \dots, p$ for covariates and $k = 1 \dots, m$ for events. This enables to omit much of the omnipresent indexing without compromising in clarity of exposition. Random objects are denoted in upcase and vector objects are written in bold unless it contradicts an established notation.

# 1. Survival Analysis

Let $T_1, \ldots, T_n$ be independent non-negative continuous random variables denoting the event times and let $C_1, \ldots, C_n$ be independent non-negative random variables denoting the censoring times. Denote $T_i^\star = \min(T_i, C_i)$ time to either event or censoring – whichever occurs first – and $\Delta_i = \mathbb{1}(T_i \leq C_i)$ the event indicator. Let $t_1 < \ldots < t_m$ represent the observed event times and set $t_0 = 0$. Denote $\tau$ the time of the end of the study.

## 1.1 The Case of No Covariates

First suppose that the observed information is contained in independent identically distributed random vectors $(T_i^\star, \Delta_i)^\mathsf{T}$. There are several ways to describe the distribution of time to event such as the *survival function*

$$S(t) = 1 - F(t) \equiv \mathsf{P}(T_i > t),$$

the *hazard function*

$$\lambda(t) \equiv \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \leq T_i < t + h \,|\, t \leq T_i)$$

which can be expressed in the survival function form

$$\lambda(t) = \lim_{h \to 0+} \frac{F(t+h) - F(t)}{h[1 - F(t)]} = \frac{F'(t)}{S(t)} = -\frac{S'(t)}{S(t)}$$

and the *cumulative hazard function*

$$\Lambda(t) = \int_0^t \lambda(s) \, \mathrm{d}s$$

which can also be rewritten as

$$\Lambda(t) = -\int_0^t \frac{S'(s)}{S(s)} \, \mathrm{d}s = -\log S(t).$$

It is therefore sufficient to know one of the functions $S, \lambda$ and $\Lambda$ to reconstruct the other two. Estimators and their properties can be conveniently derived using counting process theory. Denote the observed event indicator and count in the form of right-continuous counting processes

$$N_i(t) = \mathbb{1}(T_i^\star \leq t, \Delta_i = 1), \qquad \overline{N}(t) = \sum_{i=1}^n N_i(t)$$

and the at-risk indicator and count in the form of left-continuous counting processes

$$Y_i(t) = \mathbb{1}(T_i^\star \geq t), \qquad \overline{Y}(t) = \sum_{i=1}^n Y_i(t).$$

Also denote $\mathcal{F}_t = \sigma\{N_i(s), \mathbb{1}(T_i^\star \le s, \Delta_i = 0) : 0 \le s \le t,\, i = 1, \dots, n\}$. Under the *independent censoring assumption*

$$\lambda(t) \equiv \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \le T_i < t + h \,|\, t \le T_i,\, t \le C_i)$$

the Doob-Meyer theorem A.1 implies that the counting process $N_i(t)$ can be decomposed into a sum of an increasing $\mathcal{F}_t$-predictable process $A_i(t)$ representing the mean structure and a zero-mean $\mathcal{F}_t$-martingale $M_i(t)$ representing the noise. In the current setting the decomposition can be written as

$$N_i(t) = A_i(t) + M_i(t) = \int_0^t Y_i(s)\, \mathrm{d}\Lambda(s) + M_i(t)$$

(Fleming and Harrington, 1991, p. 26). The *Nelson-Aalen estimator* of the cumulative hazard function has the form

$$\widehat{\Lambda}(t) = \int_0^t \frac{\mathrm{d}\overline{N}(s)}{\overline{Y}(s)} = \sum_{t_k \le t} \frac{\Delta \overline{N}(t_k)}{\overline{Y}(t_k)}.$$

The survival function is typically estimated with the *Kaplan-Meier estimator*

$$\widehat{S}(t) = \prod_{t_k \le t} [1 - \mathrm{d}\widehat{\Lambda}(t_k)] = \prod_{t_k \le t} \left[1 - \frac{\Delta \overline{N}(t_k)}{\overline{Y}(t_k)}\right].$$

## 1.2 The Case of Time-fixed Covariates

Next assume that the observed information is contained in the vector $(T_i^\star, \Delta_i, \mathbf{X}_i^\mathsf{T})^\mathsf{T}$ where $\mathbf{X}_i$ is a random vector of $p$ covariates associated with individual $i$. We define the conditional survival function

$$S(t \,|\, \mathbf{X}_i) = 1 - F(t \,|\, \mathbf{X}_i) = \mathsf{P}(T_i > t \,|\, \mathbf{X}_i),$$

the conditional hazard function

$$\lambda(t \,|\, \mathbf{X}_i) = \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \le T_i < t + h \,|\, t \le T_i, \mathbf{X}_i)$$

and the conditional cumulative hazard function

$$\Lambda(t \,|\, \mathbf{X}_i) = \int_0^t \lambda(s \,|\, \mathbf{X}_i)\, \mathrm{d}s = -\log S(t \,|\, \mathbf{X}_i).$$

The vector of covariates is incorporated in the model through the hazard function in the form

$$\lambda(t \,|\, \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\mathsf{T} \boldsymbol{\beta}}.$$

The *Cox proportional hazards model* arises when $\lambda_0$ is allowed to be an arbitrary non-negative function of time and is treated as a nuisance functional parameter. The *proportional hazards assumption* means that the hazard ratio

$$\frac{\lambda(t \,|\, \mathbf{X}_i)}{\lambda(t \,|\, \mathbf{X}_{i'})} = \frac{\lambda_0(t) e^{\mathbf{X}_i^\mathsf{T} \boldsymbol{\beta}}}{\lambda_0(t) e^{\mathbf{X}_{i'}^\mathsf{T} \boldsymbol{\beta}}} = \frac{e^{\mathbf{X}_i^\mathsf{T} \boldsymbol{\beta}}}{e^{\mathbf{X}_{i'}^\mathsf{T} \boldsymbol{\beta}}} \tag{1.1}$$

does not depend on time and therefore the hazards $\lambda(t \,|\, \mathbf{X}_i)$ and $\lambda(t \,|\, \mathbf{X}_{i'})$ are proportional in the sense that $\lambda(t \,|\, \mathbf{X}_i) = c(\mathbf{X}_i, \mathbf{X}_{i'}) \cdot \lambda(t \,|\, \mathbf{X}_{i'})$ where $c(\mathbf{X}_i, \mathbf{X}_{i'}) = e^{(\mathbf{X}_i^\mathsf{T} - \mathbf{X}_{i'}^\mathsf{T})\boldsymbol{\beta}}$ is time-invariant. Increasing the $j$-th component of $\mathbf{X}_i$ by one therefore causes the hazard to increase $e^{\beta_j}$ times. Under the independent censoring assumption

$$\lambda(t \,|\, \mathbf{X}_i) = \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \le T_i < t + h \,|\, t \le T_i,\, t \le C_i,\, \mathbf{X}_i)$$

the Doob-Meyer decomposition with respect to filtration $\mathcal{F}_t = \sigma\{N_i(s),\, Y_i(s+),\, \mathbf{X}_i : 0 \le s \le t,\, i = 1, \ldots, n\}$ gives

$$N_i(t) = A_i(t) + M_i(t) = \int_0^t Y_i(s) e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}} \,\mathrm{d}\Lambda_0(s) + M_i(t).$$

## 1.3 The Case of Time-varying Covariates

Finally suppose that the observed information is contained in the triplet $(T_i^\star, \Delta_i, \mathbf{X}_i(t) : 0 \le t \le T_i^\star)$ where $\mathbf{X}_i(t)$ is a left-continuous covariate process. Then we define $\mathcal{F}_t^{\mathbf{X}_i} = \sigma\{\mathbf{X}_i(s), 0 \le s \le t\}$ and the conditional hazard function

$$\lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \le T_i < t + h \,|\, t \le T_i, \mathcal{F}_t^{\mathbf{X}_i}).$$

The covariate process is incorporated in the model through the hazard function in the form

$$\lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \lambda_0(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}}.$$

Under the independent censoring assumption

$$\lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \lim_{h \to 0+} \frac{1}{h} \mathsf{P}(t \le T_i < t + h \,|\, t \le T_i,\, t \le C_i,\, \mathcal{F}_t^{\mathbf{X}_i})$$

the Doob-Meyer decomposition with respect to filtration $\mathcal{F}_t = \sigma\{N_i(s),\, Y_i(s+),\, \mathbf{X}_i(s) : 0 \le s \le t,\, i = 1, \ldots, n\}$ gives

$$N_i(t) = A_i(t) + M_i(t) = \int_0^t Y_i(s) e^{\mathbf{X}_i^\mathsf{T}(s)\boldsymbol{\beta}} \,\mathrm{d}\Lambda_0(s) + M_i(t).$$

## 1.4 Partial Likelihood Theory

Inference about the Cox proportional hazards model is based on the *partial likelihood* theory introduced by Cox (1972, 1975). The partial likelihood function has the form (Therneau and Grambsch, 2000, p. 40)

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{t \ge 0} \left[ \frac{Y_i(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}}}{\sum_{i'=1}^{n} Y_{i'}(t) e^{\mathbf{X}_{i'}^\mathsf{T}(t)\boldsymbol{\beta}}} \right]^{\Delta N_i(t)}.$$

The partial log-likelihood can be written as

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\infty \left\{ Y_i(t)\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta} - \log\left[ \sum_{i'=1}^{n} Y_{i'}(t) e^{\mathbf{X}_{i'}^\mathsf{T}(t)\boldsymbol{\beta}} \right] \right\} \mathrm{d}N_i(t).$$

Differentiation with respect to $\boldsymbol{\beta}$ yields the score function

$$\mathcal{U}(\boldsymbol{\beta}) = \frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_i(t) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t)]\, \mathrm{d}N_i(t) = \sum_{k=1}^{m} [\mathbf{X}_{i(k)}(t_k) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k)]$$

where the individual experiencing an event at time $t_k$ is denoted as $i(k)$ and

$$\overline{\mathbf{X}}(\boldsymbol{\beta}, t) = \frac{\sum_{i=1}^n Y_i(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}} \mathbf{X}_i(t)}{\sum_{i=1}^n Y_i(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}}}$$

is the conditional expectation of $\mathbf{X}_i(t)$ over the risk set applicable at time $t$ where the individual hazards $\lambda_i(t_k)$ play the role of weights

$$\begin{aligned} \mathsf{E}[\mathbf{X}_{i(k)}(t_k) \mid \mathcal{F}_{t_k}] &= \frac{\sum_{i:T_i^\star \geq t_k} \lambda_i(t_k \mid \mathcal{F}_{t_k}) \mathbf{X}_i(t_k)}{\sum_{i:T_i^\star \geq t_k} \lambda_i(t_k \mid \mathcal{F}_{t_k})} \\ &= \frac{\sum_{i=1}^n Y_i(t_k) e^{\mathbf{X}_i^\mathsf{T}(t_k)\boldsymbol{\beta}} \mathbf{X}_i(t_k)}{\sum_{i=1}^n Y_i(t_k) e^{\mathbf{X}_i^\mathsf{T}(t_k)\boldsymbol{\beta}}} = \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k). \end{aligned} \tag{1.2}$$

The score function $\mathcal{U}(\boldsymbol{\beta})$ is then set to zero and the equation is solved numerically using the Newton-Raphson algorithm. The partial likelihood estimator $\widehat{\boldsymbol{\beta}}$ of parameter $\boldsymbol{\beta}$ is obtained. The observed information matrix has the form

$$\mathcal{I}(\boldsymbol{\beta}) = -\frac{\partial^2 \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\mathsf{T}} = \sum_{i=1}^{n} \int_0^\infty \mathcal{V}(\boldsymbol{\beta}, t)\, \mathrm{d}N_i(t) = \sum_{k=1}^{m} \mathcal{V}(\boldsymbol{\beta}, t_k)$$

where

$$\mathcal{V}(\boldsymbol{\beta}, t) = \frac{\sum_{i=1}^n Y_i(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}} [\mathbf{X}_i(t) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t)]^{\otimes 2}}{\sum_{i=1}^n Y_i(t) e^{\mathbf{X}_i^\mathsf{T}(t)\boldsymbol{\beta}}}$$

is the weighted variance matrix of $\mathbf{X}_i(t)$ over the risk set applicable at time $t$. By analogy with (1.2) it holds that

$$\begin{aligned} \mathsf{var}[\mathbf{X}_{i(k)}(t_k) \mid \mathcal{F}_{t_k}] &= \frac{\sum_{i:T_i^\star \geq t_k} \lambda_i(t_k \mid \mathcal{F}_{t_k}) [\mathbf{X}_i(t_k) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k)]^{\otimes 2}}{\sum_{i:T_i^\star \geq t_k} \lambda_i(t_k \mid \mathcal{F}_{t_k})} \\ &= \frac{\sum_{i=1}^n Y_i(t_k) e^{\mathbf{X}_i^\mathsf{T}(t_k)\boldsymbol{\beta}} [\mathbf{X}_i(t_k) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k)]^{\otimes 2}}{\sum_{i=1}^n Y_i(t_k) e^{\mathbf{X}_i^\mathsf{T}(t_k)\boldsymbol{\beta}}} = \mathcal{V}(\boldsymbol{\beta}, t_k). \end{aligned} \tag{1.3}$$

The appealing attribute of the partial likelihood is that the asymptotic properties of the classical likelihood are preserved. Denote $\mathcal{J}(\boldsymbol{\beta})$ the limit in probability of $n^{-1}\mathcal{I}(\boldsymbol{\beta})$. Kalbfleisch and Prentice (2002, 174–176) establish regularity conditions under which the limit exists and

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathcal{U}(\boldsymbol{\beta}) &\xrightarrow{\mathcal{D}} \mathsf{N}[\mathbf{0}, \mathcal{J}(\boldsymbol{\beta})], \\ \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &\xrightarrow{\mathcal{D}} \mathsf{N}[\mathbf{0}, \mathcal{J}^{-1}(\boldsymbol{\beta})]. \end{aligned}$$

It is therefore possible to test

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$$

with one of the three test statistics with the asymptotic $\chi^2_p$ distribution under the null hypothesis, namely the *partial likelihood ratio test* statistic

$$2[\ell(\widehat{\boldsymbol{\beta}}) - \ell(\boldsymbol{\beta}_0)] \xrightarrow{\mathcal{D}} \chi^2_p,$$

the *score test* statistic

$$\mathcal{U}^\mathsf{T}(\boldsymbol{\beta}_0)\mathcal{I}^{-1}(\boldsymbol{\beta}_0)\mathcal{U}(\boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \chi^2_p$$

and the *Wald test* statistic

$$(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^\mathsf{T}\mathcal{I}(\widehat{\boldsymbol{\beta}})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} \chi^2_p.$$

## 1.5 Residuals

Here and further we denote with a hat those expressions where $\widehat{\boldsymbol{\beta}}$ instead of $\boldsymbol{\beta}$ is plugged in and the cumulative baseline hazard function

$$\Lambda_0(t) = \int_0^t \lambda_0(s)\,\mathrm{d}s$$

is replaced with the *Breslow estimator*

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\mathrm{d}\overline{N}(s)}{\sum_{i=1}^n Y_i(s)e^{\mathbf{X}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}}}.$$

For example we write

$$\widehat{M_i}(t) = N_i(t) - \int_0^t Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}\,\mathrm{d}\widehat{\Lambda}_0(s). \tag{1.4}$$

The *martingale residuals* are defined as $\widehat{M_i} = \widehat{M_i}(\infty)$. More insight into equation (1.4) can be gained in case of time-fixed covariates where

$$\Lambda(t \,|\, \mathbf{X}_i) = \int_0^t \lambda(s \,|\, \mathbf{X}_i)\,\mathrm{d}s = \int_0^t \lambda_0(s)e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}\,\mathrm{d}s = e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}\Lambda_0(t) \tag{1.5}$$

allows to write

$$\begin{aligned} M_i(t) &= N_i(t) - \int_0^t Y_i(s)e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}\,\mathrm{d}\Lambda_0(s) = N_i(t) - \int_0^t Y_i(s)\,\mathrm{d}\Lambda(s \,|\, \mathbf{X}_i) \\ &= N_i(t) - \int_0^{T_i^\star \wedge t} \mathrm{d}\Lambda(s \,|\, \mathbf{X}_i) = N_i(t) - \Lambda(T_i^\star \wedge t \,|\, \mathbf{X}_i) \end{aligned}$$

in the form comparing event indicator with the total amount of risk undertaken summarized in

$$M_i = \Delta_i - \Lambda(T_i^\star \,|\, \mathbf{X}_i) \quad \text{and} \quad \widehat{M_i} = \Delta_i - e^{\mathbf{X}_i^\mathsf{T}\widehat{\boldsymbol{\beta}}}\,\widehat{\Lambda}_0(T_i^\star).$$

For $i \neq i'$ it holds that

$$\mathsf{E}M_i = 0 \quad \text{and} \quad \mathsf{cov}(M_i, M_{i'}) = 0$$

as well as

$$\mathsf{var}M_i(t) = \mathsf{E}M_i^2(t) = \mathsf{E}A_i(t) = \mathsf{E}N_i(t) \tag{1.6}$$

(Therneau and Grambsch, 2000, p. 22). While it also holds that $\mathsf{E}\widehat{M}_i = 0$, the martingale residuals $\widehat{M}_1, \ldots, \widehat{M}_n$ are correlated as they sum to zero

$$\begin{aligned}
\sum_{i=1}^n \widehat{M}_i(t) &= \sum_{i=1}^n \left[N_i(t) - \int_0^t Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}\, \mathrm{d}\widehat{\Lambda}_0(s)\right] \\
&= \sum_{i=1}^n \left[\int_0^t \mathrm{d}N_i(s) - \int_0^t Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}} \frac{\sum_{i'=1}^n \mathrm{d}N_{i'}(s)}{\sum_{i'=1}^n Y_{i'}(s)e^{\mathbf{X}_{i'}^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}}\right] \\
&= \left[\sum_{i=1}^n \int_0^t \mathrm{d}N_i(s)\right] - \left[\sum_{i'=1}^n \int_0^t \mathrm{d}N_{i'}(s)\right] \frac{\sum_{i=1}^n Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}}{\sum_{i'=1}^n Y_{i'}(s)e^{\mathbf{X}_{i'}^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}} = 0.
\end{aligned}$$

We define the *score process*

$$\mathbf{U}_i(t) = \int_0^t [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\boldsymbol{\beta}, s)]\, \mathrm{d}M_i(s)$$

and the *score residuals* $\widehat{\mathbf{U}}_i = \widehat{\mathbf{U}}_i(\infty)$. It follows from the definition of $\widehat{\boldsymbol{\beta}}$ that $\sum_{i=1}^n \widehat{\mathbf{U}}_i = \mathbf{0}$ since

$$\begin{aligned}
\widehat{\mathbf{U}}_i(t) &= \int_0^t [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]\, \mathrm{d}\widehat{M}_i(s) \\
&= \int_0^t [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)][\,\mathrm{d}N_i(s) + Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}\, \mathrm{d}\widehat{\Lambda}_0(s)] \\
&= \int_0^t [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)][\,\mathrm{d}N_i(s) + \frac{Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}}{\sum_{i=1}^n Y_i(s)e^{\mathbf{X}_i^\mathsf{T}(s)\widehat{\boldsymbol{\beta}}}}\, \mathrm{d}\overline{N}(s)] \\
&= \int_0^t [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]\, \mathrm{d}N_i(s).
\end{aligned} \tag{1.7}$$

Summing the score process over all individuals and partitioning with respect to the observed event times we obtain

$$\mathbf{R}_k = \sum_{i=1}^n \int_{t_{k-1}}^{t_k} [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\boldsymbol{\beta}, s)]\, \mathrm{d}M_i(s) = \sum_{i=1}^n [\mathbf{U}_i(t_k) - \mathbf{U}_i(t_{k-1})] \tag{1.8}$$

and define the *Schoenfeld residuals* as $\widehat{\mathbf{R}}_k$. In the case of no ties – which is implied by the assumed continuous event times distribution – it follows from (1.7) that the Schoenfeld residuals can be rewritten in the form (1.9) originally referred to as the *partial residuals* (Schoenfeld, 1982)

$$\begin{aligned}
\widehat{\mathbf{R}}_k &= \sum_{i=1}^n [\widehat{\mathbf{U}}_i(t_k) - \widehat{\mathbf{U}}_i(t_{k-1})] \\
&= \sum_{i=1}^n \int_{t_{k-1}}^{t_k} [\mathbf{X}_i(s) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]\, \mathrm{d}N_i(s) \\
&= \sum_{i=1}^n [\mathbf{X}_i(t_k) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, t_k)]\Delta N_i(t_k) \\
&= \mathbf{X}_{i(k)}(t_k) - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, t_k)
\end{aligned} \tag{1.9}$$

where $\Delta N_i(t_k) = N_i(t_k) - N_i(t_k-)$ is the indicator that an individual $i$ is experiencing an event at time $t_k$. From (1.2) we obtain

$$\begin{aligned}
\mathsf{E}(\mathbf{R}_k \mid \mathcal{F}_{t_k}) &= \mathsf{E}[\mathbf{X}_{i(k)}(t_k) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k) \mid \mathcal{F}_{t_k}] \\
&= \mathsf{E}[\mathbf{X}_{i(k)}(t_k) - \mathsf{E}[\mathbf{X}_{i(k)}(t_k) \mid \mathcal{F}_{t_k}] \mid \mathcal{F}_{t_k}] = \mathbf{0} \qquad (1.10)
\end{aligned}$$

and using (1.3) we get

$$\begin{aligned}
\mathsf{var}(\mathbf{R}_k \mid \mathcal{F}_{t_k}) &= \mathsf{var}[\mathbf{X}_{i(k)}(t_k) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t) \mid \mathcal{F}_{t_k}] \\
&= \mathsf{var}[\mathbf{X}_{i(k)}(t_k) - \mathsf{E}[\mathbf{X}_{i(k)}(t_k) \mid \mathcal{F}_{t_k}] \mid \mathcal{F}_{t_k}] \\
&= \mathsf{var}[\mathbf{X}_{i(k)}(t_k) \mid \mathcal{F}_{t_k}] = \mathcal{V}(\boldsymbol{\beta}, t_k). \qquad (1.11)
\end{aligned}$$

It holds that $\mathbf{R}_k$ is $\mathcal{F}_{k'}$-measurable for $k < k'$ and

$$\mathsf{E}(\mathbf{R}_{k'} \mid \mathbf{R}_k) = \mathsf{E}[\mathsf{E}(\mathbf{R}_{k'} \mid \mathcal{F}_{t_{k'}}) \mid \mathbf{R}_k] = \mathsf{E}\,\mathbf{R}_{k'}$$

implies

$$\mathsf{cov}(\mathbf{R}_k, \mathbf{R}_{k'} \mid \mathcal{F}_{t_k}) = 0. \qquad (1.12)$$

The *scaled Schoenfeld residuals* $\widehat{\mathbf{R}}_k^\star$ are obtained from the expression

$$\mathbf{R}_k^\star = \mathcal{V}^{-1}(\boldsymbol{\beta}, t_k)\mathbf{R}_k.$$

# 2. Tests of the Proportionality of Hazards

The Cox proportional hazards model

$$\lambda(t \mid \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal \boldsymbol{\beta}}$$

carries two important assumptions worth depicting explicitly:

1. The hazard ratio (1.1) does not depend on time.
2. The multiplicative effect on the hazard of unit increase in a single covariate keeping all others fixed does not depend on time.

While the first condition is typically called the *proportional hazards assumption* it is actually its consequence described in the second condition that really matters. The hazard ratio itself is of little interest as long as it is possible to summarize the effect of a covariate with a single parameter. Suppose covariate values are recorded several times over the course of the study. A suitable model for this situation has the form

$$\lambda(t \mid \mathcal{F}_t^{\mathbf{X}_i}) = \lambda_0(t) e^{\mathbf{X}_i^\intercal(t) \boldsymbol{\beta}}$$

and the hazard ratio

$$\frac{\lambda(t \mid \mathcal{F}_t^{\mathbf{X}_i})}{\lambda(t \mid \mathcal{F}_t^{\mathbf{X}_{i'}})} = \frac{\lambda_0(t) e^{\mathbf{X}_i^\intercal(t) \boldsymbol{\beta}}}{\lambda_0(t) e^{\mathbf{X}_{i'}^\intercal(t) \boldsymbol{\beta}}} = \frac{e^{\mathbf{X}_i^\intercal(t) \boldsymbol{\beta}}}{e^{\mathbf{X}_{i'}^\intercal(t) \boldsymbol{\beta}}}$$

does depend on time. Nevertheless, the multiplicative effect on the hazard of unit increase in the $j$-th covariate is still summarized by a single coefficient $e^{\beta_j}$. Some authors like Kalbfleisch and Prentice (2002) are even reluctant to use the term *proportional hazards model* as it is associated with unnecessary limitations. This interpretation suggests that the real objective of model validity verification might be to test for the time-invariance of the regression coefficients. An alternative to the proportional hazards could be formulated through the model

$$\lambda(t \mid \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal \boldsymbol{\beta}(t)}.$$

In Section 2.1 we will consider one class of such models with the most basic example in the form

$$\lambda(t \mid \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal (\boldsymbol{\beta} + t\boldsymbol{\theta})}.$$

Interestingly, all models of this class can be reformulated in terms of time-varying covariates such as

$$\lambda(t \mid \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal \boldsymbol{\beta} + (t \times \mathbf{X}_i^\intercal)\boldsymbol{\theta}}.$$

This suggests that standard tools can be applied for testing of the interaction terms significance. Indeed, it is possible to proceed this way but it had long been technically cumbersome to recreate datasets with time-transformed covariates for each separate period of time. This inconvenience has been overcome with the addition of `tt` option to the `coxph` function from `survival` library in R (Therneau and Crowson, 2013).

## 2.1 Grambsch-Therneau Class of Tests

The approach based on the time-varying regression coefficients was first adopted by Schoenfeld (1982) for a single time-varying regression coefficient and later generalized by Grambsch and Therneau (1994) for more than one time-varying regression coefficient. Let $Q_j(t)$ be a $\mathcal{F}_t$-predictable process and $G(t)$ be a diagonal matrix where

$$G_{jj}(t) = Q_j(t) - \overline{Q}_j = Q_j(t) - \frac{1}{m}\sum_{k=1}^{m} Q_j(t_k).$$

Schoenfeld (1982) considered the case where

$$\beta_j(t) = \beta_j + [Q_j(t) - \overline{Q}_j]\theta_j$$

and tested

$$H_0 : \theta_j = 0 \quad \text{against} \quad H_1 : \theta_j \neq 0.$$

Grambsch and Therneau (1994) investigated the multivariate case

$$\boldsymbol{\beta}(t) = \boldsymbol{\beta} + G(t)\boldsymbol{\theta}$$

allowing for a joint test of

$$H_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{against} \quad H_1 : \boldsymbol{\theta} \neq \mathbf{0}.$$

The null hypothesis of the proportional hazards corresponds to the model

$$\lambda(t \,|\, \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal \boldsymbol{\beta}} \tag{2.1}$$

whereas the alternative model has the form

$$\lambda(t \,|\, \mathbf{X}_i) = \lambda_0(t) e^{\mathbf{X}_i^\intercal [\boldsymbol{\beta} + G(t)\boldsymbol{\theta}]}. \tag{2.2}$$

We will work in the more general multivariate setting and derive a test of the null hypothesis of the proportional hazards against a pre-specified alternative model (2.2). The derivation is divided in two parts for illustrative purposes. First, we will make an unrealistic assumption that the regression coefficient $\boldsymbol{\beta}$ is known and later this assumption will be left out.

### The first part of the derivation

Let us first assume that the parameter $\boldsymbol{\beta}$ is known and model (2.2) is valid. We can write the Schoenfeld residuals corresponding to model (2.1) in the form

$$\mathbf{R}_k = \{\mathbf{X}_{i(k)}(t_k) - \overline{\mathbf{X}}[\boldsymbol{\beta}(t_k), t_k]\} + \{\overline{\mathbf{X}}[\boldsymbol{\beta}(t_k), t_k] - \overline{\mathbf{X}}(\boldsymbol{\beta}, t_k)\} \tag{2.3}$$

instead of $\widehat{\mathbf{R}}_k$ as $\boldsymbol{\beta}$ is assumed to be known. The first term represents the Schoenfeld residuals under the true model (2.2) with zero mean. By differentiating with respect to $\boldsymbol{\beta}$ we obtain

$$\begin{aligned}
\frac{\partial \overline{\mathbf{X}}(\boldsymbol{\beta},t)}{\partial \boldsymbol{\beta}^{\mathsf{T}}} &= \frac{\partial}{\partial \boldsymbol{\beta}^{\mathsf{T}}}\left[\frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}\mathbf{X}_i(t)}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}}\right] \\
&= \frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}\mathbf{X}_i(t)^{\otimes 2}}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}} - \left[\frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}\mathbf{X}_i(t)}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}}\right]^{\otimes 2} \\
&= \frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}\mathbf{X}_i(t)^{\otimes 2}}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}} - \overline{\mathbf{X}}(\boldsymbol{\beta},t)^{\otimes 2} \\
&= \frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}[\mathbf{X}_i(t)^{\otimes 2} - \overline{\mathbf{X}}(\boldsymbol{\beta},t)^{\otimes 2}]}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}} \\
&= \frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}[\mathbf{X}_i(t)^{\otimes 2} - \mathbf{X}_i(t)\overline{\mathbf{X}}(\boldsymbol{\beta},t)^{\mathsf{T}} + \overline{\mathbf{X}}(\boldsymbol{\beta},t)\mathbf{X}_i(t)^{\mathsf{T}} - \overline{\mathbf{X}}(\boldsymbol{\beta},t)^{\otimes 2}]}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}} \\
&= \frac{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}[\mathbf{X}_i(t) - \overline{\mathbf{X}}(\boldsymbol{\beta},t)]^{\otimes 2}}{\sum_{i=1}^n Y_i(t)e^{\mathbf{X}_i^{\mathsf{T}}(t)\boldsymbol{\beta}}} = \mathcal{V}(\boldsymbol{\beta},t).
\end{aligned}$$

The second term in (2.3) can thus be approximated using first order Taylor expansion of $\overline{\mathbf{X}}[\boldsymbol{\beta}(t_k),t_k]$ around $\boldsymbol{\beta}$ as

$$\frac{\partial \overline{\mathbf{X}}(\boldsymbol{\beta},t_k)}{\partial \boldsymbol{\beta}^{\mathsf{T}}}[\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}] = \mathcal{V}_k \times [\boldsymbol{\beta}(t_k)-\boldsymbol{\beta}] = \mathcal{V}_k G_k \boldsymbol{\theta}$$

where $G_k = G(t_k)$ and $\mathcal{V}_k = \mathcal{V}(\boldsymbol{\beta},t_k)$. It follows that

$$\mathsf{E}(\mathbf{R}_k^{\star} \,|\, \mathcal{F}_{t_k}) = \mathsf{E}(\mathcal{V}_k^{-1}\mathbf{R}_k \,|\, \mathcal{F}_{t_k}) \doteq G_k \boldsymbol{\theta}. \tag{2.4}$$

Due to the contribution of the first term in (2.3) we obtain from (1.11) that

$$\mathsf{var}(\mathbf{R}_k^{\star} \,|\, \mathcal{F}_{t_k}) = \mathcal{V}_k^{-1}\mathcal{V}[\boldsymbol{\beta}(t_k),t_k]\mathcal{V}_k^{-1} \doteq \mathcal{V}_k^{-1} \tag{2.5}$$

and from (1.12) follows that for $k < k'$

$$\mathsf{cov}(\mathbf{R}_k^{\star},\mathbf{R}_{k'}^{\star} \,|\, \mathcal{F}_{t_k}) = \mathsf{cov}(\mathbf{R}_k,\mathbf{R}_{k'} \,|\, \mathcal{F}_{t_k}) = 0.$$

The assumption of known $\boldsymbol{\beta}$ together with conditions (2.4) and (2.5) imply that $\mathbf{R}_k^{\star}$ follows a linear model (Anděl, 2007, p. 193). The best linear unbiased predictor of $\boldsymbol{\theta}$ has the form

$$\widehat{\boldsymbol{\theta}} = \left(\sum_{k=1}^m G_k \mathcal{V}_k G_k\right)^{-1} \sum_{k=1}^m G_k \mathcal{V}_k \mathbf{R}_k^{\star} = \left(\sum_{k=1}^m G_k \mathcal{V}_k G_k\right)^{-1} \sum_{k=1}^m G_k \mathbf{R}_k$$

according to the Gauss-Markov theorem (Anděl, 2007, p. 194). Under the null hypothesis $H_0 : \boldsymbol{\theta} = \mathbf{0}$ it holds that

$$\left(\sum_{k=1}^m G_k \mathbf{R}_k\right)^{\mathsf{T}} \left(\sum_{k=1}^m G_k \mathcal{V}_k G_k\right)^{-1} \sum_{k=1}^m G_k \mathbf{R}_k \xrightarrow{\mathcal{D}} \chi_p^2.$$

### The second part of the derivation

Now, we abandon the assumption that $\boldsymbol{\beta}$ is known and denote $\widehat{\mathcal{V}}_k = \mathcal{V}(\widehat{\boldsymbol{\beta}}, t_k)$. Analogous arguments as for $\mathbf{R}_k^\star$ reveal that

$$\mathsf{E}\widehat{\mathbf{R}}_k^\star = \mathsf{E}(\widehat{\mathcal{V}}_k^{-1}\widehat{\mathbf{R}}_k) \doteq G_k\widehat{\boldsymbol{\theta}}$$

but the condition

$$\sum_{k=1}^m \widehat{\mathbf{R}}_k = \mathcal{U}(\widehat{\boldsymbol{\beta}}) = \mathbf{0}$$

following from (1.7) causes that the scaled Schoenfeld residuals $\widehat{\mathbf{R}}_1^\star, \ldots, \widehat{\mathbf{R}}_m^\star$ are correlated. To investigate their covariance structure we approximate $\widehat{\mathbf{R}}_k$ using the first order Taylor expansion around $\boldsymbol{\beta}$ as follows

$$\widehat{\mathbf{R}}_k \doteq \mathbf{R}_k - \mathcal{V}_k \times (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{2.6}$$

Hence denoting $\mathbf{R} = \sum_{k=1}^m \mathbf{R}_k$, $\mathcal{I} = \sum_{k=1}^m \mathcal{V}_k$ and summing over all events we obtain

$$\mathbf{0} \doteq \mathbf{R} - \mathcal{I} \times (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{2.7}$$

Combining (2.6) and (2.7) we get

$$\widehat{\mathbf{R}}_k \doteq \mathbf{R}_k - \mathcal{V}_k\mathcal{I}^{-1}\mathbf{R}$$

and since $\mathbf{R}_1, \ldots, \mathbf{R}_m$ are uncorrelated we obtain

$$\begin{aligned}
\mathsf{cov}(\widehat{\mathbf{R}}_k, \widehat{\mathbf{R}}_{k'}) &\doteq \mathsf{cov}(\mathbf{R}_k - \mathcal{V}_k\mathcal{I}^{-1}\mathbf{R}, \mathbf{R}_{k'} - \mathcal{V}_{k'}\mathcal{I}^{-1}\mathbf{R}) \\
&= \mathsf{cov}(\mathbf{R}_k, \mathbf{R}_{k'}) - \mathsf{cov}(\mathbf{R}_k, \mathcal{V}_{k'}\mathcal{I}^{-1}\mathbf{R}) \\
&\quad -\mathsf{cov}(\mathcal{V}_k\mathcal{I}^{-1}\mathbf{R}, \mathbf{R}_k) + \mathsf{cov}(\mathcal{V}_k\mathcal{I}^{-1}\mathbf{R}, \mathcal{V}_{k'}\mathcal{I}^{-1}\mathbf{R}) \\
&= \delta_{k,k'}\mathcal{V}_k - \mathsf{cov}(\mathbf{R}_k, \mathbf{R})\mathcal{I}^{-1}\mathcal{V}_{k'} \\
&\quad -\mathcal{V}_k\mathcal{I}^{-1}\mathsf{cov}(\mathbf{R}, \mathbf{R}_{k'}) + \mathcal{V}_k\mathcal{I}^{-1}\mathsf{var}(\mathbf{R})\mathcal{I}^{-1}\mathcal{V}_{k'} \\
&= \delta_{k,k'}\mathcal{V}_k - \mathsf{var}(\mathbf{R}_k)\mathcal{I}^{-1}\mathcal{V}_{k'} \\
&\quad -\mathcal{V}_k\mathcal{I}^{-1}\mathsf{var}(\mathbf{R}_{k'}) + \mathcal{V}_k\mathcal{I}^{-1}\mathcal{V}_{k'} \\
&= \delta_{k,k'}\mathcal{V}_k - \mathcal{V}_k\mathcal{I}^{-1}\mathcal{V}_{k'}.
\end{aligned}$$

Approximating $\mathcal{V}_k$ with $\widehat{\mathcal{V}}_k$ and $\mathcal{I}$ with $\widehat{\mathcal{I}} = \mathcal{I}(\widehat{\boldsymbol{\beta}})$ we get

$$\mathsf{var}\,\widehat{\mathbf{R}}_k^\star = \mathsf{var}(\widehat{\mathcal{V}}_k^{-1}\widehat{\mathbf{R}}_k) \doteq (\widehat{\mathcal{V}}_k^{-1}\mathcal{V}_k)\widehat{\mathcal{V}}_k^{-1} - (\widehat{\mathcal{V}}_k^{-1}\mathcal{V}_k)\mathcal{I}^{-1}(\mathcal{V}_k\widehat{\mathcal{V}}_k^{-1}) \doteq \widehat{\mathcal{V}}_k^{-1} - \widehat{\mathcal{I}}^{-1}.$$

The estimator of $\boldsymbol{\theta}$ has the form

$$\widetilde{\boldsymbol{\theta}} = D^{-1}\sum_{k=1}^m G_k\widehat{\mathbf{R}}_k$$

where the inverted variance matrix estimator

$$D = \sum_{k=1}^m G_k\widehat{\mathcal{V}}_kG_k - \left(\sum_{k=1}^m G_k\widehat{\mathcal{V}}_k\right)\widehat{\mathcal{I}}^{-1}\left(\sum_{k=1}^m G_k\widehat{\mathcal{V}}_k\right)^{\mathsf{T}}$$

takes the correlation into account and leads to

$$\Psi(D) = \widetilde{\boldsymbol{\theta}}^{\mathsf{T}} D \widetilde{\boldsymbol{\theta}} = \left( \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \right)^{\mathsf{T}} D^{-1} \left( \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \right) \xrightarrow{\mathcal{D}} \chi^2_p$$

under the null hypothesis. In practise, $\mathcal{V}(\widehat{\boldsymbol{\beta}}, t)$ does not vary dramatically with time (Therneau and Grambsch, 2000, p. 134) and is usually replaced with the average value

$$\mathcal{V}(\widehat{\boldsymbol{\beta}}, t) \doteq \frac{1}{m} \sum_{k=1}^{m} \widehat{\mathcal{V}}_k = \frac{1}{m} \widehat{\mathcal{I}}. \tag{2.8}$$

Since $\sum_{k=1}^{m} G_k = 0$ it follows that

$$D = \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \tag{2.9}$$

and

$$\widetilde{\boldsymbol{\theta}} = \left[ \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \right]^{-1} \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \tag{2.10}$$

from which we get

$$\begin{aligned} \Psi(D) = \widetilde{\boldsymbol{\theta}}^{\mathsf{T}} D \widetilde{\boldsymbol{\theta}} &= \left\{ \left[ \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \right]^{-1} \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \right\}^{\mathsf{T}} \left[ \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \right] \\ &\quad \times \left[ \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \right]^{-1} \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \\ &= \left( \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k \right)^{\mathsf{T}} \left[ \frac{1}{m} \sum_{k=1}^{m} G_k \widehat{\mathcal{I}} G_k \right]^{-1} \sum_{k=1}^{m} G_k \widehat{\mathbf{R}}_k. \end{aligned}$$

Moreover, let $Q_j(t) \equiv Q(t)$ be the same for all covariates and $R$ denote the matrix of the Schoenfeld residuals whose rows are formed by $\mathbf{R}_k^{\mathsf{T}}$ for $k = 1 \ldots, m$. By analogy $R^{\star} = m\mathcal{I}^{-1} R$. Stacking $Q(t_k) - \overline{Q}$ for $k = 1, \ldots, m$ into the vector $\mathbf{Q} - \overline{\mathbf{Q}}$, it is possible to test for the overall non-proportionality of hazards with

$$\begin{aligned} \Psi(D) &= \left\{ \sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{\mathbf{R}}_k \right\}^{\mathsf{T}} \left\{ \frac{1}{m} \widehat{\mathcal{I}} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2 \right\}^{-1} \\ &\quad \times \sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{\mathbf{R}}_k \\ &= \frac{(\mathbf{Q} - \overline{\mathbf{Q}})^{\mathsf{T}} R^{\star} \widehat{\mathcal{I}}^{-1} R^{\star \mathsf{T}} (\mathbf{Q} - \overline{\mathbf{Q}})}{\frac{1}{m} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2} \\ &= \frac{(\mathbf{Q} - \overline{\mathbf{Q}})^{\mathsf{T}} R \widehat{\mathcal{I}} R^{\mathsf{T}} (\mathbf{Q} - \overline{\mathbf{Q}})}{m \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2} \xrightarrow{\mathcal{D}} \chi^2_p \end{aligned}$$

holding under the null hypothesis. Alternatively, each covariate can be investigated separately. From (2.10) we get

$$\begin{aligned}
\widetilde{\theta}_j &= \left\{ \frac{1}{m} \widehat{\mathcal{I}}_{jj} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2 \right\}^{-1} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{R}_{kj} \\
&\doteq \left\{ \frac{1}{m} \widehat{\mathcal{I}}_{jj} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2 \right\}^{-1} \frac{1}{m} \widehat{\mathcal{I}}_{jj} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{R}^{\star}_{kj} \\
&= \frac{\sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{R}^{\star}_{kj}}{\sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2} \qquad (2.11)
\end{aligned}$$

and the test statistic obtained from (2.9) and (2.11) satisfies

$$\Psi_j(D) = \frac{\widetilde{\theta}_j^2}{D_{jj}} = \frac{\{\sum_{k=1}^{m} [Q(t_k) - \overline{Q}] \widehat{R}^{\star}_{kj}\}^2}{m \widehat{\mathcal{I}}_{jj}^{-1} \sum_{k=1}^{m} [Q(t_k) - \overline{Q}]^2} \xrightarrow{\mathcal{D}} \chi_1^2 \qquad (2.12)$$

under the null hypothesis. The test is available in R library `survival` under the name `cox.zph`. Formula (2.12) reveals that it is – up to the assumption of the constant variance – equivalent to testing for zero correlation between $\widehat{R}^{\star}_{kj}$ and $Q(t_k)$ or perhaps more illustratively to testing for zero slope in the regression of $\widehat{R}^{\star}_{kj}$ with $Q(t_k)$ as a single covariate. This enables to visualize the test by plotting the scaled Schoenfeld residuals against time to see if the time transformation $Q$ is well chosen, i.e. the spread of the values of $t_k$ is wide enough for the correlation-based inference to be reliable, see Therneau and Grambsch (2000, p. 144) for an example. Grambsch and Therneau (1994) noted that it is possible to construct a whole class of tests in this fashion when various forms of $Q_j(t)$ are considered. Actually, many such tests had already been proposed before using different argumentation.

### Cox test

Cox (1972) was the first to propose a test for the significant effect of adding time-varying covariates in the form of interaction between the $j$-th covariate and a pre-specified function of time. The Cox test therefore corresponds to the test described above with deterministic functions such as $Q_j(t) \equiv t$ and $Q_j(t) \equiv \log t$ used for all covariates. This test can be performed in R via the function `cox.zph` in library `survival` setting `transform = "identity"` and `transform = "log"` respectively.

### Nagelkerke-Oosting-Hart test

Nagelkerke et al. (1984) based their test on the serial correlation of the Schoenfeld residuals in the univariate case and more generally on the serial correlation of the linear combinations $\mathbf{R}_k^{\mathsf{T}} \widehat{\boldsymbol{\beta}}$ in the multivariate case. Should the proportional hazards assumption be violated due to the time-varying effect of $\mathbf{X}_i^{\mathsf{T}} \widehat{\boldsymbol{\beta}}$ the terms $\mathbf{R}_k^{\mathsf{T}} \widehat{\boldsymbol{\beta}}$ would be positively correlated (Nagelkerke et al., 1984). The test is equivalent to using the lagged Schoenfeld residuals $\widehat{R}_{kj}$ to construct $Q_j(t_1) = 0$ and $Q_j(t_{k+1}) = \widehat{\beta}_j^2 \widehat{R}_{kj}$ for $k \geq 1$ (Grambsch and Therneau, 1994).

### Breslow-Edler-Berger test

The test originally proposed by Breslow et al. (1984) in the context of two-sample problem aimed to present an alternative to the log-rank test which would be capable to detect the non-proportionality based on the crossing survival curves. The test generalized into the regression context is equivalent to using the ranks of the observed event times in $Q_j(t) \equiv \overline{N}(t-)$. This test can be performed in R via `cox.zph` setting `transform = "rank"`. A modification of the test developed by Harrell and Lee (1986) uses high absolute values of Fischer $z$-transform of the correlation between the Schoenfeld residuals and the corresponding ranks of the observed event times as evidence of monotone trend in the hazard ratio (Ng'andu, 1994). This variant was implemented in a former version of SAS within the procedure `phglm`.

### Moreau-O'Quigley-Mesbah test

Moreau et al. (1985) presented a test based on the partitioning of the time axis into several disjoint intervals. In the special case of the two-sample problem the test is equivalent to the chi-square goodness-of-fit test by Schoenfeld (1980). Let $J_1, \ldots, J_q$ be a disjoint decomposition of $[0, \infty)$ into $q$ intervals and let $\mathbf{a}_1, \ldots, \mathbf{a}_q$ be fixed vectors of $p$ components. The resulting test may be formulated in terms of piecewise constant function $Q_j(t) = \sum_{g=1}^{q} a_{gj} \mathbb{1}(t \in J_g)$. This parametrization allows to choose different functions for different covariates but the intervals $J_1, \ldots, J_q$ on which those functions are constant remain the same. A drawback of the test is that the partitioning of the time axis has to be done in advance in a rather arbitrary fashion. The test was further generalized by O'Quigley and Pessione (1989).

### Lin test

Lin (1991) suggested a test which is a generalization of the two-sample test by Gill and Schumacher (1987). The idea of the original Gill-Schumacher test was to construct two different estimators one of which puts more weight on early events than the other. The difference between the two estimators served as evidence against the proportionality of hazards. The Lin test compares the estimator $\widehat{\boldsymbol{\beta}}$ following from the score equations

$$\mathcal{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^\infty [\mathbf{X}_i(t) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t)] \, \mathrm{d}N_i(t) = \mathbf{0}$$

with the estimator $\widetilde{\boldsymbol{\beta}}$ following from the estimating equations

$$\sum_{i=1}^{n} \int_0^\infty Q(t)[\mathbf{X}_i(t) - \overline{\mathbf{X}}(\boldsymbol{\beta}, t)] \, \mathrm{d}N_i(t) = \mathbf{0}.$$

The test statistic $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \widetilde{\boldsymbol{\beta}})$ has asymptotically zero-mean multivariate normal distribution under the null hypothesis of the proportional hazards (Lin, 1991). If the estimator $\widetilde{\boldsymbol{\beta}}$ is based on the first iteration of the Newton-Raphson algorithm starting from $\widehat{\boldsymbol{\beta}}$ the test belongs to the Grambsch-Therneau class where $Q_j(t) \equiv$

$Q(t)$ is the same for all covariates (Grambsch and Therneau, 1994). In practice, $Q(t) = \widehat{S}(t-)$ where $\widehat{S}(t)$ is the Kaplan-Meier estimator is often used to detect hazard ratios monotonously dependent on time. It is the default option for R function `cox.zph`. Another possibility is to use $Q(t) = \widehat{S}(t-)[1 - \widehat{S}(t-)]$ which may be more appropriate to detect quadratic departures from the proportionality of hazards. We have implemented the quadratic variant in R as an extension of the function `cox.zph`.

**Khondoker-Islam test**

Recently, Khondoker and Islam (2009) suggested a test based on the log-log transformation of the survival function claiming it is a direct analogy of the routinely used graphical approach of plotting the log-log survival curves generalizing the concept from a single binary covariate setting and removing the arbitrariness of visual inspection. In fact, it is just another variant of the Lin test described above. Therefore, we have implemented the test in R as another extension of the function `cox.zph`. It uses $Q(t_k) = \log[-\log \widehat{S}(t_k-)]$ for $k = 1, \ldots, m$ to ensure that $Q$ is predictable with the only exception of setting $Q(t_0) = Q(t_1)$ to avoid zero argument in the outer logarithm.

**Grønnesby-Borgan test**

A test beyond the scope of the Grambsch-Therneau class was presented by Grønnesby and Borgan (1996), yet it was revealed by May and Hosmer (1998) that it is quite close in spirit. The idea of the Grønnesby-Borgan test is to to divide the subjects into a fixed number of groups according to their risk score $\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}$ similarly as in the case of the Moreau-O'Quigley-Mesbah test where the subjects are grouped with respect to the event times rather than the risk scores. The disadvantage that both of the tests share is that the number of the groups is arbitrary and has to be specified in advance. The groups usually contain about the same number of observations.

## 2.2 Lin-Wei-Ying Test

In the following, the central role is played by the components of the $p$-dimensional empirical score process

$$\widehat{\mathbf{U}}(t) = \sum_{i=1}^{n} \int_0^t [\mathbf{X}_i - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)] \, \mathrm{d}\widehat{M}_i(s)$$

denoted as $\widehat{U}_j(t) = \sum_{i=1}^{n} \widehat{U}_{ij}(t)$. The research of their asymptotic properties led to tests for which no functional form of the alternative hypothesis has to be specified. Before we proceed with the Lin-Wei-Ying test itself we try to motivate it with the historical evolution of the topic. The original idea of the test emerged in the work by Wei (1984) who showed that for a single binary variable the weighted empirical score process

$$\widehat{\mathcal{I}}_{11}^{-1/2} \widehat{U}_1(t)$$

is asymptotically equivalent to the Brownian bridge $W^0(t)$ on $[0, 1]$ under the proportional hazards assumption. Therneau et al. (1990) argued that the asymptotic equivalence is attained more generally for

$$\widehat{\mathcal{I}}_{jj}^{-1/2} \widehat{U}_j(t)$$

provided the asymptotic variance matrix $\mathcal{J}(t)$ of $n^{-1/2}\mathbf{U}(t)$ satisfies $\mathcal{J}_{jj'}(t) = 0$ for all $j' \neq j$ which means that the $j$-th covariate is asymptotically uncorrelated with all others. We present the reasoning behind the result of Therneau et al. (1990) from which the one by Wei (1984) follows. We first extend the definition of $\mathcal{I}(\boldsymbol{\beta})$ to allow for dependence on time

$$\mathcal{I}(\boldsymbol{\beta}, t) = \sum_{i=1}^{n} \int_0^t \mathcal{V}(\boldsymbol{\beta}, t) \, \mathrm{d}N_i(t)$$

and write $\mathcal{I}(t) = \mathcal{I}(\boldsymbol{\beta}, t)$. Also recall that $\mathcal{I} = \mathcal{I}(\infty)$. We combine two first order Taylor expansions of the empirical score process $\widehat{\mathbf{U}}(t) = \mathbf{U}(\widehat{\boldsymbol{\beta}}, t)$ around $\boldsymbol{\beta}$

$$\widehat{\mathbf{U}}(t) \doteq \mathbf{U}(t) - \mathcal{I}(t)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

and

$$\mathbf{0} = \widehat{\mathbf{U}}(\infty) \doteq \mathbf{U}(\infty) - \mathcal{I}(\infty)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Eliminating $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}$ we obtain

$$\widehat{\mathbf{U}}(t) \doteq \mathbf{U}(t) - \mathcal{I}(t)\mathcal{I}^{-1}(\infty)\mathbf{U}(\infty) \tag{2.13}$$

which can be suitably rewritten as

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{U}}(t) \doteq \frac{1}{\sqrt{n}}\mathbf{U}(t) - \frac{1}{n}\mathcal{I}(t) \left[\frac{1}{n}\mathcal{I}(\infty)\right]^{-1} \frac{1}{\sqrt{n}}\mathbf{U}(\infty).$$

Let $\mathbf{W}(t)$ be a $p$-dimensional zero-mean Gaussian process with independent increments and the variance matrix $\mathcal{J}(t)$. (Fleming and Harrington, 1991, p. 289–291) establish regularity conditions under which

$$\frac{1}{\sqrt{n}}\mathbf{U}(t) \xrightarrow{\mathcal{D}} \mathbf{W}(t) \tag{2.14}$$

holds as well as

$$\frac{1}{n}\mathcal{I}(t) \xrightarrow{\mathcal{P}} \mathcal{J}(t) \tag{2.15}$$

and

$$\frac{1}{n}\widehat{\mathcal{I}}(t) \xrightarrow{\mathcal{P}} \mathcal{J}(t). \tag{2.16}$$

This means that both $n^{-1}\mathcal{I}(t)$ and $n^{-1}\widehat{\mathcal{I}}(t)$ are consistent estimators of $\mathcal{J}(t)$. It follows from the Cramér-Slutsky theorem (Anděl, 2007, p. 333) that

$$\frac{1}{\sqrt{n}}\widehat{\mathbf{U}}(t) \xrightarrow{\mathcal{D}} \mathbf{W}(t) - \mathcal{J}(t)\mathcal{J}^{-1}(\infty)\mathbf{W}(\infty).$$

The asymptotic covariance matrix of the process $n^{-1/2}\widehat{\mathbf{U}}(t)$ has the form

$$\begin{aligned}\mathcal{C}(s,t) &= \mathsf{E}[\mathbf{W}(s) - \mathcal{J}(s)\mathcal{J}^{-1}(\infty)\mathbf{W}(\infty)][\mathbf{W}(t) - \mathcal{J}(t)\mathcal{J}^{-1}(\infty)\mathbf{W}(\infty)]^{\mathsf{T}} \\ &= \mathsf{cov}[\mathbf{W}(s), \mathbf{W}(t)] - \mathcal{J}(s)\mathcal{J}^{-1}(\infty)\mathsf{cov}[\mathbf{W}(\infty), \mathbf{W}(t)] \\ &\quad - \mathsf{cov}[\mathbf{W}(s), \mathbf{W}(\infty)]\mathcal{J}^{-1}(\infty)\mathcal{J}(t) \\ &\quad + \mathcal{J}(s)\mathcal{J}^{-1}(\infty)\mathsf{var}[\mathbf{W}(\infty)]\mathcal{J}^{-1}(\infty)\mathcal{J}(t)\end{aligned}$$

and for $s < t$ follows that

$$\mathcal{C}(s,t) = \mathcal{J}(s) - \mathcal{J}(s)\mathcal{J}^{-1}(\infty)\mathcal{J}(t)$$

because of the independent increments property

$$\begin{aligned}\mathsf{cov}[\mathbf{W}(s), \mathbf{W}(t)] &= \mathsf{cov}[\mathbf{W}(s), \mathbf{W}(t) - \mathbf{W}(s) + \mathbf{W}(s)] \\ &= \mathsf{cov}[\mathbf{W}(s), \mathbf{W}(t) - \mathbf{W}(s)] + \mathsf{var}\,\mathbf{W}(s) = \mathsf{var}\,\mathbf{W}(s).\end{aligned}$$

From (2.16) we get a special case

$$\frac{1}{n}\widehat{\mathcal{I}}(\infty) \xrightarrow{\mathcal{P}} \mathcal{J}(\infty)$$

and conclude using the Cramér-Slutsky theorem that when $\mathcal{J}_{jj'}(t) = 0$ for some $j = 1, \ldots, p$ and all $j' \neq j$ then

$$\widehat{\mathcal{I}}_{jj}^{-1/2}\widehat{U}_j(t) = \left[\frac{1}{n}\widehat{\mathcal{I}}_{jj}(\infty)\right]^{-1/2}\frac{1}{\sqrt{n}}\widehat{U}_j(t)$$

converges in distribution to

$$\mathcal{J}_{jj}^{-1/2}(\infty)\left[B_j(t) - \frac{\mathcal{J}_{jj}(t)}{\mathcal{J}_{jj}(\infty)}B_j(\infty)\right] = W^0\left[\frac{\mathcal{J}_{jj}(t)}{\mathcal{J}_{jj}(\infty)}\right]$$

where $W^0(t)$ is distributed as a Brownian bridge on $[0,1]$. However, the assumption of asymptotically uncorrelated covariates is often too strong to hold in practice which is the reason why the Lin-Wei-Ying test based on Monte Carlo simulation has been developed. In line with equation (2.13) Lin et al. (1993)

revealed that the process $n^{-1/2}\widehat{\mathbf{U}}(t)$ is asymptotically equivalent to the process $n^{-1/2}\widetilde{\mathbf{U}}(t)$ where

$$\begin{aligned}\widetilde{\mathbf{U}}(t) &= \sum_{i=1}^{n}\int_0^t [\mathbf{X}_i - \widetilde{\mathbf{X}}(\boldsymbol{\beta}, s)]\,\mathrm{d}M_i(s) \\ &\quad -\sum_{i=1}^{n}\int_0^t Y_i(s)e^{\mathbf{X}_i^{\mathsf{T}}\boldsymbol{\beta}}\mathbf{X}_i[\mathbf{X}_i - \widetilde{\mathbf{X}}(\boldsymbol{\beta}, s)]^{\mathsf{T}}\,\mathrm{d}\Lambda_0(s) \\ &\quad \times \mathcal{I}^{-1}\sum_{i=1}^{n}\int_0^\infty [\mathbf{X}_i - \widetilde{\mathbf{X}}(\boldsymbol{\beta}, s)]\,\mathrm{d}M_i(s)\end{aligned}$$

and $\widetilde{\mathbf{X}}(\boldsymbol{\beta}, t)$ denotes the limit in probability of $\overline{\mathbf{X}}(\boldsymbol{\beta}, t)$. Lin et al. (1993) further showed that $n^{-1/2}\widetilde{\mathbf{U}}(t)$ converges in distribution to a zero-mean Gaussian process. While the distribution of $M_i(t)$ is not known it follows from (1.6) that $\mathsf{var}M_i(t) = \mathsf{E}N_i(t)$. It is therefore possible to approximate the distribution of $\widetilde{\mathbf{U}}(t)$ by replacing $M_i(t)$ with $N_i(t)Z_i$ where $Z_i$ are independent random variables with the distribution $\mathsf{N}(0, 1)$. The empirical version of process $\widetilde{\mathbf{U}}(t)$ is constructed as

$$\begin{aligned}\widehat{\mathbf{U}}_{sim}(t) &= \sum_{i=1}^{n}[\mathbf{X}_i - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]N_i(t)Z_i \\ &\quad -\sum_{i=1}^{n}\int_0^t Y_i(s)e^{\mathbf{X}_i^{\mathsf{T}}\widehat{\boldsymbol{\beta}}}\mathbf{X}_i[\mathbf{X}_i - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]^{\mathsf{T}}\,\mathrm{d}\widehat{\Lambda}_0(s) \\ &\quad \times \widehat{\mathcal{I}}^{-1}\sum_{i=1}^{n}[\mathbf{X}_i - \overline{\mathbf{X}}(\widehat{\boldsymbol{\beta}}, s)]\Delta_i Z_i.\end{aligned}$$

The distribution of $n^{-1/2}\widehat{\mathbf{U}}(t)$ is approximated by sampling $n^{-1/2}\widehat{\mathbf{U}}_{sim}(t)$ via $Z_i$ while $(T_i^\star, \Delta_i,\ \mathbf{X}_i)$ is kept fixed. This approach is enabled by the fact that the unconditional distribution of $n^{-1/2}\widehat{\mathbf{U}}(t)$ and its conditional distribution given $(T_i^\star, \Delta_i, \mathbf{X}_i)$ coincide in the limit (Lin et al., 1993). Being able to simulate the score process under the proportional hazards assumption we can estimate the quantiles of the distributions of the test statistics

$$\sup_{t\geq 0}|\widehat{U}_j(t)| \quad \text{and} \quad \sup_{t\geq 0}\widehat{\mathcal{I}}_{jj}^{-1/2}|\widehat{U}_j(t)|.$$

It is also possible to consider omnibus test statistics such as

$$\sup_{t\geq 0}||\widehat{\mathbf{U}}(t)|| = \sup_{t\geq 0}\sqrt{\widehat{U}_1^2(t) + \cdots + \widehat{U}_p^2(t)}$$

and

$$\sup_{t\geq 0}\sum_{j=1}^{p}\widehat{\mathcal{I}}_{jj}^{-1/2}|\widehat{U}_j(t)|.$$

The test is implemented in R library `timereg` within the function `cox.aalen` and in SAS procedure `phreg` within the statement `assess` with option `ph`.

## 2.3 Scheike-Martinussen Test

Scheike and Martinussen (2004) introduced a test of the time-varying effect of some of the regression coefficients. This indicates an important feature of the test. All other presented tests suffer from the drawback that they test whether a certain covariate violates the proportional hazards assumption while assuming that other covariates do not. The test developed by Scheike and Martinussen (2004) works within the framework of the extended Cox model

$$\lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \lambda_0(t) e^{\mathbf{X}_{1i}^\intercal(t)\boldsymbol{\alpha} + \mathbf{X}_{2i}^\intercal(t)\boldsymbol{\beta}(t)} \tag{2.17}$$

incorporating both the time-fixed regression coefficients $\boldsymbol{\alpha}$ corresponding to covariates $\mathbf{X}_{1i}(t)$ and the time-varying regression coefficients $\boldsymbol{\beta}(t)$ corresponding to a different set of covariates $\mathbf{X}_{2i}(t)$. It is recommended to start with model allowing all regression coefficients to depend on time and sequentially test for the significance of the time effect. We can thus avoid the assumption that all regression coefficients apart from the one currently investigated are time-fixed. It is of interest to test

$$H_0 : \boldsymbol{\beta}(t) \equiv \boldsymbol{\beta} \quad \text{against} \quad H_1 : \boldsymbol{\beta}(t) \not\equiv \boldsymbol{\beta}.$$

Equivalently, we define the cumulative regression coefficients as

$$\mathbf{B}(t) = \int_0^t \boldsymbol{\beta}(s) \, \mathrm{d}s$$

and test

$$H_0 : \mathbf{B}(t) = t\boldsymbol{\beta} \quad \text{against} \quad H_1 : \mathbf{B}(t) \neq t\boldsymbol{\beta}.$$

The central idea of the test is analogous with the Lin-Wei-Ying test and the most noticeable difference is that the expressions are no longer neat. To avoid plethora of new notation or half-page expressions we focus on the principles here and refer the courageous reader to the original sources (Scheike and Martinussen, 2004; Martinussen and Scheike, 2006). The estimators $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\beta}}(t)$ of the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}(t)$ respectively are based on the maximum likelihood principle. The iterative procedure to obtain the estimators consists of the following steps:

0. Set the initial values of $\widehat{\boldsymbol{\alpha}}^{(0)}$ and $\widehat{\boldsymbol{\beta}}^{(0)}(t)$.

1. Use the estimators from the previous step $\widehat{\boldsymbol{\alpha}}^{(r)}$ and $\widehat{\boldsymbol{\beta}}^{(r)}(t)$ to compute the Breslow estimator

$$\widehat{\Lambda}_0^{(r)}(t) = \int_0^t \frac{\mathrm{d}\overline{N}(s)}{\sum_{i=1}^n Y_i(s) e^{\mathbf{X}_{1i}^\intercal(s)\widehat{\boldsymbol{\alpha}}^{(r)} + \mathbf{X}_{2i}^\intercal(s)\widehat{\boldsymbol{\beta}}^{(r)}(s)}}.$$

2. Smooth $\widehat{\Lambda}_0^{(r)}(t)$ to obtain $\widehat{\lambda}_0^{(r)}(t)$ using kernel estimation

$$\widehat{\lambda}_0^{(r)}(t) = \frac{1}{b} \int_0^\tau K\left(\frac{s-t}{b}\right) \mathrm{d}\widehat{\Lambda}_0^{(r)}(s)$$

where $b > 0$ is a bandwidth parameter and $K$ is a suitable kernel function.

3. Perform one step of the Newton-Raphson algorithm based on the score equations

$$\begin{aligned}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{X}_{1i}(t)-\overline{\mathbf{X}}_1[(\boldsymbol{\alpha}^\mathsf{T},\boldsymbol{\beta}^\mathsf{T}(t))^\mathsf{T},t]\}\,\mathrm{d}N_i(t) &= \mathbf{0},\\ \sum_{i=1}^{n}\{\mathbf{X}_{2i}(t)-\overline{\mathbf{X}}_2[(\boldsymbol{\alpha}^\mathsf{T},\boldsymbol{\beta}^\mathsf{T}(t))^\mathsf{T},t]\}\Delta N_i(t) &\equiv \mathbf{0} \qquad \forall\, t\in[0,\tau].\end{aligned}$$

4. Use the updating step of the Newton-Raphson algorithm to proceed form the current estimator $\widehat{\boldsymbol{\alpha}}^{(r)}$ to $\widehat{\boldsymbol{\alpha}}^{(r+1)}$.

5. Compute $\widehat{\mathbf{B}}^{(r+1)}(t)$ based on the Newton-Raphson updating equations with $\widehat{\boldsymbol{\alpha}}^{(r+1)}$ plugged in.

6. Smooth $\widehat{\mathbf{B}}^{(r+1)}(t)$ to obtain $\widehat{\boldsymbol{\beta}}^{(r+1)}(t)$ using kernel estimation

$$\widehat{\boldsymbol{\beta}}^{(r)}(t)=\frac{1}{b}\int_0^\tau K\left(\frac{s-t}{b}\right)\,\mathrm{d}\widehat{\mathbf{B}}^{(r)}(s)$$

and return to step 1.

Finally, we set $\widehat{\boldsymbol{\beta}}(t)=\widehat{\boldsymbol{\beta}}^{(\infty)}(t)$ and $\widehat{\mathbf{B}}(t)=\widehat{\mathbf{B}}^{(\infty)}(t)$. Martinussen and Scheike (2006, p. 208) establish conditions under which the process $\sqrt{n}[\widehat{\mathbf{B}}(t)-\mathbf{B}(t)]$ converges in distribution to a zero-mean Gaussian process. In the same spirit as for the Lin-Wei-Ying test Scheike and Martinussen (2004) show that the process $\sqrt{n}[\widehat{\mathbf{B}}(t)-\mathbf{B}(t)]$ is asymptotically equivalent to a certain process with remarkably technical expression whose empirical version can then be modified by multiplication of its $n$ summands with independent standard normal random variables to obtain a new process with the same limit distribution as $\sqrt{n}[\widehat{\mathbf{B}}(t)-\mathbf{B}(t)]$. To test the null hypothesis

$$H_0:\beta_j(t)\equiv\beta_j \quad \text{against} \quad H_1:\beta_j(t)\not\equiv\beta_j$$

we construct a test statistic of the Kolmogorov-Smirnov type

$$\sup_{0\le t\le\tau}\left|\widehat{B}_j(t)-\frac{t}{\tau}\widehat{B}_j(\tau)\right|$$

or a test statistic of the Cramér-von Mises type

$$\int_0^\tau\left[\widehat{B}_j(t)-\frac{t}{\tau}\widehat{B}_j(\tau)\right]^2\,\mathrm{d}t.$$

Kvaløy and Neef (2004) propose a test statistic of the Anderson-Darling type in the context of the proportional hazards testing as well. The quantiles of the test statistics distributions are approximated by complete analogy with the Lin-Wei-Ying test through sampling independent random variables from $\mathsf{N}(0,1)$ and transforming them into random variables with a distribution asymptotically equivalent to $\sqrt{n}[\widehat{\mathbf{B}}(t)-\mathbf{B}(t)]$. The test is implemented in R library `timereg` within the function `timecox`.

## 2.4 Lin-Zhang-Davidian Test

The members of the Grambsch-Therneau class of tests share the disadvantage of the need for an arbitrary specification of the alternative hypothesis. To overcome this obstacle consider a model with a single time-varying covariate

$$\lambda(t \mid \mathbf{X}_{1i}, X_{2i}) = \lambda_0(t) e^{\mathbf{X}_{1i}^{\mathsf{T}}\boldsymbol{\beta} + f(t)X_{2i}}$$

and allow $f$ to be an arbitrary smooth function of time with square integrable second derivative. The estimation is performed maximizing the penalized partial likelihood

$$\ell_{pen}(\boldsymbol{\beta}, f, \theta) = \ell(\boldsymbol{\beta}, f) - \frac{\theta}{2} \int [f''(t)]^2 \,\mathrm{d}t \tag{2.18}$$

where the second term penalizes variation in $f$ to avoid over-fitting when using large number of parameters. Before we give details on the recent test of Lin et al. (2006) we set the scene considering some earlier approaches. Sleeper and Harrington (1990) proposed to use non-penalized regression splines to model $f$ but the result was criticized by Gray (1992) for undue sensitivity to the number and location of the knots as well as overall instability. O'Sullivan (1988) stated that the maximum of the penalized partial likelihood (2.18) is attained for $f$ in the form of a cubic smoothing spline with the total of $m$ knots and provides methods for parameters estimation using B-spline basis functions (de Boor, 1978). However, allocating a parameter for each event time is not very efficient. Reduction of the dimensionality of the parameter space was proposed by Gray (1992, 1994) who developed a spline-based tests for the proportional hazards which only uses a moderate number of basis functions $1 < \mu \ll m$ so that

$$f(t) = \sum_{\vartheta=1}^{\mu} \gamma_\vartheta b_\vartheta(t, q) = \boldsymbol{\gamma}^{\mathsf{T}} \mathbf{b}(t, q)$$

where $b_\vartheta(t, q)$, $\vartheta = 1, \ldots, \mu$ are B-spline basis functions of order $q$. The choice of $\mu$ is arbitrary but Gray (1992) argues that using more than 10–20 knots hardly ever enhances the fit substantially. The penalized partial log-likelihood has the form

$$\begin{aligned}
\ell_{pen}(\boldsymbol{\beta}, f, \theta) &= \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \frac{\theta}{2} \int [f''(t)]^2 \,\mathrm{d}t \\
&= \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \frac{\theta}{2} \int [\boldsymbol{\gamma}^{\mathsf{T}} \mathbf{b}''(t, q)]^2 \,\mathrm{d}t \\
&= \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \frac{\theta}{2} \boldsymbol{\gamma}^{\mathsf{T}} \left[ \int \mathbf{b}''(t, q)^{\otimes 2} \,\mathrm{d}t \right] \boldsymbol{\gamma} \\
&= \ell(\boldsymbol{\beta}, \boldsymbol{\gamma}) - \frac{\theta}{2} \boldsymbol{\gamma}^{\mathsf{T}} P_B \boldsymbol{\gamma}
\end{aligned}$$

where $P_B = \int \mathbf{b}''(t, q)^{\otimes 2} \,\mathrm{d}t$ is a positive semi-definite matrix. Significant computational advantage could be gained if the penalty function is slightly modified as explained in the following. Denote $h$ the fixed distance between the adjacent knots. It has been shown (de Boor, 1978, p. 116) that the B-spline basis functions satisfy

$$\begin{aligned}\sum_{\vartheta} \gamma_\vartheta b'_\vartheta(t,q) &= \frac{1}{h}\sum_{\vartheta}(\gamma_{\vartheta+1}-\gamma_\vartheta) b_\vartheta(t,q-1)\\ &= \frac{1}{h}\sum_{\vartheta}\Delta\gamma_\vartheta b_\vartheta(t,q-1).\end{aligned}$$

Eilers and Marx (1996) note that by induction

$$\begin{aligned}\sum_{\vartheta} \gamma_\vartheta b''_\vartheta(t,q) &= \frac{1}{h}\sum_{\vartheta}(\gamma_{\vartheta+1}-\gamma_\vartheta) b'_\vartheta(t,q-1)\\ &= \frac{1}{h^2}\sum_{\vartheta}(\gamma_{\vartheta+2}-2\gamma_{\vartheta+1}+\gamma_\vartheta) b_\vartheta(t,q-2)\\ &= \frac{1}{h^2}\sum_{\vartheta}\Delta^2\gamma_\vartheta b_\vartheta(t,q-2). \qquad (2.19)\end{aligned}$$

The cubic B-spline penalty has the form

$$\begin{aligned}\frac{\theta}{2}\int\left[\sum_{\vartheta}\gamma_\vartheta b''_\vartheta(t,3)\right]^2 \mathrm{d}t &= \frac{\theta}{2h^4}\int\left[\sum_{\vartheta}\Delta^2\gamma_\vartheta b_\vartheta(t,1)\right]^2 \mathrm{d}t\\ &= \frac{\theta}{2h^4}\int\sum_{\vartheta}\sum_{\vartheta'}\Delta^2\gamma_\vartheta\Delta^2\gamma_{\vartheta'} b_\vartheta(t,1) b_{\vartheta'}(t,1)\,\mathrm{d}t.\end{aligned}$$

Only the adjacent linear B-splines overlap and therefore the only non-zero terms are either of the form $b_\vartheta^2(t,1)$ or $b_\vartheta(t,1)b_{\vartheta+1}(t,1)$ so the penalty term can be written as a sum of the difference penalty obtained for $\vartheta = \vartheta'$

$$\underbrace{\frac{\theta}{2h^4}\int b_\vartheta^2(t,1)\,\mathrm{d}t}_{\text{constant}} \times \sum_{\vartheta}(\Delta^2\gamma_\vartheta)^2$$

and a term depending only on the products of the adjacent second differences obtained for $\vartheta = \vartheta' \pm 1$

$$\underbrace{\frac{\theta}{h^4}\int b_\vartheta(t,1)b_{\vartheta+1}(t,1)\,\mathrm{d}t}_{\text{constant}} \times \sum_{\vartheta}\Delta^2\gamma_\vartheta\Delta^2\gamma_{\vartheta+1}$$

which is eventually excluded from the P-spline penalty (Eilers and Marx, 1996). Denote the $(\mu-2)\times\mu$ dimensional second differences matrix

$$K_P = \begin{pmatrix} 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1\end{pmatrix} \quad \text{and} \quad P_P = K_P^\mathsf{T} K_P.$$

The penalized partial log-likelihood with P-spline penalty is formed as

$$\begin{aligned}\ell_{pen}(\boldsymbol{\beta}, f, \theta) &= \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) - \frac{\theta}{2}\sum_{\vartheta=1}^{\mu-2}(\Delta^2\gamma_\vartheta)^2\\ &= \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) - \frac{\theta}{2}(K_P\boldsymbol{\gamma})^{\otimes 2}\\ &= \ell(\boldsymbol{\beta},\boldsymbol{\gamma}) - \frac{\theta}{2}\boldsymbol{\gamma}^\mathsf{T} P_P\boldsymbol{\gamma}.\end{aligned}$$

Define $\boldsymbol{\eta} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\gamma}^{\mathsf{T}})^{\mathsf{T}}$ and

$$\begin{aligned}
\mathcal{U}_{pen} &= \frac{\partial \ell_{pen}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\mathsf{T}}} = \frac{\partial \ell(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}^{\mathsf{T}}} - \begin{pmatrix} \mathbf{0} \\ P_P \boldsymbol{\gamma} \end{pmatrix}, \\
\mathcal{I}_{pen} &= -\frac{\partial^2 \ell_{pen}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^{\mathsf{T}}} = \mathcal{I} - \begin{pmatrix} 0 & 0 \\ 0 & -P_P \end{pmatrix}.
\end{aligned}$$

The variance matrix of the regression coefficients is typically estimated with $\widehat{V}$ in the form of either

$$\widehat{V}_1 = \widehat{\mathcal{I}}_{pen}^{-1} \quad \text{or} \quad \widehat{V}_2 = \widehat{\mathcal{I}}_{pen}^{-1} \widehat{\mathcal{I}} \widehat{\mathcal{I}}_{pen}^{-1},$$

the former being conservative and usually preferred (Therneau and Grambsch, 2000, p. 122). The null hypothesis of the proportional hazards corresponds to the case $\gamma_1 = \cdots = \gamma_\mu$ since the B-spline basis functions sum to one. Gray (1992) proposed to use a general Wald type test of

$$H_0 : C\boldsymbol{\eta} = 0 \quad \text{against} \quad H_1 : C\boldsymbol{\eta} \neq 0$$

where the $(\mu - 1) \times (p + \mu)$ constraint matrix $C$ corresponding to the proportional hazards assumption can be written as

$$C = \begin{pmatrix} 0 & \cdots & 0 & 1 & -1 & & \\ \vdots & \ddots & \vdots & & \ddots & \ddots & \\ 0 & \cdots & 0 & & & 1 & -1 \end{pmatrix}.$$

The test statistic which employs the maximum penalized partial likelihood estimator $\widehat{\boldsymbol{\eta}}$ has the form

$$(C\widehat{\boldsymbol{\eta}})^{\mathsf{T}} [C\mathcal{I}_{pen}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) C^{\mathsf{T}}]^{-1} C\widehat{\boldsymbol{\eta}}$$

and the null hypothesis is rejected on the significance level $\alpha$ if the test statistic exceeds the value of the $1 - \alpha$ quantile of the $\chi^2$ distribution with

$$\mathsf{tr}[C\mathcal{I}_{pen}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) C^{\mathsf{T}}]^{-1} C\widehat{V} C^{\mathsf{T}}. \tag{2.20}$$

generalized degrees of freedom. Therneau and Grambsch (2000, p. 121) note that under the null hypothesis the test statistics is asymptotically equivalent to

$$\sum_{\omega=1}^{\nu} \mathsf{ev}_\omega Z_\omega^2$$

where $\mathsf{ev}_\omega$, $\omega = 1, \ldots, \nu$ are the eigenvalues of

$$[C\widehat{\mathcal{I}}_{pen}^{-1}(\widehat{\boldsymbol{\beta}}, \widehat{\gamma}) C^{\mathsf{T}}]^{-1} C\widehat{V} C^{\mathsf{T}}$$

and $Z_\omega$, $\omega = 1, \ldots, \nu$ are independent standard normal random variables. For the classical partial likelihood estimation it holds that $\mathsf{ev}_\omega \in \{0, 1\}$ and the asymptotic distribution of the test statistic is chi-square with $\sum_{\omega=1}^{\nu} \mathsf{ev}_\omega$ degrees of freedom. For the penalized partial likelihood estimation it holds that $\mathsf{ev}_\omega \in [0, 1]$ and while the expected value of the test statistic is still $\sum_{\omega=1}^{\nu} \mathsf{ev}_\omega$ the variance $2 \sum_{\omega=1}^{\nu} \mathsf{ev}_\omega^2$

does not exceed the value $2\sum_{\omega=1}^{\nu} \mathsf{ev}_\omega$ obtained for non-penalized models. Therneau and Grambsch (2000, p. 121) therefore conclude that it is conservative to use the traditional $\chi^2$ limit distribution with $\sum_{\omega=1}^{\nu} \mathsf{ev}_\omega$ generalized degrees of freedom as equivalently expressed in (2.20).

P-splines are implemented in the R function `coxph` from library `survival` under the option `pspline`. The recent addition of the option `tt` (Therneau and Crowson, 2013) theoretically allows to use P-splines as time transformations and construct the partial likelihood ratio test of the proportional hazards. However, the current implementation still does not seem completely reassuring and the test has not been included in the analysis. Some other penalties may be considered as well including those examined by Gray (1992) specifically in the context of the proportional hazards testing where fewer events are likely to appear near the end of the study due to censoring.

It is worth noting that the penalized partial likelihood framework also encompasses models such as ridge regression using identity matrix instead of $P_B$ or $P_P$ and frailty models with Gaussian and gamma frailty terms (Therneau and Grambsch, 2000, p. 123 and 233).

### The derivation of the Lin-Zhang-Davidian test

Lin et al. (2006) point out a common issue related to the spline-based tests described above. They argue that the tests only achieve high power when the smoothing parameter $\theta$ is tweaked with respect to the true alternative. Instead, they propose a test utilizing a frailty models concept which requires no such tuning. Consider the penalized partial likelihood with penalty based on the first derivative

$$\ell_{pen}(\boldsymbol{\beta}, f, \theta) = \ell(\boldsymbol{\beta}, f) - \frac{\theta}{2}\int [f'(t)]^2 \,\mathrm{d}t.$$

Lin et al. (2006) state that the maximum of the penalized partial likelihood is attained for $f$ in the form of the first order smoothing spline which can be parametrized as

$$f(t) = \delta_0 + \sum_{k=1}^{m} \delta_k \min(t, t_k) \tag{2.21}$$

where $\delta_0, \ldots, \delta_m \in \mathbb{R}$. We can rewrite equation (2.21) as

$$\boldsymbol{\varphi} = \delta_0 \mathbf{1} + \Sigma\boldsymbol{\delta}$$

where $\boldsymbol{\varphi} = (f(t_1), \ldots, f(t_m))^\mathsf{T}$, $\mathbf{1} = (1, \ldots, 1)^\mathsf{T}$ and

$$\Sigma = \begin{pmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_m \end{pmatrix}.$$

Setting $\rho = \theta^{-1}$ we can write

$$\ell_{pen}(\boldsymbol{\beta}, \delta_0, \boldsymbol{\delta}, \rho) = \ell[\boldsymbol{\beta}, \boldsymbol{\varphi}(\delta_0, \boldsymbol{\delta})] - \frac{1}{2\rho}\boldsymbol{\delta}^\mathsf{T}\Sigma\boldsymbol{\delta} \tag{2.22}$$

or

$$\mathcal{L}_{pen}(\boldsymbol{\beta}, \delta_0, \boldsymbol{\delta}, \rho) = \mathcal{L}[\boldsymbol{\beta}, \boldsymbol{\varphi}(\delta_0, \boldsymbol{\delta})] \times e^{-\frac{1}{2\rho}\boldsymbol{\delta}^{\mathsf{T}}\Sigma\boldsymbol{\delta}}$$

which can be looked at as the product of the partial likelihood, the density of $\mathsf{N}(\mathbf{0}, \rho\Sigma^{-1})$ and a constant. This suggests $\boldsymbol{\delta}$ can be interpreted as a random vector with the $\mathsf{N}(\mathbf{0}, \rho\Sigma^{-1})$ distribution and $\mathcal{L}_{pen}$ can be interpreted as the conditional penalized partial likelihood given $\boldsymbol{\delta}$. The null hypothesis of the proportional hazards corresponds to $\boldsymbol{\delta} = \mathbf{0}$. We denote

$$\mathbf{U}_{\boldsymbol{\varphi}} = \frac{\partial \ell[\boldsymbol{\beta}, \boldsymbol{\varphi}(\delta_0, \mathbf{0})]}{\partial \boldsymbol{\varphi}} \quad \text{and} \quad V_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = \frac{\partial^2 \ell[\boldsymbol{\beta}, \boldsymbol{\varphi}(\delta_0, \mathbf{0})]}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^{\mathsf{T}}} \tag{2.23}$$

and consider a score type test of

$$H_0 : \boldsymbol{\delta} = \mathbf{0} \quad \text{against} \quad H_0 : \boldsymbol{\delta} \neq \mathbf{0}$$

based on $\widehat{\Phi}_{\boldsymbol{\varphi}} = n^{-1}\widehat{\mathbf{U}}_{\boldsymbol{\varphi}}^{\mathsf{T}}\Sigma\widehat{\mathbf{U}}_{\boldsymbol{\varphi}}$ with the maximum partial likelihood estimators under the null hypothesis $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}_0$ plugged in. We start with a spectral decomposition of $\Sigma$ setting

$$\Phi_{\boldsymbol{\varphi}} = \frac{1}{n}\mathbf{U}_{\boldsymbol{\varphi}}^{\mathsf{T}}\Sigma\mathbf{U}_{\boldsymbol{\varphi}} = \frac{1}{n}\mathbf{U}_{\boldsymbol{\varphi}}^{\mathsf{T}}PDP^{\mathsf{T}}\mathbf{U}_{\boldsymbol{\varphi}} = \sum_{k=1}^{m} d_k \left(\frac{1}{\sqrt{n}}\mathbf{p}_k^{\mathsf{T}}\mathbf{U}_{\boldsymbol{\varphi}}\right)^2$$

where the diagonal elements $d_k$ of $D$ are the eigenvalues of $\Sigma$ and the column vectors $\mathbf{p}_k$ of $P$ are the corresponding eigenvectors. The second equation in (1.8) shows that the $k$-th Schoenfeld residual is the increment of the $p$-dimensional score process at the $k$-th event time. The first equation in (2.23) describes this relationship in $m$-dimensional vector form for $p = 1$ and $k = 1, \ldots, m$. Therefore, $\widehat{\mathbf{U}}_{\boldsymbol{\varphi}}$ is a vector of the Schoenfeld residuals corresponding to the single time-varying covariate. Equations (1.10), (1.11) and (1.12) show that the components of $\mathbf{U}_{\boldsymbol{\varphi}}$ have zero mean and are uncorrelated, with variances given by the diagonal elements of $V_{\boldsymbol{\varphi}\boldsymbol{\varphi}}$. Therefore $n^{-1/2}\mathbf{p}_k^{\mathsf{T}}\mathbf{U}_{\boldsymbol{\varphi}}$ can be interpreted as a weighted sum of large number of uncorrelated increments and the central limit theorem can be used to approximate its distribution as $\mathsf{N}(\mathbf{0}, w_k)$ where $w_k = n^{-1}\mathbf{p}_k^{\mathsf{T}}V_{\boldsymbol{\varphi}\boldsymbol{\varphi}}\mathbf{p}_k$. Hence

$$\widehat{\Phi}_{\boldsymbol{\varphi}} \doteq \sum_{k=1}^{m} d_k w_k \left(\frac{1}{\sqrt{n w_k}}\mathbf{p}_k^{\mathsf{T}}\mathbf{U}_{\boldsymbol{\varphi}}\right)^2$$

is approximately a sum of the weighted $\chi_1^2$ variables. The distribution of the whole sum is approximated using the Satterthwaite method to match the first two moments of the scaled chi-square distribution $c \cdot \chi^2$ with $d$ degrees of freedom (Satterthwaite, 1946). It holds that

$$\mathsf{E}\,\Phi_{\boldsymbol{\varphi}} = \mathsf{tr}\,\mathsf{E}\,\Phi_{\boldsymbol{\varphi}} = \mathsf{E}\,\mathsf{tr}\,\Phi_{\boldsymbol{\varphi}} = \frac{1}{n}\mathsf{E}\,\mathsf{tr}\,\mathbf{U}_{\boldsymbol{\varphi}}^{\mathsf{T}}\Sigma\mathbf{U}_{\boldsymbol{\varphi}} = \frac{1}{n}\mathsf{E}\,\mathsf{tr}\,\mathbf{U}_{\boldsymbol{\varphi}}\mathbf{U}_{\boldsymbol{\varphi}}^{\mathsf{T}}\Sigma = \frac{1}{n}\mathsf{tr}\,V_{\boldsymbol{\varphi}\boldsymbol{\varphi}}\Sigma$$

and Lin et al. (2006) reveal that

$$\mathsf{var}\,\Phi_{\boldsymbol{\varphi}} = \frac{2}{n^2}\mathsf{tr}[(V_{\boldsymbol{\varphi}\boldsymbol{\varphi}}\Sigma)^2].$$

The method of moments gives

$$c = \frac{\mathsf{var}\,\Phi_{\boldsymbol{\varphi}}}{2\mathsf{E}\,\Phi_{\boldsymbol{\varphi}}} = \frac{\mathsf{tr}[(V_{\boldsymbol{\varphi\varphi}}\Sigma)^2]}{n\mathsf{tr}\,V_{\boldsymbol{\varphi\varphi}}\Sigma} \quad \text{and} \quad d = \frac{(2\mathsf{E}\,\Phi_{\boldsymbol{\varphi}})^2}{\mathsf{var}\,\Phi_{\boldsymbol{\varphi}}} = \frac{(\mathsf{tr}\,V_{\boldsymbol{\varphi\varphi}}\Sigma)^2}{\mathsf{tr}[(V_{\boldsymbol{\varphi\varphi}}\Sigma)^2]}$$

which means that the test statistic $\widehat{c}^{-1}\widehat{\Phi}_{\boldsymbol{\varphi}}$ can be approximated with the $\chi^2$ distribution with $\widehat{d}$ degrees of freedom. Furthermore, Lin et al. (2006) proposed a small sample correction for the variance introduced by estimating the regression coefficients using

$$\widehat{c}^\star = \frac{\mathsf{tr}[(\widehat{V}^\star_{\boldsymbol{\varphi\varphi}}\Sigma)^2]}{n\mathsf{tr}\,\widehat{V}^\star_{\boldsymbol{\varphi\varphi}}\Sigma} \quad \text{and} \quad \widehat{d}^\star = \frac{(\mathsf{tr}\,\widehat{V}^\star_{\boldsymbol{\varphi\varphi}}\Sigma)^2}{\mathsf{tr}[(\widehat{V}^\star_{\boldsymbol{\varphi\varphi}}\Sigma)^2]}$$

where

$$\widehat{V}^\star_{\boldsymbol{\varphi\varphi}} = \widehat{V}_{\boldsymbol{\varphi\varphi}} - \begin{pmatrix} \widehat{V}_{\boldsymbol{\varphi\beta}} & \widehat{V}_{\boldsymbol{\varphi\varphi}}\mathbf{1} \end{pmatrix} \begin{pmatrix} \widehat{V}_{\boldsymbol{\beta\beta}} & \widehat{V}_{\boldsymbol{\beta\varphi}}\mathbf{1} \\ \mathbf{1}^\mathsf{T}\widehat{V}_{\boldsymbol{\varphi\beta}} & \mathbf{1}^\mathsf{T}\widehat{V}_{\boldsymbol{\varphi\varphi}}\mathbf{1} \end{pmatrix}^{-1} \begin{pmatrix} \widehat{V}_{\boldsymbol{\beta\varphi}} \\ \mathbf{1}^\mathsf{T}\widehat{V}_{\boldsymbol{\varphi\varphi}} \end{pmatrix}$$

and $\widehat{V}_{\boldsymbol{\beta\beta}}, \widehat{V}_{\boldsymbol{\beta\varphi}}, \widehat{V}_{\boldsymbol{\varphi\beta}}$ are defined by analogy with (2.23) plugging $\widehat{\boldsymbol{\beta}}$ and $\widehat{\delta}_0$ in.

The test is not yet implemented in any R library to our knowledge. Therefore, we have implemented the test in R by ourselves based on the STATA code from the simulation study of Grant et al. (2013). In this article it is tacitly assumed that the variance matrix is time-invariant as was done for the Grambsch-Therneau class of tests in (2.8). This variant has been programmed as well.

# 3. Application to Myocardial Infarction Data

The methods described in the previous chapter are illustrated on medical data. The data are taken from a prospective cohort study of cardiovascular diseases in American patients of age 65 or higher (Fried et al., 1991). The study lasted from 1989 to 2002 with 3 954 patients enrolled. We are interested in testing the proportional hazards assumption rather than making medical conclusions so we select a sub-population of 3 374 patients (385 events) based on data completeness without investigating the origin of the missing data. The objective is to investigate the effect of a set of covariates on the time to myocardial infarction. We build a model including five covariates with highly significant effects (p-values $< 0.001$) on the hazard of myocardial infarction: gender, sub-clinical atherosclerosis, pack-years (the amount of cigarettes smoked throughout life), systolic blood pressure and HDL cholesterol level. The best model selection is not the objective in this case. All covariate values are recorded at baseline. The significance level is set to 0.05.

**Results of the Tests**

The tests of the Grambsch-Therneau class are based on the correlation between the scaled Schoenfeld residuals and the corresponding event times after a suitable transformation. Equivalently, they are based on the slope of the regression line using the scaled Schoenfeld residuals as response and the time transformation as a single covariate, see Figures 3.1 and 3.2. The selection of the time transformation plays a vital role as an inappropriate choice may create a couple of influential observations with a great effect on the slope of the regression line, see Figure 3.1.

The two tests where the logarithmic transformation is applied – the Cox test and the Khodonker-Islam test – prove inadequate, as illustrated for the first one on Figure 3.1. All other tests give similar results. For example, the p-values of the R default Lin test are 0.048 for gender, 0.058 for the systolic blood pressure and higher for others. The combined p-value for all covariates is 0.098. The scaled Schoenfeld residuals corresponding to gender approach zero after about six years and the scaled Schoenfeld residuals corresponding to the systolic blood pressure gradually decrease towards zero suggesting the effects of these covariates diminish over time, see Figure 3.2.

The unweighted version of the Lin-Wei-Ying test rejects the null hypothesis of the proportional hazards for gender (p-value 0.050) but not for the systolic blood pressure (p-value 0.078) or any other covariate, see Figure 3.3. The weighted version of the test does not reject the null hypothesis for any of the covariates (p-values $\geq 0.090$), see Figure 3.4.

The Scheike-Martinussen test is performed within the extended Cox model which assumes the time-varying effect for all covariates and allows to impose the constant effect constraint on covariates one by one. This way we proceed to a model where the effects of sub-clinical atherosclerosis, pack-years and HDL cholesterol are assumed time-fixed and the effects of gender and the systolic blood pressure

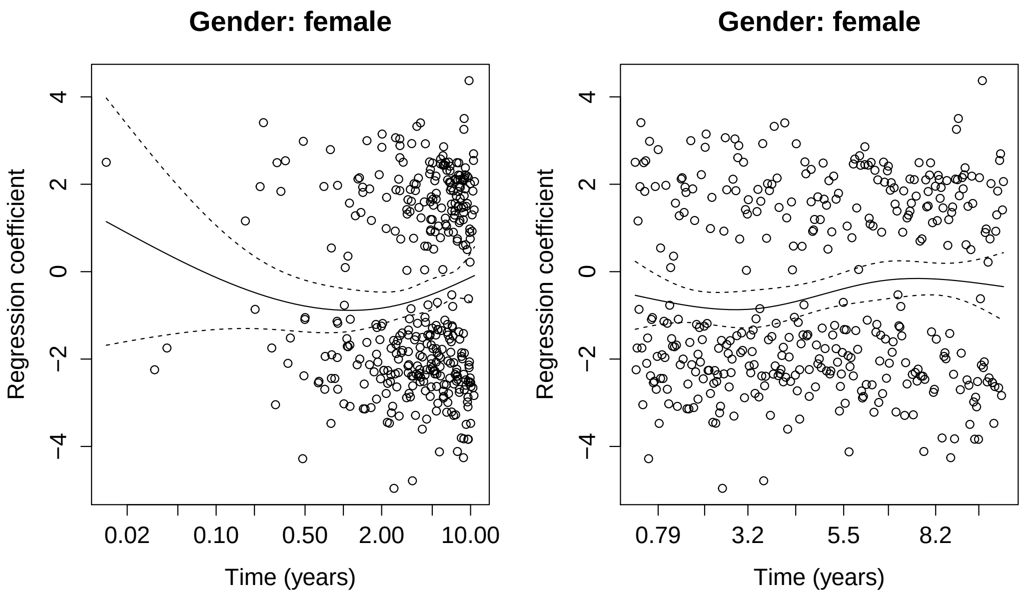


Figure 3.1: Grambsch-Therneau class of tests: the scaled Schoenfeld residuals against the event times after the logarithmic transformation (left plot, p-value 0.266) and the rank transformation (right plot, p-value 0.049). The shape of the regression coefficient is visualized by a spline fit (solid line) and endowed with two-standard-error pointwise confidence bands (dashed lines) as described by Therneau and Grambsch (2000, p. 134–135). The first test is strongly influenced by the three earliest observations.

are left possibly time-varying. The Kolmogorov-Smirnov type test rejects the null hypothesis of the proportional hazards for gender (p-value 0.027) but not for the systolic blood pressure (p-value 0.088) or any other covariate. The Cramér-von Mises type test rejects the null hypothesis for gender (p-value 0.009) and the systolic blood pressure (p-value 0.030) but not for any other covariate.

The original version of the Lin-Zhang-Davidian test rejects the null hypothesis of the proportional hazards for gender (p-value 0.015) but not for the systolic blood pressure (p-value 0.101) or any other covariate. The differences between the original and the constant variance version of the test are small (p-value differences $\leq 0.006$).

#### Summary of the Results

There is no evidence of the violation of the proportional hazards assumption for sub-clinical atherosclerosis, pack-years and HDL cholesterol level. The effects of gender and the systolic blood pressure are revealed to be weakened over time. Since gender is a binary covariate we may compare the results with the common approach of plotting the log-log survival curves in a model stratified with respect to gender. The two curves overlap for the first half-year and then they start to drift apart. Their distance stabilizes after about one year and the curves remain approximately parallel for about five more years after which they start to get closer again, see Figure 3.6. The final approaching of the curves is in line with the findings above. The effect of gender in the first half-year is inconclusive.

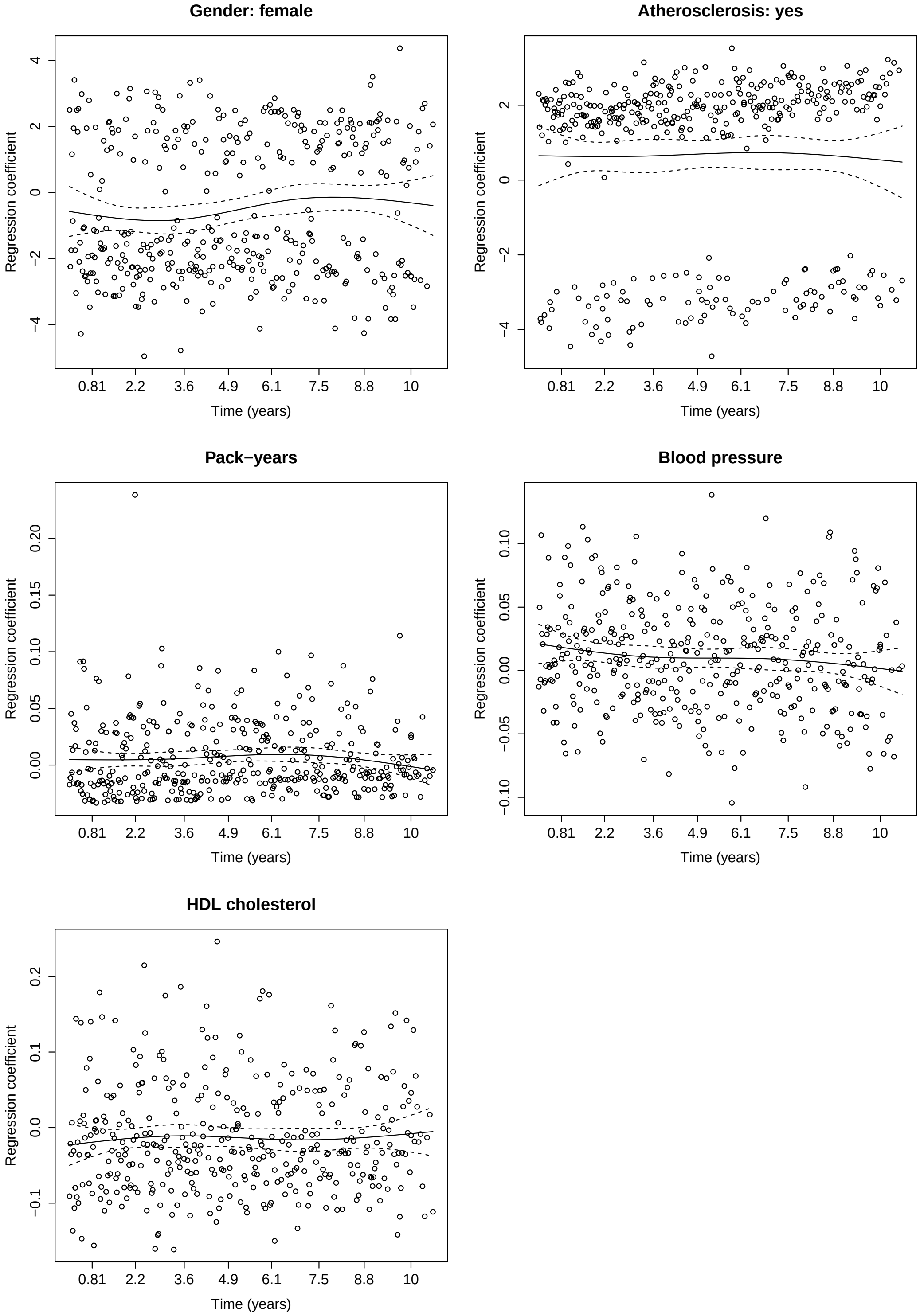


Figure 3.2: Grambsch-Therneau class of tests: the scaled Schoenfeld residuals against the event times after the Kaplan-Meier transformation. The shapes of the regression coefficients are visualized by spline fits (solid lines) and endowed with two-standard-error pointwise confidence bands (dashed lines). The effect of gender nearly vanishes after about six years. The effect of the systolic blood pressure is gradually reduced to zero.

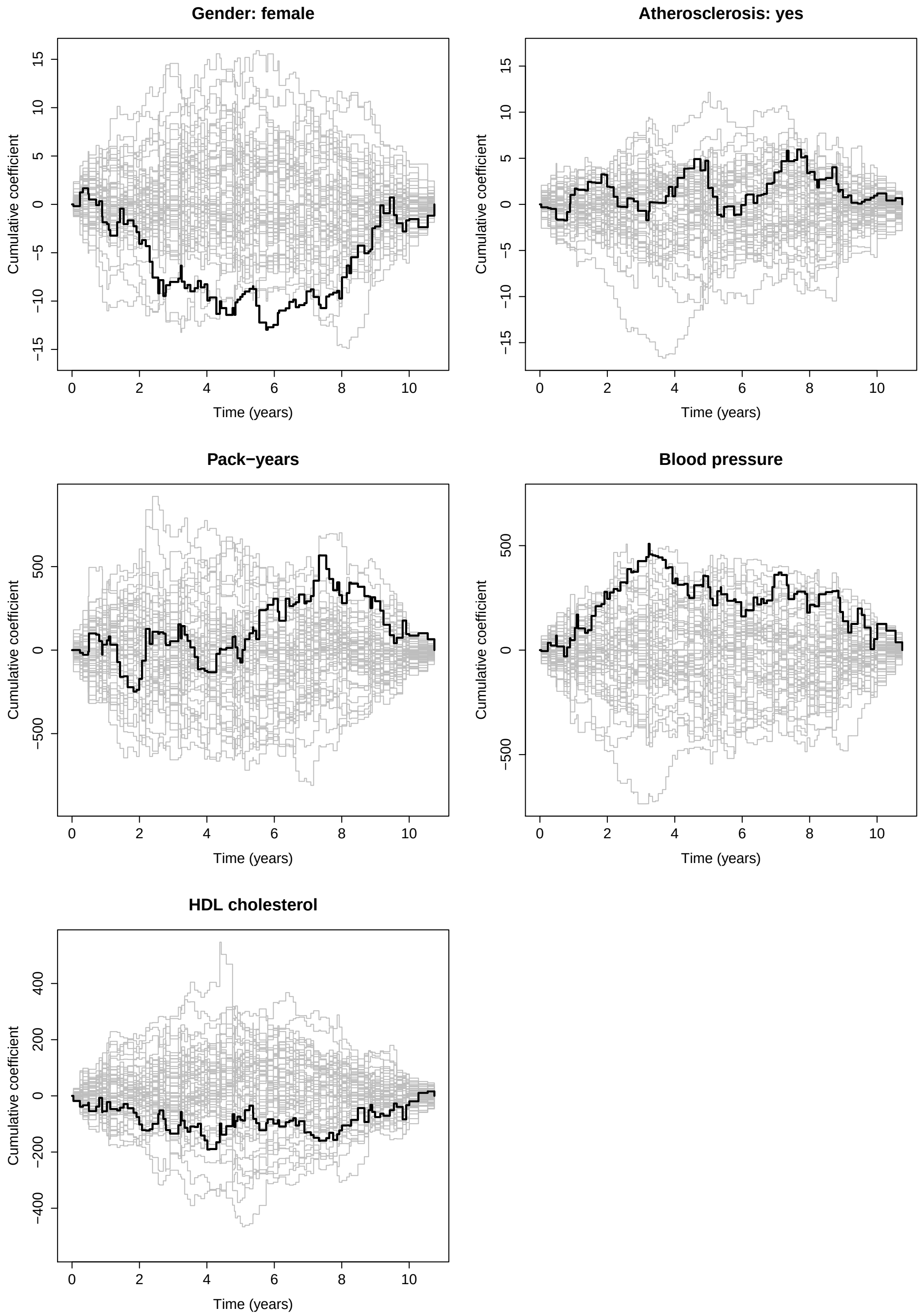


Figure 3.3: Unweighted version of the Lin-Wei-Ying test: Empirical cumulative score process (black) with 50 cumulative score processes simulated under the null hypothesis (gray). No extensively divergent behaviour is apparent. The maximum value for the systolic blood pressure is close to the overall maximum value simulated.

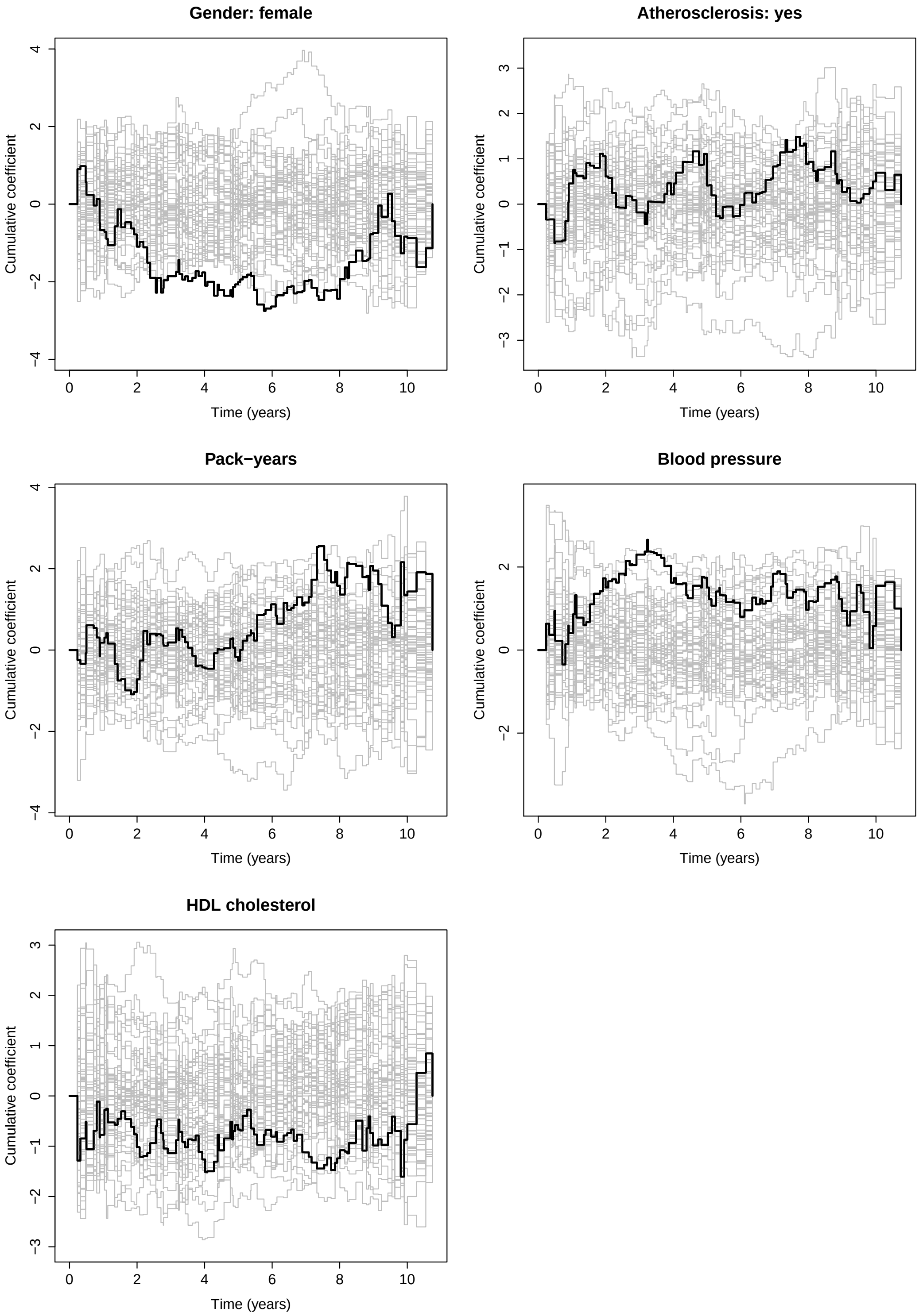


Figure 3.4: Weighted version of the Lin-Wei-Ying test: Empirical cumulative score process (black) with 50 cumulative score processes simulated under the null hypothesis (gray). No extensively divergent behaviour is apparent. The minimum value for gender is close to the overall minimum value simulated.

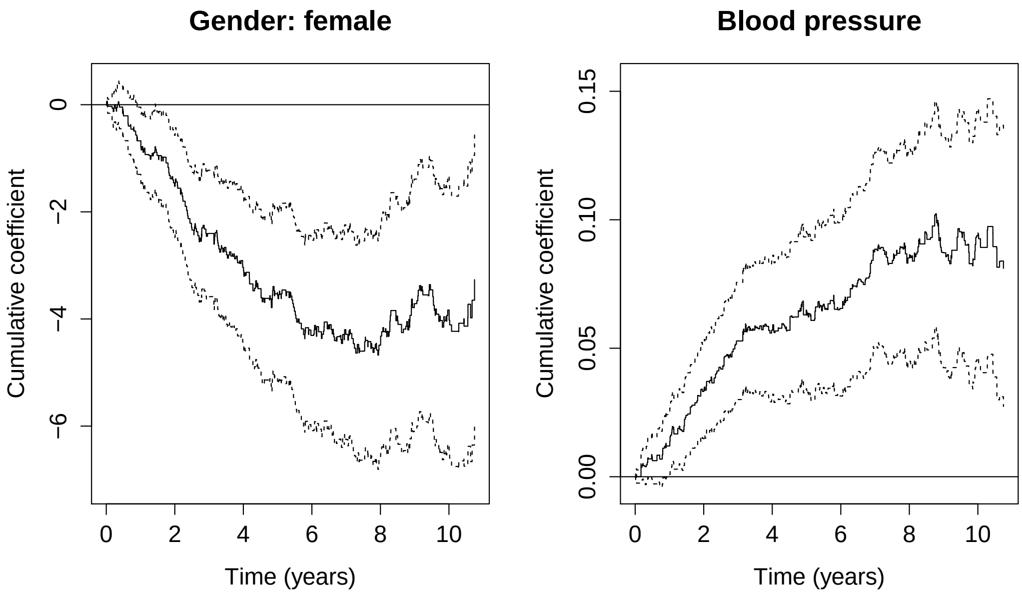


Figure 3.5: Scheike-Martinussen test: Estimates of the cumulative regression coefficients (solid line) endowed with 95 % pointwise confidence bands (dashed lines) as described by Martinussen and Scheike (2006, p. 221). Considerable divergence from linearity is detected suggesting that the effects of gender (left plot; p-values 0.027 [KS], 0.009 [CvM]) and the systolic blood pressure (right plot; p-values 0.088 [KS], 0.030 [CvM]) diminish over time.

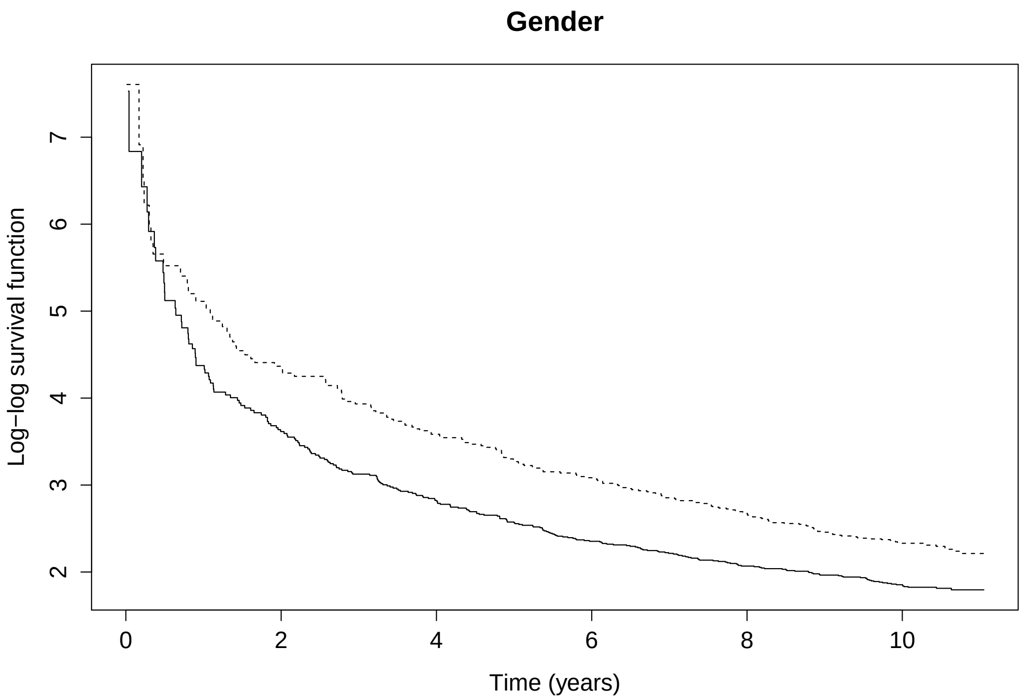


Figure 3.6: Estimated log-log survival curves for men (solid line) and women (dashed line). After one-year stabilization period the curves become approximately parallel indicating little evidence of the non-proportional hazards until about six years. Then the curves start to get closer confirming the lack of fit towards the end of the time range. The effect in the first half-year is inconclusive.

# 4. Simulation Study

## 4.1 Theoretical Considerations

Bender et al. (2005) present the following approach to simulation of the event times. Let $T_i$ be random variables with increasing continuous distribution function $F$. Then the random variables $F(T_i)$ have the uniform distribution on $[0, 1]$ and also

$$1 - F(T_i) = S(T_i) = e^{-\Lambda(T_i)} \sim \mathsf{Unif}(0, 1).$$

Therefore, if the cumulative baseline hazard function is invertible we can use formula (1.5) to simulate the event times in the regression setting as

$$T_i = \Lambda^{-1}(-\log U_i \,|\, \mathbf{X}_i) = \Lambda_0^{-1}(-\log U_i \cdot e^{-\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}})$$

where $U_i \sim \mathsf{Unif}(0, 1)$ or equivalently

$$T_i = \Lambda^{-1}(E_i \,|\, \mathbf{X}_i) = \Lambda_0^{-1}(E_i \cdot e^{-\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}) \tag{4.1}$$

where $E_i \sim \mathsf{Exp}(1)$. Specifically, if $T_i$ have the exponential distribution with rate parameter $\lambda$ then

$$\lambda_0(t) \equiv \lambda, \quad \Lambda_0(t) = \lambda t \quad \text{and} \quad \Lambda_0^{-1}(t) = \frac{t}{\lambda}.$$

Thus, the event times can be simulated as

$$T_i = \frac{E_i}{\lambda e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}}. \tag{4.2}$$

Since the hazard function takes the form $\lambda(t \,|\, \mathbf{X}_i) = \lambda e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}$ the event times can be interpreted as exponentially distributed random variables with rate parameters $\lambda e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}$. More generally, the event times with the Weibull distribution and the baseline hazard $\lambda_0(t) = \lambda\gamma t^{\gamma-1}$ can be simulated as $T_i = (E_i \cdot \lambda^{-1} e^{-\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}})^{1/\gamma}$. In addition, the event times with the Gompertz distribution and the baseline hazard $\lambda_0(t) = \lambda e^{\gamma t}$ can be simulated as

$$T_i = \frac{1}{\gamma} \log\left(1 + \frac{\gamma E_i}{\lambda e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}}\right).$$

### 4.1.1 Time-varying Covariates

Austin (2012) investigates the simulation of the event times in the context of time-varying covariates. An extension of this approach allows to include time-varying covariates $\mathbf{X}_i(t) = (\mathbf{X}_{1i}^\mathsf{T}, \mathbf{X}_{2i}^\mathsf{T}(t))^\mathsf{T} = (\mathbf{X}_{1i}^\mathsf{T}, h(t)\mathbf{X}_{2i}^\mathsf{T})^\mathsf{T}$. To sample the event times with the exponential distribution and linear function $h(t) = t$, it suffices to invert the conditional cumulative hazard function as in (4.1). For $\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta} > 0$ it holds that

$$\begin{aligned}
\Lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) &= \int_0^t \lambda e^{\mathbf{X}_{1i}^\mathsf{T}\boldsymbol{\alpha} + \mathbf{X}_{2i}^\mathsf{T}(s)\boldsymbol{\beta}} \,\mathrm{d}s = \lambda e^{\mathbf{X}_{1i}^\mathsf{T}\boldsymbol{\alpha}} \int_0^t e^{s\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta}} \,\mathrm{d}s \\
&= \lambda e^{\mathbf{X}_{1i}^\mathsf{T}\boldsymbol{\alpha}} \left[\frac{1}{\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta}} e^{s\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta}}\right]_0^t = \frac{1}{\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta}} \lambda e^{\mathbf{X}_{1i}^\mathsf{T}\boldsymbol{\alpha}} (e^{t\mathbf{X}_{2i}^\mathsf{T}\boldsymbol{\beta}} - 1)
\end{aligned}$$

while for $\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} = 0$ we obtain

$$\Lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \int_0^t \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \,\mathrm{d}s = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} t.$$

Hence the event times can be simulated as

$$T_i = \Lambda^{-1}(E_i \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \begin{cases} \frac{1}{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta}} \log\left[1 + \frac{E_i \cdot \mathbf{X}_{2i}^\intercal \boldsymbol{\beta}}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})}\right] & \text{for} \quad \mathbf{X}_{2i}^\intercal \boldsymbol{\beta} > 0, \\ \frac{E_i}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})} & \text{for} \quad \mathbf{X}_{2i}^\intercal \boldsymbol{\beta} = 0. \end{cases}$$

To sample the event times with the exponential distribution and a step-function $h(t) = \mathbb{1}(t > t_b)$ with unit upwards jump at breakpoint time $t_b$ applied to a single covariate we compute

$$\begin{aligned} \Lambda(\,t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) &= \int_0^t \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha} + \beta X_{2i}(s)} \,\mathrm{d}s = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \int_0^t e^{\beta X_{2i}(s)} \,\mathrm{d}s \\ &= \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \left( \int_0^{t_b} \mathrm{d}s + \int_{t_b}^t e^{\beta X_{2i}} \,\mathrm{d}s \right) = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} [t_b + e^{\beta X_{2i}}(t - t_b)] \end{aligned}$$

for $t_b < t$ while formula (4.2) remains valid for $t_b \geq t$. After inverting the conditional cumulative hazard function for $t_b < t$ we calculate the overall formula

$$T_i = \begin{cases} \frac{E_i}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})} & \text{for} \quad \frac{E_i}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})} < t_b, \\ \frac{E_i + \lambda t_b \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})[\exp(\beta X_{2i}) - 1]}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha} + \beta X_{2i})} & \text{for} \quad \frac{E_i}{\lambda \exp(\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha})} \geq t_b. \end{cases}$$

Besides generalizing take the results published by Austin (2012) we also introduce the logarithmic transformation $h(t) = \log t$. For $\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} \neq -1$ it holds that

$$\begin{aligned} \Lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) &= \int_0^t \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha} + \mathbf{X}_{2i}^\intercal(s) \boldsymbol{\beta}} \,\mathrm{d}s = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \int_0^t e^{\log(s) \mathbf{X}_{2i}^\intercal \boldsymbol{\beta}} \,\mathrm{d}s \\ &= \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \int_0^t s^{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta}} \,\mathrm{d}s = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \left[ \frac{s^{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1}}{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1} \right]_0^t = \lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}} \frac{t^{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1}}{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1} \end{aligned}$$

and

$$T_i = \exp\left\{ \frac{1}{\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1} \log\left[ \frac{E_i \cdot (\mathbf{X}_{2i}^\intercal \boldsymbol{\beta} + 1)}{\lambda e^{\mathbf{X}_{1i}^\intercal \boldsymbol{\alpha}}} \right] \right\}.$$

Austin (2012) also shows how to generalize the formula for arbitrarily switching function with range $\{0, 1\}$ and includes the Weibull and Gompertz distributed event times as well.

### 4.1.2 Model Misspecification

A different way to violate the proportional hazards assumption is to generate the event times from the *accelerated failure time model*

$$\lambda(t \,|\, \mathbf{X}_i) = \lambda_0(te^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}})e^{\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}}.$$

One typical choice of the distribution for the baseline hazard is the log-normal distribution which allows to sample the event times as

$$T_i = \exp(\log T_{0i} - \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta})$$

where $\log T_{0i}$ are drawn from the normal distribution. Another model misspecification occurs when the event times are generated by the *additive hazards model*

$$\lambda(t \,|\, \mathbf{X}_i) = \lambda_0(t) + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}$$

introduced by Lin and Ying (1994). In the case of the constant baseline hazard function $\lambda_0(t) \equiv \lambda$ it holds that

$$\Lambda(T_i \,|\, \mathbf{X}_i) = \int_0^{T_i} (\lambda + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta})\,\mathrm{d}s = T_i(\lambda + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}) \sim \mathsf{Exp}(1)$$

and it is possible to sample the event times as

$$T_i = \frac{E_i}{\lambda + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}} \tag{4.3}$$

where $E_i \sim \mathsf{Exp}(1)$. Comparing (4.2) and (4.3) we see that the event times are generated from the exponential distribution with rate parameter $\lambda + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}$ and the proportional hazards assumption is not violated. Therefore, we investigate the case of the linear hazard function. It holds that

$$\Lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) = \int_0^{T_i} (\lambda t + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta})\,\mathrm{d}s = \frac{\lambda}{2} \cdot T_i^2 + \mathbf{X}_i^\mathsf{T}\boldsymbol{\beta} \cdot T_i.$$

It follows that

$$\left(\sqrt{\frac{\lambda}{2}} \cdot T_i + \frac{1}{\sqrt{2\lambda}}\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}\right)^2 = \Lambda(t \,|\, \mathcal{F}_t^{\mathbf{X}_i}) + \frac{1}{2\lambda}(\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta})^2.$$

Hence the positive event times can be sampled via $E_i \sim \mathsf{Exp}(1)$ as

$$T_i = \sqrt{\frac{2}{\lambda} \cdot E_i + \left(\frac{1}{\lambda}\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}\right)^2} - \frac{1}{\lambda}\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}.$$

## 4.2 Simulation Design

Monte Carlo simulations with 1 000 replications are undertaken to evaluate the size and the power of the tests presented above. The first simulation is performed using the standard setting with sample size $n = 250$, all covariates independent, the censoring times independent of covariates and the censoring rate, i.e. the proportion of the observations where no event is recorded, around 30 %. Next, these conditions are varied one by one. We examine the influence of low sample size, correlated covariates, heavy censoring and covariate-dependent censoring. The empirical size is calculated as the proportion of null hypotheses rejected when the null hypothesis is true. The empirical power is calculated as the proportion of null hypotheses rejected under one of the alternatives specified below.

### 4.2.1 Sampling of the Covariate Values

Three covariates are considered, two of them continuous and one discrete. Covariate values $X_{1i}$ are sampled from the normal distribution with mean 1 and standard deviation 0.1 to avoid generating values out of the range $[0, 2]$ (probability $< 10^{-23}$). In the standard setting, covariate values $X_{2i}$ are sampled from the uniform distribution on $[0, 1]$ and covariates $X_{3i}$ are sampled from the Bernoulli distribution with parameter $p_i \equiv 0.5$ independently of all other covariates. Alternatively, the covariates are assumed to be correlated. Covariate values $X_{2i}$ are then sampled from the beta distribution where

$$a_i = 5X_{1i}, \quad b_i = \frac{5}{X_{1i}}, \quad \mathsf{E}(X_{2i} \,|\, X_{1i}) = \frac{a_i}{a_i + b_i} = \frac{1}{1 + \frac{1}{X_{1i}^2}},$$

see Figure 4.1. Covariate values $X_{3i}$ are sampled from the Bernoulli distribution with parameter $p_i = 0.5X_{1i}$. The regression coefficients corresponding to the time-fixed covariates are set to $\beta_1 = \beta_2 = \beta_3 = 0.1$. The linear predictor $\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}$ has empirical mean 0.2 and empirical standard deviation 0.06 for both uncorrelated and correlated covariates based on the additional simulation of 10 000 replications, see Figure 4.2.

### 4.2.2 Sampling of the Event Times

First, the event times are generated from the exponential model with rate parameter 0.25. The covariates are then allowed to depend on time one by one through linear, logarithmic and step-function interactions. The regression coefficients for the time-varying covariates are selected with the aim of attaining the power around 0.5 for which differences among the tests are best revealed, see Table 4.1. The breakpoint in the step-function is set to $t_b = 2.5$ so that it is reached by approximately half of the individuals. Second, the Gompertz proportional hazards model with the baseline hazard function $\lambda_0(t) = 0.25e^{0.1t}$, the log-normal accelerated failure time model with parameters $\mu = 1, \sigma = 0.1$ and the additive hazards model with the linear baseline hazard function $\lambda_0(t) = 0.25t$ are investigated.

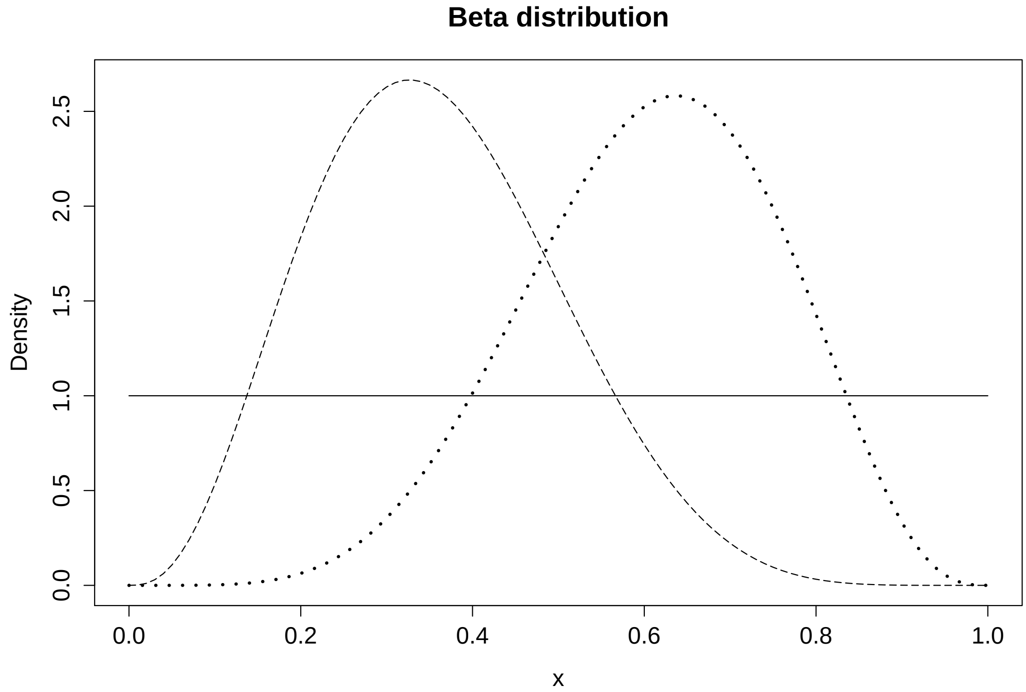


Figure 4.1: Density of the beta distribution for fixed values of $X_{1i}$ set to 0.75 (dashed line) and 1.25 (dotted line) approximately corresponding to the minimum and the maximum values of the first covariate for sample size $n = 250$, respectively. The density of the uniform distribution (solid line) used in the case of independent covariates is a special case of the beta distribution with unit parameters.

### 4.2.3 Sampling of the Censoring Times

The censoring times used in the standard setting are generated from the exponential distribution with rate parameter 0.15 corresponding to the relatively low censoring rate of about 30 %. An alternative option is to set the parameter to 0.5 which corresponds to relatively heavy censoring of about 60 % of the event times. Apart from the censoring times independent of covariate values, we also construct light censoring depending on the first covariate as $C_i \sim \mathsf{Exp}(0.15X_{1i})$.

Table 4.1: Regression coefficients for the time-varying covariates. Transform: the type of the time transformation used in the interaction term; TVC1–TVC3: the regression coefficient for the first–third covariate allowed to depend on time.

| **Transform** | **TVC1** | **TVC2** | **TVC3** |
|---|---|---|---|
| linear | 0.50 | 0.50 | 0.25 |
| step-function | 3.75 | 1.50 | 0.75 |
| logarithmic | 3.25 | 0.50 | 0.25 |

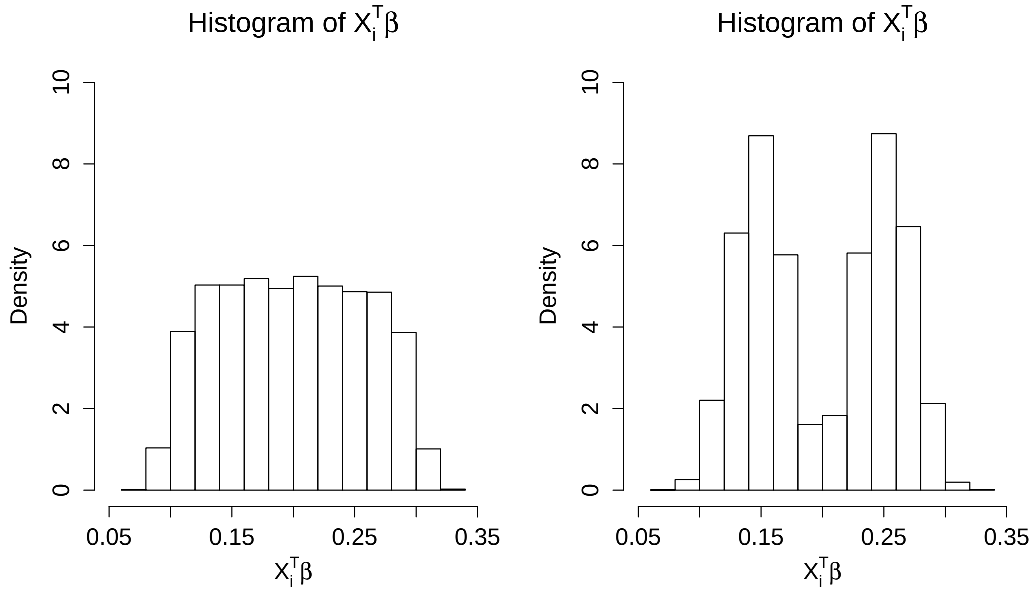


Figure 4.2: Histogram of the linear predictor $\mathbf{X}_i^\mathsf{T}\boldsymbol{\beta}$ in the case of independent covariates (left plot) and correlated covariates (right plot). An additional sample of 10 000 replications was used for its construction.

### 4.2.4 Overview of the Tests under Study

The ability to reveal the violation of the proportional hazards assumption is investigated for the following tests:

1. Grambsch-Therneau class of tests as implemented in R library `survival` within the function `cox.zph`
   (a) Cox test (`transform = "identity"`)
   (b) Cox test (`transform = "log"`)
   (c) Breslow-Edler-Berger test (`transform = "rank"`)
   (d) Lin test (`transform = "km"`)
   (e) Lin test (`transform = "km(1-km)"`)[1]
   (f) Khondoker-Islam test (`transform = "log(-log(km))"`)[1]
2. Lin-Wei-Ying test as implemented in R library `timereg` within the function `cox.aalen`
   (a) Unweighted version of the test
   (b) Weighted version of the test
3. Lin-Zhang-Davidian test self-implemented in R[2]
   (a) The original version of the test
   (b) Constant variance version of the test

[1]Self-implemented extension of the function `cox.zph`.
[2]Implementation based on the STATA code from the simulation study of Grant et al. (2013).

The Scheike-Martinussen test as implemented in R library `timereg` within the function `timecox` was excluded from the study since it repeatedly encountered numerical problems. The same issue was reported by Buchholz (2010) who suggested to restrict the time range used for fitting to $[0, \tau_{max}]$ for some $\tau_{max} < \tau$. Unfortunately, this approach has not solved the difficulties to such an extent that the simulation would be plausible. It was noted by the authors of the test (Scheike and Martinussen, 2004, p. 177) that it might indeed be difficult to fit the extended Cox model for small sample sizes with all regression coefficients possibly time-varying. They recommend setting some of the regression coefficients as time-fixed in such cases which has not removed the problem either.

## 4.3 Simulation Results

The complete simulation results are summarized in the Appendix. The tables of results are divided into the first part investigating the time-varying covariate effects, see Tables A.1–A.10 and the second part investigating the model misspecification, see Tables A.11–A.15.

### 4.3.1 Test-specific Results

The weighted version of the Lin-Wei-Ying test does not have valid size for sample size $n = 100$, see Tables A.3, A.4 and A.12. It also fails to attain valid size for the third covariate with sample size $n = 250$, see e.g. Tables A.1, A.2 and A.11. Therefore, it is not considered for the subsequent power analysis. The test has valid size for large sample sizes though. Using the standard setting with enlarged sample size $n = 1\,000$ the empirical sizes under the exponential model are 0.038, 0.063 and 0.076 respectively for each of the covariates. All tests under study apart from the weighted version of the Lin-Wei-Ying test have valid sizes.

The quadratic variant of the Lin test has extremely low power compared to other tests from the Grambsch-Therneau class and cannot be recommended unless there is a prior indication of quadratic departure from the proportionality of hazards. The Cox test with no time transformation appears to be more conservative than other tests from the Grambsch-Therneau class apart from the above mentioned version of the Lin test. The constant-variance version of the Lin-Zhang-Davidian test turns out to be more conservative than the original version which means that the approximation does reduce power.

The most powerful test is always one of the following: the original version of the Lin-Zhang-Davidian test (32 times), the Khondoker-Islam test (20 times), the unweighted version of the Lin-Wei-Ying test (17 times) and the Cox test with the logarithmic transformation (9 times). The highest power among the tests of the Grambsch-Therneau class is attained by the two tests based on the logarithmic transformation – the Cox test and the Khodonker-Islam test. Nevertheless, it was suggested in Chapter 3 that these two tests are the least reliable ones, see Figure 3.1. The last two tests from the Grambsch-Therneau class with the rank transformation and the Kaplan-Meier transformation typically have only slightly lower power than the overall maximum attained.

### 4.3.2 Alternative-specific Results

The most powerful test for detecting the time-varying covariate effect is the Lin-Zhang-Davidian test in most cases (23 of 45 times). The most powerful test for detecting the accelerated failure time model misspecification is the Khondoker-Islam test in most cases (7 of 15 times) followed by the Lin-Zhang-Davidian test (6 times). The most powerful test for detecting the additive hazards model misspecification is the unweighted version of the Lin-Wei-Ying test in most cases (8 of 15 times), yet it should be noted that no parameter values yielding powers of any practical relevance were found in the current setting even after inclusion of the intercept in the linear baseline hazard function. Similarly, the range of the first covariate is strongly restricted, see Subsection 4.2.1, and the power of all tests is very low when sampling the event times in the presence of linear interaction of the first covariate with time.

### 4.3.3 Design-specific Results

The results for various simulation design alterations are contrasted with the results for the standard setting, see Tables A.1, A.2 and A.11. The immediate consequence of decreasing the sample size is the loss of power. The decline is about one half of the power under the standard setting irrespective of the test type, see Tables A.3, A.4 and A.12. The results for correlated covariates suggest that the higher the correlation between covariates the higher the loss of power. The first covariate is fixed and its correlation with the second covariate is about 0.3. The power of the tests for the time-invariance of the second covariate is mostly ranging between one third and one half of the power under the standard setting. In contrast, the correlation between the first and the third covariate is about 0.1 and there is hardly any difference in power, see Tables A.5, A.6 and A.13. The effect of heavy censoring is of the same type as the effect of lowering the sample size as it effectively reduces the number of the observed events, see Tables A.7, A.8 and A.14. Censoring dependent on the first covariate has no visible effect on the power of the tests, see Tables A.9, A.10 and A.15.

# Conclusion

The proportional hazards model is introduced and the proportional hazards assumption is explained in detail. The tests of the Grambsch-Therneau class as well as the Lin-Wei-Ying test, the Scheike-Martinussen test and the Lin-Zhang-Davidian test are derived and described. The tests are motivated with both historical development and their benefits over other tests whenever possible. The appearance of the potential connections between the tests is accentuated by adhering to unified terminology and notation.

The application to medical data illustrates the tests along with the graphical diagnostics which are shown to provide a valuable insight into the mechanism based on which the test statistics are calculated. They might reveal the way in which the proportional hazards assumption is violated, see Figure 3.5, or they might hint at the reliability of the test, see Figure 3.1.

The simulation study suggests that the Lin-Zhang-Davidian test, the unweighted version of the Lin-Wei-Ying test and the two tests of the Grambsch-Therneau class based on the logarithmic transformation – the Cox test and the Khodonker-Islam test – have superior power. Nevertheless, the latter two tests should only be used when supported by the graphical diagnostics because of their undue sensitivity to influential observations, see Figure 3.1. The Lin test with the Kaplan-Meier transformation is more reliable in this aspect (see also Grambsch and Therneau, 1994, p. 136) and while it does not provide the maximum power it is always close enough to be considered a reasonable default option. It was further revealed that the weighted version of the Lin-Wei-Ying test may not have valid size for sample sizes up to $n = 250$ and should not be used in such cases. The applicability of the Scheike-Martinussen test depends on whether the sample size is large enough to fit the extended Cox model (2.17).

The parameters in the simulation study are set with the aim of revealing the differences in the power of the tests to the fullest extent. The simulation results are therefore limited to the test comparisons and are not supposed to be used to assess whether the power of a certain test would be large enough in a real-life setting. Simulation results concerning the practical applicability of the tests are provided by Grant et al. (2013).

There are aspects of the topic beyond the scope of the thesis. For example, there might be more than one variable violating the proportional hazards assumption at a time. From the vast number of possible alterations in the simulation design we highlight the possibility to test for a specified alternative to the constant treatment effect such as gradual decline, delayed effect, switching effect etc.

Several other tests of the proportional hazards have been proposed some of which are summarized by Buchholz (2010) including *fractional polynomial models* (Sauerbrei et al., 2007), *reduced rank models* (Perperoglou et al., 2006) and methods based on *local linear estimation* (Cai and Sun, 2003) and *average hazard ratios* (Schemper et al., 2009). Kraus (2007) applied the idea of *data-driven smooth tests* (Ledwina, 1994) in the context of the proportional hazards testing.

# A. Appendix

## A.1 Probability Theory

We give a brief overview of the basic terms related to the counting processes and martingale theory. The exposition closely follows the framework introduced by Fleming and Harrington (1991). The terminology is related to the interval $[0, \infty)$.

**Definition A.1.** A class of $\sigma$-algebras $\{\mathcal{F}_t, t \geq 0\}$ satisfying $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s \leq t$ is called a *filtration.* For filtration $\mathcal{F}_t$ we define

1. $\mathcal{F}_{t+} = \bigcap_{h>0} \mathcal{F}_{t+h}$,
2. $\mathcal{F}_{t-} = \sigma\left(\bigcup_{h>0} \mathcal{F}_{t-h}\right)$.

**Definition A.2.** A family of random variables $\{X(t), t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is called a *stochastic process.*

**Definition A.3.** A stochastic process $\{X(t), t \geq 0\}$ where $X(t)$ is $\mathcal{F}_t$-measurable for all $t \geq 0$ is called *adapted to $\mathcal{F}_t$.*

**Definition A.4.** A stochastic process $\{X(t), t \geq 0\}$ adapted to $\mathcal{F}_{t+}$ is called a *right-continuous process.*

**Definition A.5.** A right-continuous process with limits from the left adapted to $\mathcal{F}_t$ satisfying $\mathsf{E}|X(t)| < \infty$ for all $t < \infty$ and one of the following conditions:

1. $\mathsf{E}[X(t+s) \,|\, \mathcal{F}_t] = X(t)$ a.s. for all $s, t \geq 0$,
2. $\mathsf{E}[X(t+s) \,|\, \mathcal{F}_t] \geq X(t)$ a.s. for all $s, t \geq 0$,
3. $\mathsf{E}[X(t+s) \,|\, \mathcal{F}_t] \leq X(t)$ a.s. for all $s, t \geq 0$,

is called a *$\mathcal{F}_t$-martingale*, a *$\mathcal{F}_t$-submartingale* or a *$\mathcal{F}_t$-supermartingale* respectively.

**Definition A.6.** A stochastic process $\{X(t), t \geq 0\}$ where $X(t)$ is $\mathcal{F}_{t-}$-measurable for all $t \geq 0$ is called a *$\mathcal{F}_t$-predictable process.*

**Theorem A.1** (Doob-Meyer Theorem)**.** *Let $\{N(t), t \geq 0\}$ be a right-continuous non-negative $\mathcal{F}_t$-submartingale satisfying $N(0) = 0$ and $\mathsf{E}N(t) < \infty$ for all $t \geq 0$. Then there is a unique decomposition into a right-continuous $\mathcal{F}_t$-martingale $\{M(t), t \geq 0\}$ and an increasing right-continuous $\mathcal{F}_t$-predictable process $\{A(t), t \geq 0\}$ satisfying $A(0) = 0$ a.s. and $\mathsf{E}A(t) < \infty$ such that*

$$N(t) = M(t) + A(t) \quad a.s.$$

*for all $t \geq 0$.* The stochastic process $\{A(t), t \geq 0\}$ is called a *compensator* of submartingale $\{N(t), t \geq 0\}$.

## A.2 Tables of Simulation Results

The complete results of the simulation study are provided, see Tables A.1–A.15.

Table A.1: Results of the simulation study for the time-varying covariate effects under the standard setting. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (exp): the size of the test under the exponential model; TVC1–TVC3 (linear/step/log): the power or the size (in italics) of the test under the exponential model with the first–third covariate allowed to depend on time through linear/step/log interaction.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT (a) | 1st | *0.033* | 0.089 | 0.489 | 0.247 | *0.019* | *0.022* | *0.028* | *0.022* | *0.028* | *0.033* |
| | 2nd | *0.036* | *0.040* | *0.048* | *0.039* | 0.452 | 0.477 | 0.247 | *0.025* | *0.022* | *0.032* |
| | 3rd | *0.026* | *0.042* | *0.047* | *0.032* | *0.026* | *0.027* | *0.024* | 0.356 | 0.320 | 0.221 |
| GT (b) | 1st | *0.049* | 0.098 | 0.314 | 0.280 | *0.040* | *0.049* | *0.051* | *0.044* | *0.049* | *0.059* |
| | 2nd | *0.047* | *0.058* | *0.046* | *0.046* | 0.453 | 0.385 | 0.436 | *0.040* | *0.052* | *0.052* |
| | 3rd | *0.051* | *0.049* | *0.052* | *0.038* | *0.050* | *0.056* | *0.052* | 0.444 | 0.297 | 0.402 |
| GT (c) | 1st | *0.057* | 0.105 | 0.501 | 0.246 | *0.038* | *0.043* | *0.050* | *0.040* | *0.050* | *0.050* |
| | 2nd | *0.049* | *0.050* | *0.038* | *0.050* | 0.562 | 0.581 | 0.411 | *0.040* | *0.038* | *0.055* |
| | 3rd | *0.058* | *0.052* | *0.042* | *0.039* | *0.047* | *0.050* | *0.052* | 0.545 | 0.432 | 0.365 |
| GT (d) | 1st | *0.050* | 0.100 | 0.513 | 0.243 | *0.027* | *0.042* | *0.050* | *0.039* | *0.047* | *0.050* |
| | 2nd | *0.047* | *0.047* | *0.041* | *0.049* | 0.562 | 0.601 | 0.381 | *0.036* | *0.036* | *0.052* |
| | 3rd | *0.051* | *0.049* | *0.038* | *0.039* | *0.044* | *0.049* | *0.047* | 0.547 | 0.443 | 0.356 |
| GT (e) | 1st | *0.051* | 0.051 | 0.062 | 0.077 | *0.051* | *0.048* | *0.052* | *0.050* | *0.052* | *0.054* |
| | 2nd | *0.047* | *0.047* | *0.051* | *0.051* | 0.104 | 0.096 | 0.224 | *0.044* | *0.055* | *0.051* |
| | 3rd | *0.056* | *0.044* | *0.039* | *0.057* | *0.048* | *0.047* | *0.057* | 0.117 | 0.085 | 0.208 |
| GT (f) | 1st | *0.048* | 0.104 | 0.426 | 0.280 | *0.045* | *0.051* | *0.059* | *0.041* | *0.051* | *0.061* |
| | 2nd | *0.046* | *0.055* | *0.041* | *0.053* | 0.471 | 0.436 | 0.438 | *0.038* | *0.054* | *0.053* |
| | 3rd | *0.055* | *0.046* | *0.042* | *0.042* | *0.048* | *0.055* | *0.046* | 0.449 | 0.330 | 0.399 |

Table A.2: Continuation of Table A.1. The censoring rate gives the proportion of the observations where no event is recorded.

| Test | Cov | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LWY (a) | $1^{\text{st}}$ | *0.058* | 0.115 | 0.591 | 0.227 | *0.052* | *0.060* | *0.057* | *0.058* | *0.061* | *0.053* |
| | $2^{\text{nd}}$ | *0.059* | *0.058* | *0.054* | *0.058* | 0.535 | 0.659 | 0.339 | *0.051* | *0.051* | *0.072* |
| | $3^{\text{rd}}$ | *0.048* | *0.055* | *0.059* | *0.049* | *0.059* | *0.055* | *0.063* | 0.499 | 0.511 | 0.335 |
| LWY (b) | $1^{\text{st}}$ | *0.035* | 0.084 | 0.472 | 0.187 | *0.036* | *0.043* | *0.037* | *0.045* | *0.038* | *0.026* |
| | $2^{\text{nd}}$ | *0.042* | *0.044* | *0.035* | *0.047* | 0.511 | 0.579 | 0.282 | *0.043* | *0.051* | *0.055* |
| | $3^{\text{rd}}$ | *0.215* | *0.283* | *0.320* | *0.327* | *0.281* | *0.249* | *0.249* | 0.784 | 0.632 | 0.465 |
| LZD (a) | $1^{\text{st}}$ | *0.053* | 0.114 | 0.541 | 0.266 | *0.034* | *0.050* | *0.061* | *0.050* | *0.053* | *0.046* |
| | $2^{\text{nd}}$ | *0.051* | *0.054* | *0.047* | *0.054* | 0.608 | 0.665 | 0.367 | *0.045* | *0.043* | *0.057* |
| | $3^{\text{rd}}$ | *0.055* | *0.055* | *0.049* | *0.042* | *0.050* | *0.049* | *0.053* | 0.609 | 0.532 | 0.340 |
| LZD (b) | $1^{\text{st}}$ | *0.041* | 0.098 | 0.539 | 0.248 | *0.024* | *0.040* | *0.049* | *0.041* | *0.042* | *0.038* |
| | $2^{\text{nd}}$ | *0.039* | *0.045* | *0.047* | *0.048* | 0.525 | 0.620 | 0.324 | *0.036* | *0.038* | *0.039* |
| | $3^{\text{rd}}$ | *0.038* | *0.047* | *0.048* | *0.039* | *0.044* | *0.041* | *0.033* | 0.506 | 0.464 | 0.306 |
| **Censoring rate** | | 0.330 | 0.219 | 0.235 | 0.227 | 0.263 | 0.294 | 0.324 | 0.300 | 0.318 | 0.333 |

Table A.3: Results of the simulation study for the time-varying covariate effects with low sample size. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (exp): the size of the test under the exponential model; TVC1–TVC3 (linear/step/log): the power or the size (in italics) of the test under the exponential model with the first–third covariate allowed to depend on time through linear/step/log interaction.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT (a) | $1^{st}$ | *0.021* | 0.034 | 0.197 | 0.101 | *0.028* | *0.022* | *0.018* | *0.021* | *0.018* | *0.021* |
| | $2^{nd}$ | *0.015* | *0.032* | *0.043* | *0.037* | 0.126 | 0.154 | 0.075 | *0.020* | *0.023* | *0.014* |
| | $3^{rd}$ | *0.020* | *0.034* | *0.058* | *0.042* | *0.020* | *0.022* | *0.027* | 0.103 | 0.094 | 0.076 |
| GT (b) | $1^{st}$ | *0.057* | 0.076 | 0.157 | 0.133 | *0.054* | *0.051* | *0.045* | *0.056* | *0.046* | *0.055* |
| | $2^{nd}$ | *0.036* | *0.058* | *0.059* | *0.044* | 0.198 | 0.195 | 0.182 | *0.048* | *0.046* | *0.037* |
| | $3^{rd}$ | *0.038* | *0.057* | *0.069* | *0.051* | *0.040* | *0.047* | *0.053* | 0.219 | 0.144 | 0.195 |
| GT (c) | $1^{st}$ | *0.049* | 0.057 | 0.188 | 0.110 | *0.048* | *0.040* | *0.047* | *0.041* | *0.041* | *0.054* |
| | $2^{nd}$ | *0.036* | *0.041* | *0.036* | *0.046* | 0.239 | 0.246 | 0.161 | *0.044* | *0.054* | *0.033* |
| | $3^{rd}$ | *0.036* | *0.052* | *0.047* | *0.044* | *0.037* | *0.046* | *0.051* | 0.240 | 0.184 | 0.169 |
| GT (d) | $1^{st}$ | *0.040* | 0.054 | 0.186 | 0.103 | *0.043* | *0.039* | *0.038* | *0.036* | *0.037* | *0.047* |
| | $2^{nd}$ | *0.033* | *0.041* | *0.037* | *0.044* | 0.231 | 0.242 | 0.138 | *0.043* | *0.049* | *0.026* |
| | $3^{rd}$ | *0.029* | *0.052* | *0.046* | *0.041* | *0.032* | *0.041* | *0.051* | 0.226 | 0.186 | 0.154 |
| GT (e) | $1^{st}$ | *0.056* | 0.072 | 0.058 | 0.076 | *0.056* | *0.051* | *0.053* | *0.065* | *0.053* | *0.054* |
| | $2^{nd}$ | *0.043* | *0.052* | *0.053* | *0.042* | 0.084 | 0.085 | 0.125 | *0.057* | *0.063* | *0.051* |
| | $3^{rd}$ | *0.046* | *0.066* | *0.049* | *0.034* | *0.059* | *0.050* | *0.059* | 0.103 | 0.089 | 0.128 |
| GT (f) | $1^{st}$ | *0.055* | 0.067 | 0.176 | 0.140 | *0.055* | *0.051* | *0.051* | *0.050* | *0.056* | *0.051* |
| | $2^{nd}$ | *0.040* | *0.050* | *0.047* | *0.047* | 0.216 | 0.208 | 0.181 | *0.049* | *0.049* | *0.034* |
| | $3^{rd}$ | *0.040* | *0.060* | *0.055* | *0.047* | *0.048* | *0.046* | *0.052* | 0.232 | 0.162 | 0.178 |

Table A.4: Continuation of Table A.3. The censoring rate gives the proportion of the observations where no event is recorded.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LWY (a) | 1st | *0.071* | 0.082 | 0.247 | 0.118 | *0.062* | *0.064* | *0.064* | *0.071* | *0.060* | *0.060* |
| | 2nd | *0.051* | *0.056* | *0.064* | *0.058* | 0.272 | 0.316 | 0.162 | *0.051* | *0.067* | *0.050* |
| | 3rd | *0.057* | *0.060* | *0.072* | *0.055* | *0.059* | *0.057* | *0.068* | 0.257 | 0.223 | 0.182 |
| LWY (b) | 1st | *0.275* | 0.468 | 0.572 | 0.525 | *0.329* | *0.340* | *0.325* | *0.287* | *0.283* | *0.310* |
| | 2nd | *0.303* | *0.543* | *0.551* | *0.550* | 0.523 | 0.545 | 0.436 | *0.320* | *0.329* | *0.340* |
| | 3rd | *0.413* | *0.574* | *0.583* | *0.559* | *0.472* | *0.448* | *0.448* | 0.619 | 0.573 | 0.515 |
| LZD (a) | 1st | *0.049* | 0.081 | 0.224 | 0.138 | *0.059* | *0.052* | *0.053* | *0.047* | *0.048* | *0.062* |
| | 2nd | *0.058* | *0.048* | *0.059* | *0.051* | 0.289 | 0.320 | 0.165 | *0.055* | *0.058* | *0.040* |
| | 3rd | *0.045* | *0.059* | *0.058* | *0.052* | *0.040* | *0.048* | *0.061* | 0.294 | 0.238 | 0.174 |
| LZD (b) | 1st | *0.032* | 0.048 | 0.215 | 0.110 | *0.038* | *0.042* | *0.034* | *0.034* | *0.034* | *0.042* |
| | 2nd | *0.026* | *0.032* | *0.055* | *0.046* | 0.205 | 0.248 | 0.112 | *0.029* | *0.042* | *0.019* |
| | 3rd | *0.024* | *0.042* | *0.055* | *0.044* | *0.029* | *0.033* | *0.038* | 0.189 | 0.166 | 0.121 |
| **Censoring rate** | | 0.330 | 0.218 | 0.236 | 0.227 | 0.263 | 0.295 | 0.322 | 0.299 | 0.316 | 0.330 |

Table A.5: Results of the simulation study for the time-varying covariate effects with correlated covariates. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (exp): the size of the test under the exponential model; TVC1–TVC3 (linear/step/log): the power or the size (in italics) of the test under the exponential model with the first–third covariate allowed to depend on time through linear/step/log interaction.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT (a) | $1^{\text{st}}$ | *0.024* | 0.077 | 0.460 | 0.192 | *0.032* | *0.025* | *0.027* | *0.020* | *0.029* | *0.032* |
| | $2^{\text{nd}}$ | *0.024* | *0.032* | *0.049* | *0.050* | 0.168 | 0.138 | 0.084 | *0.021* | *0.035* | *0.029* |
| | $3^{\text{rd}}$ | *0.021* | *0.030* | *0.055* | *0.052* | *0.034* | *0.026* | *0.030* | 0.406 | 0.314 | 0.252 |
| GT (b) | $1^{\text{st}}$ | *0.054* | 0.096 | 0.284 | 0.234 | *0.053* | *0.047* | *0.050* | *0.046* | *0.064* | *0.058* |
| | $2^{\text{nd}}$ | *0.037* | *0.049* | *0.048* | *0.063* | 0.183 | 0.145 | 0.150 | *0.053* | *0.057* | *0.044* |
| | $3^{\text{rd}}$ | *0.042* | *0.043* | *0.053* | *0.055* | *0.051* | *0.043* | *0.051* | 0.451 | 0.304 | 0.439 |
| GT (c) | $1^{\text{st}}$ | *0.043* | 0.097 | 0.472 | 0.185 | *0.044* | *0.038* | *0.037* | *0.038* | *0.051* | *0.054* |
| | $2^{\text{nd}}$ | *0.038* | *0.038* | *0.030* | *0.057* | 0.217 | 0.190 | 0.134 | *0.043* | *0.056* | *0.047* |
| | $3^{\text{rd}}$ | *0.052* | *0.038* | *0.052* | *0.054* | *0.045* | *0.042* | *0.049* | 0.580 | 0.432 | 0.390 |
| GT (d) | $1^{\text{st}}$ | *0.040* | 0.097 | 0.486 | 0.177 | *0.039* | *0.037* | *0.036* | *0.033* | *0.046* | *0.053* |
| | $2^{\text{nd}}$ | *0.035* | *0.036* | *0.027* | *0.056* | 0.214 | 0.199 | 0.136 | *0.037* | *0.054* | *0.045* |
| | $3^{\text{rd}}$ | *0.044* | *0.036* | *0.051* | *0.054* | *0.043* | *0.044* | *0.047* | 0.588 | 0.439 | 0.369 |
| GT (e) | $1^{\text{st}}$ | *0.054* | 0.064 | 0.058 | 0.090 | *0.049* | *0.046* | *0.056* | *0.049* | *0.056* | *0.046* |
| | $2^{\text{nd}}$ | *0.048* | *0.046* | *0.046* | *0.058* | 0.062 | 0.063 | 0.097 | *0.060* | *0.058* | *0.048* |
| | $3^{\text{rd}}$ | *0.045* | *0.037* | *0.037* | *0.041* | *0.052* | *0.061* | *0.045* | 0.126 | 0.087 | 0.198 |
| GT (f) | $1^{\text{st}}$ | *0.054* | 0.089 | 0.372 | 0.236 | *0.048* | *0.042* | *0.054* | *0.043* | *0.062* | *0.059* |
| | $2^{\text{nd}}$ | *0.035* | *0.039* | *0.038* | *0.065* | 0.199 | 0.151 | 0.153 | *0.049* | *0.055* | *0.042* |
| | $3^{\text{rd}}$ | *0.042* | *0.043* | *0.052* | *0.050* | *0.048* | *0.040* | *0.050* | 0.474 | 0.318 | 0.433 |

Table A.6: Continuation of Table A.5. The censoring rate gives the proportion of the observations where no event is recorded.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LWY (a) | 1st | *0.063* | 0.110 | 0.607 | 0.217 | *0.073* | *0.063* | *0.053* | *0.050* | *0.061* | *0.055* |
| | 2nd | *0.043* | *0.047* | *0.101* | *0.070* | 0.236 | 0.248 | 0.145 | *0.062* | *0.064* | *0.047* |
| | 3rd | *0.046* | *0.064* | *0.067* | *0.069* | *0.055* | *0.058* | *0.060* | 0.568 | 0.512 | 0.324 |
| LWY (b) | 1st | *0.037* | 0.087 | 0.470 | 0.159 | *0.050* | *0.036* | *0.041* | *0.031* | *0.042* | *0.033* |
| | 2nd | *0.025* | *0.046* | *0.069* | *0.064* | 0.181 | 0.189 | 0.118 | *0.042* | *0.045* | *0.025* |
| | 3rd | *0.229* | *0.275* | *0.331* | *0.336* | *0.305* | *0.257* | *0.249* | 0.811 | 0.630 | 0.487 |
| LZD (a) | 1st | *0.050* | 0.117 | 0.546 | 0.233 | *0.066* | *0.067* | *0.049* | *0.045* | *0.059* | *0.058* |
| | 2nd | *0.038* | *0.048* | *0.094* | *0.082* | 0.284 | 0.263 | 0.149 | *0.047* | *0.061* | *0.054* |
| | 3rd | *0.052* | *0.047* | *0.055* | *0.061* | *0.047* | *0.044* | *0.057* | 0.661 | 0.535 | 0.369 |
| LZD (b) | 1st | *0.040* | 0.103 | 0.541 | 0.208 | *0.053* | *0.053* | *0.034* | *0.032* | *0.046* | *0.043* |
| | 2nd | *0.031* | *0.034* | *0.094* | *0.070* | 0.225 | 0.222 | 0.120 | *0.034* | *0.047* | *0.043* |
| | 3rd | *0.040* | *0.039* | *0.052* | *0.055* | *0.039* | *0.035* | *0.050* | 0.570 | 0.467 | 0.326 |
| **Censoring rate** | | 0.329 | 0.220 | 0.236 | 0.228 | 0.259 | 0.295 | 0.325 | 0.298 | 0.316 | 0.332 |

Table A.7: Results of the simulation study for the time-varying covariate effects with heavy censoring. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (exp): the size of the test under the exponential model; TVC1–TVC3 (linear/step/log): the power or the size (in italics) of the test under the exponential model with the first–third covariate allowed to depend on time through linear/step/log interaction.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT (a) | $1^{st}$ | *0.038* | 0.049 | 0.232 | 0.147 | *0.025* | *0.034* | *0.023* | *0.022* | *0.022* | *0.028* |
| | $2^{nd}$ | *0.042* | *0.022* | *0.052* | *0.043* | 0.215 | 0.149 | 0.175 | *0.033* | *0.024* | *0.029* |
| | $3^{rd}$ | *0.033* | *0.035* | *0.043* | *0.041* | *0.027* | *0.039* | *0.039* | 0.169 | 0.118 | 0.128 |
| GT (b) | $1^{st}$ | *0.049* | 0.072 | 0.152 | 0.179 | *0.037* | *0.055* | *0.042* | *0.045* | *0.040* | *0.057* |
| | $2^{nd}$ | *0.052* | *0.041* | *0.054* | *0.052* | 0.224 | 0.134 | 0.294 | *0.046* | *0.051* | *0.059* |
| | $3^{rd}$ | *0.048* | *0.047* | *0.047* | *0.042* | *0.039* | *0.051* | *0.049* | 0.183 | 0.101 | 0.227 |
| GT (c) | $1^{st}$ | *0.047* | 0.060 | 0.210 | 0.145 | *0.046* | *0.049* | *0.040* | *0.041* | *0.038* | *0.051* |
| | $2^{nd}$ | *0.056* | *0.037* | *0.048* | *0.044* | 0.278 | 0.156 | 0.263 | *0.056* | *0.046* | *0.057* |
| | $3^{rd}$ | *0.052* | *0.044* | *0.040* | *0.045* | *0.043* | *0.050* | *0.051* | 0.217 | 0.105 | 0.207 |
| GT (d) | $1^{st}$ | *0.048* | 0.054 | 0.227 | 0.139 | *0.028* | *0.037* | *0.032* | *0.032* | *0.029* | *0.040* |
| | $2^{nd}$ | *0.050* | *0.032* | *0.044* | *0.043* | 0.260 | 0.178 | 0.221 | *0.048* | *0.030* | *0.046* |
| | $3^{rd}$ | *0.046* | *0.042* | *0.030* | *0.041* | *0.033* | *0.049* | *0.043* | 0.216 | 0.131 | 0.172 |
| GT (e) | $1^{st}$ | *0.054* | 0.049 | 0.089 | 0.083 | *0.046* | *0.052* | *0.051* | *0.048* | *0.050* | *0.061* |
| | $2^{nd}$ | *0.059* | *0.044* | *0.052* | *0.061* | 0.178 | 0.093 | 0.256 | *0.060* | *0.057* | *0.064* |
| | $3^{rd}$ | *0.055* | *0.042* | *0.052* | *0.048* | *0.050* | *0.052* | *0.053* | 0.166 | 0.081 | 0.208 |
| GT (f) | $1^{st}$ | *0.049* | 0.076 | 0.172 | 0.184 | *0.044* | *0.049* | *0.040* | *0.046* | *0.036* | *0.051* |
| | $2^{nd}$ | *0.046* | *0.040* | *0.043* | *0.053* | 0.230 | 0.141 | 0.290 | *0.044* | *0.048* | *0.061* |
| | $3^{rd}$ | *0.050* | *0.042* | *0.040* | *0.049* | *0.036* | *0.045* | *0.052* | 0.194 | 0.099 | 0.229 |

Table A.8: Continuation of Table A.7. The censoring rate gives the proportion of the observations where no event is recorded.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LWY (a) | 1st | *0.048* | 0.069 | 0.271 | 0.152 | *0.058* | *0.062* | *0.052* | *0.047* | *0.042* | *0.060* |
| | 2nd | *0.065* | *0.051* | *0.069* | *0.063* | 0.256 | 0.169 | 0.239 | *0.065* | *0.057* | *0.060* |
| | 3rd | *0.060* | *0.052* | *0.064* | *0.065* | *0.059* | *0.061* | *0.067* | 0.207 | 0.125 | 0.212 |
| LWY (b) | 1st | *0.123* | 0.044 | 0.253 | 0.168 | *0.057* | *0.132* | *0.185* | *0.089* | *0.128* | *0.171* |
| | 2nd | *0.147* | *0.044* | *0.092* | *0.118* | 0.227 | 0.273 | 0.298 | *0.096* | *0.148* | *0.167* |
| | 3rd | *0.207* | *0.213* | *0.276* | *0.305* | *0.157* | *0.246* | *0.320* | 0.389 | 0.372 | 0.357 |
| LZD (a) | 1st | *0.057* | 0.071 | 0.273 | 0.165 | *0.052* | *0.057* | *0.043* | *0.046* | *0.050* | *0.058* |
| | 2nd | *0.065* | *0.042* | *0.062* | *0.056* | 0.301 | 0.223 | 0.240 | *0.061* | *0.044* | *0.051* |
| | 3rd | *0.048* | *0.048* | *0.050* | *0.055* | *0.048* | *0.059* | *0.061* | 0.239 | 0.161 | 0.186 |
| LZD (b) | 1st | *0.048* | 0.057 | 0.247 | 0.143 | *0.040* | *0.041* | *0.032* | *0.037* | *0.033* | *0.046* |
| | 2nd | *0.046* | *0.031* | *0.056* | *0.046* | 0.255 | 0.189 | 0.207 | *0.052* | *0.034* | *0.033* |
| | 3rd | *0.035* | *0.035* | *0.041* | *0.046* | *0.039* | *0.048* | *0.049* | 0.194 | 0.137 | 0.162 |
| **Censoring rate** | | 0.623 | 0.519 | 0.558 | 0.569 | 0.567 | 0.607 | 0.643 | 0.598 | 0.621 | 0.638 |

Table A.9: Results of the simulation study for the time-varying covariate effects with covariate-dependent censoring. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (exp): the size of the test under the exponential model; TVC1–TVC3 (linear/step/log): the power or the size (in italics) of the test under the exponential model with the first–third covariate allowed to depend on time through linear/step/log interaction.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GT (a) | $1^{\text{st}}$ | *0.025* | 0.101 | 0.495 | 0.250 | *0.016* | *0.029* | *0.030* | *0.024* | *0.029* | *0.033* |
| | $2^{\text{nd}}$ | *0.034* | *0.036* | *0.043* | *0.037* | 0.447 | 0.471 | 0.255 | *0.027* | *0.016* | *0.035* |
| | $3^{\text{rd}}$ | *0.026* | *0.047* | *0.041* | *0.035* | *0.027* | *0.024* | *0.028* | 0.358 | 0.318 | 0.241 |
| GT (b) | $1^{\text{st}}$ | *0.050* | 0.101 | 0.312 | 0.289 | *0.039* | *0.045* | *0.050* | *0.044* | *0.056* | *0.060* |
| | $2^{\text{nd}}$ | *0.048* | *0.056* | *0.046* | *0.043* | 0.444 | 0.391 | 0.441 | *0.039* | *0.056* | *0.054* |
| | $3^{\text{rd}}$ | *0.053* | *0.047* | *0.050* | *0.039* | *0.046* | *0.052* | *0.056* | 0.442 | 0.302 | 0.406 |
| GT (c) | $1^{\text{st}}$ | *0.048* | 0.110 | 0.518 | 0.257 | *0.035* | *0.044* | *0.054* | *0.042* | *0.046* | *0.053* |
| | $2^{\text{nd}}$ | *0.053* | *0.046* | *0.036* | *0.049* | 0.560 | 0.588 | 0.402 | *0.036* | *0.042* | *0.061* |
| | $3^{\text{rd}}$ | *0.052* | *0.050* | *0.040* | *0.043* | *0.045* | *0.051* | *0.046* | 0.535 | 0.436 | 0.379 |
| GT (d) | $1^{\text{st}}$ | *0.045* | 0.109 | 0.526 | 0.250 | *0.030* | *0.041* | *0.047* | *0.039* | *0.042* | *0.051* |
| | $2^{\text{nd}}$ | *0.052* | *0.043* | *0.035* | *0.047* | 0.563 | 0.595 | 0.378 | *0.035* | *0.043* | *0.056* |
| | $3^{\text{rd}}$ | *0.054* | *0.050* | *0.041* | *0.043* | *0.044* | *0.045* | *0.045* | 0.542 | 0.452 | 0.357 |
| GT (e) | $1^{\text{st}}$ | *0.049* | 0.055 | 0.059 | 0.083 | *0.052* | *0.054* | *0.050* | *0.049* | *0.050* | *0.047* |
| | $2^{\text{nd}}$ | *0.047* | *0.047* | *0.047* | *0.045* | 0.099 | 0.092 | 0.219 | *0.041* | *0.061* | *0.048* |
| | $3^{\text{rd}}$ | *0.057* | *0.042* | *0.034* | *0.056* | *0.045* | *0.052* | *0.056* | 0.118 | 0.087 | 0.206 |
| GT (f) | $1^{\text{st}}$ | *0.044* | 0.110 | 0.428 | 0.287 | *0.036* | *0.046* | *0.056* | *0.044* | *0.057* | *0.065* |
| | $2^{\text{nd}}$ | *0.049* | *0.055* | *0.041* | *0.048* | 0.477 | 0.434 | 0.445 | *0.041* | *0.058* | *0.055* |
| | $3^{\text{rd}}$ | *0.057* | *0.047* | *0.041* | *0.039* | *0.048* | *0.052* | *0.055* | 0.457 | 0.323 | 0.407 |

Table A.10: Continuation of Table A.9. The censoring rate gives the proportion of the observations where no event is recorded.

| **Test** | **Cov** | **PH** (exp) | **TVC1** (linear) | **TVC1** (step) | **TVC1** (log) | **TVC2** (linear) | **TVC2** (step) | **TVC2** (log) | **TVC3** (linear) | **TVC3** (step) | **TVC3** (log) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LWY (a) | 1st | *0.052* | 0.121 | 0.592 | 0.239 | *0.054* | *0.057* | *0.057* | *0.058* | *0.057* | *0.053* |
| | 2nd | *0.059* | *0.055* | *0.050* | *0.059* | 0.523 | 0.646 | 0.350 | *0.045* | *0.050* | *0.070* |
| | 3rd | *0.054* | *0.054* | *0.057* | *0.055* | *0.054* | *0.057* | *0.055* | 0.512 | 0.506 | 0.344 |
| LWY (b) | 1st | *0.031* | 0.088 | 0.468 | 0.196 | *0.037* | *0.044* | *0.046* | *0.049* | *0.033* | *0.042* |
| | 2nd | *0.040* | *0.044* | *0.035* | *0.050* | 0.509 | 0.575 | 0.289 | *0.039* | *0.051* | *0.053* |
| | 3rd | *0.227* | *0.275* | *0.319* | *0.327* | *0.273* | *0.266* | *0.233* | 0.800 | 0.626 | 0.480 |
| LZD (a) | 1st | *0.044* | 0.117 | 0.542 | 0.271 | *0.034* | *0.048* | *0.062* | *0.052* | *0.049* | *0.049* |
| | 2nd | *0.055* | *0.054* | *0.051* | *0.050* | 0.612 | 0.661 | 0.370 | *0.047* | *0.049* | *0.061* |
| | 3rd | *0.055* | *0.057* | *0.054* | *0.047* | *0.049* | *0.049* | *0.050* | 0.612 | 0.527 | 0.352 |
| LZD (b) | 1st | *0.033* | 0.102 | 0.537 | 0.247 | *0.028* | *0.040* | *0.048* | *0.040* | *0.041* | *0.040* |
| | 2nd | *0.043* | *0.042* | *0.049* | *0.042* | 0.524 | 0.613 | 0.323 | *0.034* | *0.038* | *0.045* |
| | 3rd | *0.043* | *0.053* | *0.053* | *0.046* | *0.040* | *0.040* | *0.037* | 0.510 | 0.475 | 0.313 |
| **Censoring rate** | | 0.329 | 0.218 | 0.235 | 0.226 | 0.263 | 0.294 | 0.323 | 0.299 | 0.317 | 0.332 |

Table A.11: Results of the simulation study for the model misspecification under the standard setting. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (Gomp): the size of the test under the Gompertz model; AFT (log-norm): the power of the test under the log-normal accelerated failure time model; AH (linear): the power of the test under the additive hazards model with the linear baseline hazard.

| **Test** | **Cov** | **PH** (Gomp) | **AFT** (log-norm) | **AH** (linear) |
|---|---|---|---|---|
| GT (a) | 1st | *0.041* | 0.061 | 0.044 |
| | 2nd | *0.027* | 0.223 | 0.042 |
| | 3rd | *0.029* | 0.453 | 0.073 |
| GT (b) | 1st | *0.050* | 0.062 | 0.052 |
| | 2nd | *0.048* | 0.242 | 0.064 |
| | 3rd | *0.051* | 0.490 | 0.100 |
| GT (c) | 1st | *0.060* | 0.067 | 0.056 |
| | 2nd | *0.046* | 0.216 | 0.058 |
| | 3rd | *0.047* | 0.504 | 0.087 |
| GT (d) | 1st | *0.056* | 0.065 | 0.053 |
| | 2nd | *0.039* | 0.216 | 0.054 |
| | 3rd | *0.047* | 0.499 | 0.084 |
| GT (e) | 1st | *0.043* | 0.059 | 0.049 |
| | 2nd | *0.045* | 0.098 | 0.061 |
| | 3rd | *0.043* | 0.181 | 0.064 |
| GT (f) | 1st | *0.050* | 0.073 | 0.053 |
| | 2nd | *0.051* | 0.274 | 0.063 |
| | 3rd | *0.053* | 0.565 | 0.101 |
| LWY (a) | 1st | *0.065* | 0.066 | 0.066 |
| | 2nd | *0.055* | 0.151 | 0.074 |
| | 3rd | *0.043* | 0.456 | 0.092 |
| LWY (b) | 1st | *0.035* | 0.054 | 0.043 |
| | 2nd | *0.029* | 0.169 | 0.050 |
| | 3rd | *0.274* | 0.768 | 0.357 |
| LZD (a) | 1st | *0.066* | 0.075 | 0.064 |
| | 2nd | *0.051* | 0.177 | 0.064 |
| | 3rd | *0.049* | 0.516 | 0.092 |
| LZD (b) | 1st | *0.051* | 0.060 | 0.052 |
| | 2nd | *0.032* | 0.165 | 0.054 |
| | 3rd | *0.036* | 0.441 | 0.083 |
| **Censoring rate** | | 0.290 | 0.285 | 0.234 |

Table A.12: Results of the simulation study for the model misspecification with low sample size. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (Gomp): the size of the test under the Gompertz model; AFT (log-norm): the power of the test under the log-normal accelerated failure time model; AH (linear): the power of the test under the additive hazards model with the linear baseline hazard.

| **Test** | **Cov** | **PH** (Gomp) | **AFT** (log-norm) | **AH** (linear) |
|---|---|---|---|---|
| GT (a) | 1st | *0.028* | 0.052 | 0.025 |
| | 2nd | *0.016* | 0.084 | 0.035 |
| | 3rd | *0.027* | 0.149 | 0.045 |
| GT (b) | 1st | *0.062* | 0.057 | 0.056 |
| | 2nd | *0.055* | 0.094 | 0.053 |
| | 3rd | *0.058* | 0.160 | 0.065 |
| GT (c) | 1st | *0.050* | 0.060 | 0.046 |
| | 2nd | *0.044* | 0.093 | 0.051 |
| | 3rd | *0.048* | 0.165 | 0.068 |
| GT (d) | 1st | *0.046* | 0.057 | 0.042 |
| | 2nd | *0.038* | 0.090 | 0.047 |
| | 3rd | *0.047* | 0.162 | 0.067 |
| GT (e) | 1st | *0.063* | 0.047 | 0.058 |
| | 2nd | *0.058* | 0.061 | 0.047 |
| | 3rd | *0.059* | 0.091 | 0.051 |
| GT (f) | 1st | *0.061* | 0.062 | 0.052 |
| | 2nd | *0.048* | 0.113 | 0.044 |
| | 3rd | *0.062* | 0.175 | 0.068 |
| LWY (a) | 1st | *0.049* | 0.064 | 0.060 |
| | 2nd | *0.070* | 0.091 | 0.053 |
| | 3rd | *0.060* | 0.190 | 0.082 |
| LWY (b) | 1st | *0.367* | 0.437 | 0.432 |
| | 2nd | *0.413* | 0.507 | 0.459 |
| | 3rd | *0.506* | 0.582 | 0.561 |
| LZD (a) | 1st | *0.061* | 0.068 | 0.063 |
| | 2nd | *0.058* | 0.096 | 0.052 |
| | 3rd | *0.063* | 0.182 | 0.079 |
| LZD (b) | 1st | *0.041* | 0.050 | 0.041 |
| | 2nd | *0.037* | 0.067 | 0.036 |
| | 3rd | *0.037* | 0.142 | 0.053 |
| **Censoring rate** | | 0.290 | 0.287 | 0.232 |

Table A.13: Results of the simulation study for the model misspecification with correlated covariates. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (Gomp): the size of the test under the Gompertz model; AFT (log-norm): the power of the test under the log-normal accelerated failure time model; AH (linear): the power of the test under the additive hazards model with the linear baseline hazard.

| **Test** | **Cov** | **PH** (Gomp) | **AFT** (log-norm) | **AH** (linear) |
|---|---|---|---|---|
| GT (a) | $1^{st}$ | *0.036* | 0.063 | 0.053 |
| | $2^{nd}$ | *0.025* | 0.093 | 0.038 |
| | $3^{rd}$ | *0.036* | 0.431 | 0.071 |
| GT (b) | $1^{st}$ | *0.056* | 0.067 | 0.058 |
| | $2^{nd}$ | *0.037* | 0.097 | 0.045 |
| | $3^{rd}$ | *0.048* | 0.466 | 0.090 |
| GT (c) | $1^{st}$ | *0.055* | 0.077 | 0.061 |
| | $2^{nd}$ | *0.035* | 0.093 | 0.059 |
| | $3^{rd}$ | *0.056* | 0.482 | 0.086 |
| GT (d) | $1^{st}$ | *0.054* | 0.077 | 0.061 |
| | $2^{nd}$ | *0.034* | 0.094 | 0.058 |
| | $3^{rd}$ | *0.058* | 0.478 | 0.088 |
| GT (e) | $1^{st}$ | *0.048* | 0.054 | 0.042 |
| | $2^{nd}$ | *0.039* | 0.055 | 0.038 |
| | $3^{rd}$ | *0.052* | 0.165 | 0.075 |
| GT (f) | $1^{st}$ | *0.050* | 0.071 | 0.057 |
| | $2^{nd}$ | *0.038* | 0.111 | 0.046 |
| | $3^{rd}$ | *0.050* | 0.563 | 0.089 |
| LWY (a) | $1^{st}$ | *0.057* | 0.113 | 0.063 |
| | $2^{nd}$ | *0.052* | 0.103 | 0.072 |
| | $3^{rd}$ | *0.057* | 0.493 | 0.098 |
| LWY (b) | $1^{st}$ | *0.038* | 0.091 | 0.039 |
| | $2^{nd}$ | *0.026* | 0.083 | 0.056 |
| | $3^{rd}$ | *0.261* | 0.783 | 0.366 |
| LZD (a) | $1^{st}$ | *0.056* | 0.127 | 0.066 |
| | $2^{nd}$ | *0.044* | 0.123 | 0.063 |
| | $3^{rd}$ | *0.057* | 0.588 | 0.098 |
| LZD (b) | $1^{st}$ | *0.046* | 0.112 | 0.053 |
| | $2^{nd}$ | *0.036* | 0.104 | 0.051 |
| | $3^{rd}$ | *0.044* | 0.489 | 0.083 |
| **Censoring rate** | | 0.289 | 0.285 | 0.234 |

Table A.14: Results of the simulation study for the model misspecification with heavy censoring. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (Gomp): the size of the test under the Gompertz model; AFT (log-norm): the power of the test under the log-normal accelerated failure time model; AH (linear): the power of the test under the additive hazards model with the linear baseline hazard.

| **Test** | **Cov** | **PH** (Gomp) | **AFT** (log-norm) | **AH** (linear) |
|---|---|---|---|---|
| GT (a) | 1st | *0.033* | 0.046 | 0.034 |
| | 2nd | *0.034* | 0.102 | 0.037 |
| | 3rd | *0.031* | 0.194 | 0.056 |
| GT (b) | 1st | *0.048* | 0.055 | 0.047 |
| | 2nd | *0.059* | 0.118 | 0.064 |
| | 3rd | *0.051* | 0.212 | 0.081 |
| GT (c) | 1st | *0.047* | 0.060 | 0.051 |
| | 2nd | *0.050* | 0.114 | 0.050 |
| | 3rd | *0.045* | 0.227 | 0.077 |
| GT (d) | 1st | *0.037* | 0.057 | 0.040 |
| | 2nd | *0.044* | 0.106 | 0.046 |
| | 3rd | *0.040* | 0.217 | 0.067 |
| GT (e) | 1st | *0.055* | 0.061 | 0.053 |
| | 2nd | *0.057* | 0.083 | 0.054 |
| | 3rd | *0.060* | 0.117 | 0.065 |
| GT (f) | 1st | *0.054* | 0.071 | 0.050 |
| | 2nd | *0.054* | 0.139 | 0.058 |
| | 3rd | *0.051* | 0.224 | 0.081 |
| LWY (a) | 1st | *0.052* | 0.066 | 0.060 |
| | 2nd | *0.069* | 0.113 | 0.061 |
| | 3rd | *0.050* | 0.238 | 0.089 |
| LWY (b) | 1st | *0.101* | 0.438 | 0.043 |
| | 2nd | *0.126* | 0.508 | 0.043 |
| | 3rd | *0.224* | 0.640 | 0.221 |
| LZD (a) | 1st | *0.052* | 0.073 | 0.052 |
| | 2nd | *0.064* | 0.101 | 0.056 |
| | 3rd | *0.051* | 0.228 | 0.081 |
| LZD (b) | 1st | *0.040* | 0.049 | 0.042 |
| | 2nd | *0.041* | 0.079 | 0.041 |
| | 3rd | *0.042* | 0.181 | 0.066 |
| **Censoring rate** | | 0.593 | 0.673 | 0.541 |

Table A.15: Results of the simulation study for the model misspecification with covariate-dependent censoring. Test: the type of the test as labelled in Subsection 4.2.4; Cov: the covariate for which the power or the size is calculated; PH (Gomp): the size of the test under the Gompertz model; AFT (log-norm): the power of the test under the log-normal accelerated failure time model; AH (linear): the power of the test under the additive hazards model with the linear baseline hazard.

| **Test** | **Cov** | **PH** (Gomp) | **AFT** (log-norm) | **AH** (linear) |
|---|---|---|---|---|
| GT (a) | 1st | *0.040* | 0.058 | 0.045 |
| | 2nd | *0.030* | 0.218 | 0.058 |
| | 3rd | *0.027* | 0.462 | 0.070 |
| GT (b) | 1st | *0.055* | 0.066 | 0.055 |
| | 2nd | *0.053* | 0.228 | 0.066 |
| | 3rd | *0.053* | 0.497 | 0.094 |
| GT (c) | 1st | *0.052* | 0.066 | 0.052 |
| | 2nd | *0.047* | 0.218 | 0.067 |
| | 3rd | *0.049* | 0.505 | 0.088 |
| GT (d) | 1st | *0.050* | 0.066 | 0.054 |
| | 2nd | *0.042* | 0.214 | 0.065 |
| | 3rd | *0.045* | 0.498 | 0.086 |
| GT (e) | 1st | *0.045* | 0.048 | 0.044 |
| | 2nd | *0.047* | 0.095 | 0.049 |
| | 3rd | *0.048* | 0.173 | 0.064 |
| GT (f) | 1st | *0.055* | 0.077 | 0.056 |
| | 2nd | *0.052* | 0.273 | 0.071 |
| | 3rd | *0.053* | 0.570 | 0.103 |
| LWY (a) | 1st | *0.056* | 0.063 | 0.055 |
| | 2nd | *0.055* | 0.153 | 0.075 |
| | 3rd | *0.049* | 0.447 | 0.090 |
| LWY (b) | 1st | *0.037* | 0.056 | 0.046 |
| | 2nd | *0.029* | 0.167 | 0.065 |
| | 3rd | *0.290* | 0.763 | 0.365 |
| LZD (a) | 1st | *0.066* | 0.064 | 0.053 |
| | 2nd | *0.048* | 0.181 | 0.080 |
| | 3rd | *0.048* | 0.510 | 0.089 |
| LZD (b) | 1st | *0.046* | 0.060 | 0.044 |
| | 2nd | *0.039* | 0.162 | 0.067 |
| | 3rd | *0.039* | 0.447 | 0.079 |
| **Censoring rate** | | 0.289 | 0.284 | 0.231 |

# Bibliography

J. Anděl. *Základy matematické statistiky.* MatfyzPress, 2007.

P. C. Austin. Generating survival times to simulate Cox proportional hazards models with time-varying covariates. *Statistics in Medicine*, 31(29):3946–3958, 2012.

R. Bender, T. Augustin, and M. Blettner. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, 24(11):1713–1723, 2005.

N. E. Breslow, L. Edler, and J. Berger. A two-sample censored-data rank test for acceleration. *Biometrics*, pages 1049–1062, 1984.

A. Buchholz. *Assessment of time-varying long-term effects of therapies and prognostic factors.* PhD thesis, Technische Universität Dortmund, 2010.

Z. Cai and Y. Sun. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics*, 30(1):93–111, 2003.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, 1975.

C. de Boor. *A practical guide to splines*, volume 27. Springer-Verlag New York, 1978.

P. H. C. Eilers and B. D. Marx. Flexible smoothing with B-splines and penalties. *Statistical Science*, pages 89–102, 1996.

T. R. Fleming and D. P. Harrington. *Counting processes and survival analysis.* John Wiley & Sons, 1991.

L. P. Fried, N. O. Borhani, P. Enright, et al. The cardiovascular health study: Design and rationale. *Annals of Epidemiology*, 1(3):263–276, 1991.

R. Gill and M. Schumacher. A simple test of the proportional hazards assumption. *Biometrika*, 74(2):289–300, 1987.

P. M. Grambsch and T. M. Therneau. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3):515–526, 1994.

S. Grant, Y. Q. Chen, and S. May. Performance of goodness-of-fit tests for the Cox proportional hazards model with time-varying covariates. *Lifetime Data Analysis*, pages 1–14, 2013.

R. J. Gray. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association*, 87(420):942–951, 1992.

R. J. Gray. Spline-based tests in survival analysis. *Biometrics*, pages 640–652, 1994.

J. K. Grønnesby and Ø. Borgan. A method for checking regression models in survival analysis based on the risk score. *Lifetime Data Analysis*, 2(4):315–328, 1996.

F. E. Harrell and K. L. Lee. Verifying assumptions of the Cox proportional hazards model. In *Proceedings of the 11th Annual SAS User's Group International Conference. Cary, NC, SAS Institute*, 1986.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data.* John Wiley & Sons, 2002.

M. R. Khondoker and M. A. Islam. Use of log-log survival function in modeling time-covariate interactions in Cox regression. *Journal of Statistical Planning and Inference*, 139(6):1968–1973, 2009.

D. Kraus. Data-driven smooth tests of the proportional hazards assumption. *Lifetime Data Analysis*, 13(1):1–16, 2007.

J. T. Kvaløy and L. R. Neef. Tests for the proportional intensity assumption based on the score process. *Lifetime Data Analysis*, 10(2):139–157, 2004.

T. Ledwina. Data-driven version of Neyman's smooth test of fit. *Journal of the American Statistical Association*, 89(427):1000–1005, 1994.

D. Y. Lin. Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association*, 86 (415):725–728, 1991.

D. Y. Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71, 1994.

D. Y. Lin, L. J. Wei, and Z. Ying. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572, 1993.

J. Lin, D. Zhang, and M. Davidian. Smoothing spline-based score tests for proportional hazards models. *Biometrics*, 62(3):803–812, 2006.

T. Martinussen and T. H. Scheike. *Dynamic regression models for survival data.* Springer, 2006.

S. May and D. W. Hosmer. A simplified method of calculating an overall goodness-of-fit test for the Cox proportional hazards model. *Lifetime Data Analysis*, 4(2):109–120, 1998.

T. Moreau, J. O'Quigley, and M. Mesbah. A global goodness-of-fit statistic for the proportional hazards model. *Applied Statistics*, pages 212–218, 1985.

N. J. D. Nagelkerke, J. Oosting, and A. A. M. Hart. A simple test for goodness of fit of Cox's proportional hazards model. *Biometrics*, pages 483–486, 1984.

N. H. Ng'andu. *A comparison of test statistics for assessing the proportional hazards assumption of Cox's model.* PhD thesis, University of North Carolina, 1994.

J. O'Quigley and F. Pessione. Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, pages 135–144, 1989.

F. O'Sullivan. Nonparametric estimation of relative risk using splines and cross-validation. *SIAM Journal on Scientific and Statistical Computing*, 9(3):531–542, 1988.

A. Perperoglou, S. le Cessie, and H. C. van Houwelingen. Reduced-rank hazard regression for modelling non-proportional hazards. *Statistics in Medicine*, 25 (16):2831–2845, 2006.

F. E. Satterthwaite. An approximate distribution of estimates of variance components. *Biometrics Bulletin*, 2(6):110–114, 1946.

W. Sauerbrei, P. Royston, and M. Look. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. *Biometrical Journal*, 49(3):453–473, 2007.

T. H. Scheike and T. Martinussen. On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics*, 31(1): 51–62, 2004.

M. Schemper, S. Wakounig, and G. Heinze. The estimation of average hazard ratios by weighted Cox regression. *Statistics in Medicine*, 28(19):2473–2489, 2009.

D. Schoenfeld. Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, 67(1):145–153, 1980.

D. Schoenfeld. Partial residuals for the proportional hazards regression model. *Biometrika*, 69(1):239–241, 1982.

L. A. Sleeper and D. P. Harrington. Regression splines in the Cox model with application to covariate effects in liver disease. *Journal of the American Statistical Association*, 85(412):941–949, 1990.

T. M. Therneau and C. Crowson. Using time dependent covariates and time dependent coefficients in the Cox model. *Red*, 2:1, 2013.

T. M. Therneau and P. M. Grambsch. *Modeling survival data: Extending the Cox model.* Springer, 2000.

T. M. Therneau, P. M. Grambsch, and T. R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, 1990.

L. J. Wei. Testing goodness of fit for proportional hazards model with censored observations. *Journal of the American Statistical Association*, 79(387):649–652, 1984.

# List of Figures

# List of Tables

# List of Symbols

| | |
|---|---|
| $T_i$ | event time for individual $i$ |
| $C_i$ | censoring time for individual $i$ |
| $N_i(t)$ | observed event indicator for individual $i$ |
| $Y_i(t)$ | at-risk indicator for individual $i$ |
| $\overline{N}(t)$ | number of observed events |
| $\overline{Y}(t)$ | number of individuals at risk |
| $t_k$ | time of the $k$-th event |
| $\tau$ | time of the end of the study |
| $\lambda(t)$ | hazard function |
| $\lambda_0(t)$ | baseline hazard function |
| $\Lambda(t)$ | cumulative hazard function |
| $\Lambda_0(t)$ | cumulative baseline hazard function |
| $S(t)$ | survival function |
| $\mathcal{L}(\boldsymbol{\beta})$ | partial likelihood function |
| $\ell(\boldsymbol{\beta})$ | partial log-likelihood function |
| $\ell_{pen}(\boldsymbol{\beta})$ | penalized partial log-likelihood function |
| $\mathcal{U}(\boldsymbol{\beta})$ | score function |
| $\mathcal{I}(\boldsymbol{\beta})$ | observed information matrix |
| $\overline{\mathbf{X}}(\boldsymbol{\beta}, t)$ | conditional weighted mean of $\mathbf{X}_i(t)$ |
| $\mathcal{V}(\boldsymbol{\beta}, t)$ | conditional weighted variance matrix of $\mathbf{X}_i(t)$ |
| $\widehat{\boldsymbol{\beta}}$ | partial likelihood estimate of $\boldsymbol{\beta}$ |
| $\mathbf{a}^{\otimes 2}$ | outer product $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^{\mathsf{T}}$ of vector $\mathbf{a}$ |
| $n$ | number of individuals $(i = 1, \ldots, n)$ |
| $m$ | number of events $(k = 1, \ldots, m)$ |
| $p$ | number of covariates $(j = 1, \ldots, p)$ |
| $\sim$ | is distributed as |
| $\dot{\sim}$ | is approximately distributed as |
| $\doteq$ | is approximately equal to |
| $\mathbb{1}$ | indicator function |
| $\delta_{kk'}$ | Kronecker delta $\delta_{kk'} = \mathbb{1}(k = k')$ |
| $a \wedge b$ | minimum of $a$ and $b$ |
| $\mathsf{tr}$ | trace of a matrix |
| $\mathsf{ev}_\omega$ | $\omega$-th eigenvalue of a matrix |
| $\chi^2_p$ | chi-square distribution with $p$ degrees of freedom |
| $\mathsf{N}(\mu, \sigma^2)$ | normal distribution with mean $\mu$ and variance $\sigma^2$ |
| $\mathsf{N}(\boldsymbol{\mu}, \Sigma)$ | multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance matrix $\Sigma$ |
| $\mathsf{Exp}(\lambda)$ | exponential distribution with rate parameter $\lambda$ |
| $\mathsf{Unif}(a, b)$ | uniform distribution on $[a, b]$ |
| $\mathbb{R}$ | set of real numbers |
| R | statistical software |

# Attachments

The complete code in R containing the implementation of the tests of the proportional hazards assumption, the analysis of medical data and the simulation study is attached on the CD. The myocardial infarction data are not included as they are not approved for public release.