FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Anna Filipová

Regression analysis of current status data

Department of Probability and Mathematical Statistics

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

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

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

In .... date ....

Author's signature

I would like to thank doc. Mgr. Michal Kulich, Ph.D. for his kind and professional guidance and advice. I also thank my family and friends for their support during my studies.

Title: Regression analysis of current status data

Author: Anna Filipová

Department: Department of Probability and Mathematical Statistics

Supervisor: doc. Mgr. Michal Kulich, Ph.D., Department of Probability and Mathematical Statistics

Abstract: Survival analysis often includes dealing with data that are censored. This thesis focuses on censoring in the form of current status data. We discuss several methods of regression analysis of current status data and focus mainly on a method that assumes that the time to event follows the additive hazards model. Under the assumption of proportional hazards for the monitoring time, this method does not require knowing the baseline hazard function and allows us to use the theory and software which were developed for Cox model. We also present a modification of this method, a two-step estimator, and show that it is asymptotically normal and has the advantage of lower asymptotic variance.

Keywords: current status data, additive hazards model, Cox model, survival analysis

Název práce: Regresní analýza dat o současném stavu

Autor: Anna Filipová

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: doc. Mgr. Michal Kulich, Ph.D., Katedra pravděpodobnosti a matematické statistiky

Abstrakt: Součástí analýzy přežití je často manipulace s cenzorovanými daty. Tato práce se soustředí na cenzorování ve formě dat o současném stavu. Uvádíme několik metod regresní analýzy s tímto typem dat a soustředíme se převážně na metodu, která pro čas do události předpokládá model aditivních rizik. Pokud navíc pro monitorovací čas předpokládáme model proporcionálních rizik, nemusíme už specifikovat základní rizikovou funkci a můžeme použít teorii a software, které byly vytvořené pro Coxův model. Dále prezentujeme modifikaci této metody, konkrétně dvoufázový odhad, a ukazujeme, že tento odhad je taktéž asymptoticky normální a navíc má nižší asymptotický rozptyl.

Klíčová slova: data o současném stavu, model aditivního rizika, Coxův model, analýza přežití

Contents

Introduction

The aim of survival analysis is to analyze time-to-event data. Time to event is a random variable measuring the time until an event happens, this can be for example occurence of a disease in a medical study. The data are very often censored — we are prevented from observing the exact time to event in some/all of the observations. In the case of a medical study, the subject may leave the study before the event occurs — this is an example of right-censoring. In this thesis, we are focusing on censoring in the form of current status data. In this type of data, we only observe the subject once at so-called monitoring time and see whether the event has occured before the monitoring time or not. We will be interested in estimating the effect of covariates on time to event from such data.

One of the most popular models for performing regression on time-to-event data is the Cox model, which is used on right-censored data. Its advantage is that one is able to estimate the effect of (possibly time-varying) covariates without having to estimate the distribution of the time to event itself. Many authors have introduced methods of performing regression analysis on current status data and some of them (Finkelstein [1986], Diamond and McDonald [1991], Huang [1996]) are based on the Cox proportional hazards model. However, these methods require estimation of the distribution of time to event to be able to proceed to estimation of the effect of covariates and these covariates can only be constant. The main topic of this thesis is the method proposed by Lin et al. [1998], which performs regression on current status data using an additive hazards model for time to event and proportional hazards model for monitoring time. The application of this method is surprisingly simple and does not require any estimation of the distribution of time to event.

In the first chapter of the thesis, we introduce the basics of survival analysis and also sum up the most important results of martingale theory that are used in regression models. The second chapter describes the theory regarding the Cox proportional hazards model and the additive hazards model in the case of right-censored data. In the third chapter, we discuss regression analysis on current status data, focusing deeply on the method proposed by Lin et al. [1998]. We study their method thoroughly and supplement it with details and proofs that were not contained in the article. The fourth chapter presents a simulation study, in which we use the method by Lin et al. [1998] for different kinds of data.

1. Theoretical foundations for survival analysis

1.1 Time to event and censoring

Time-to-event data arise when we await an event that should occur with (some of) the subjects. We determine the beginning or, in other words, time zero (this can be for example start of the study, which is common for all the subjects, or the date of birth of individuals, which differs among subjects) and define $T \geq 0$ as the time from the beginning to the event. Examples of such events are heart attack of a person, death of a person caused by a certain disease, explosion of a lightbulb or breakdown of a machine. Based on the character of the event or the scientific field, T can be called time to event, survival time or failure time. We will now summarize basic definitions and results from analysis of such data (survival analysis), which will be used later in the thesis. To keep the summary brief, we will only restrict ourselves to continuous T.

Definition 1. *Let T be a continuous and non-negative random variable. Denote by f its density and by F its cumulative distribution function. We define the survival function as*

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

the hazard function as

$$\lambda(t) = \lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq T < t + h | T \geq t)$$

and the cumulative hazard function as

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

for $t \geq 0$.

The following lemma describes useful relationships between these functions.

Lemma 1. *Let T be a continuous and non-negative random variable and let f, F, S, λ and Λ be functions defined as in Definition 1. Then*

$$\lambda(t) = \frac{f(t)}{S(t_-)} = \frac{f(t)}{S(t)}$$

and

$$S(t) = e^{-\Lambda(t)}$$

for $t \geq 0$.

We will now extend the topic with the possibility of censoring. A subject is censored if we are prevented from observing the time of the failure. We will (in the first two chapters) focus on right-censoring. A subject is right-censored if we observe it from the beginning and it leaves the study before the time to event T.

Examples of this are the subject's own decision to leave the study, the subject's death (by a different cause than the one defining the failure, if the failure is death) or end of the study itself (e.g. if the study is designed to last 5 years). Let us denote by $C \geq 0$ the time of censoring. Instead of observing the time of failure T of every subject, we observe random variables $X = \min(T, C)$ and $\delta = \mathbb{I}(T \leq C)$. To be able to derive any useful properties of the distribution of failure time, we need to assume some form of independence between T and C.

Definition 2. *Let T and C be the failure time and censoring time defined above. We will say that the independent censoring condition holds if*

$$\lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq T < t + h | T \geq t) = \lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq T < t + h | T \geq t, C \geq t).$$

In other words, the independent censoring condition requires that the hazard of failure at time t does not depend on whether the subject has already been censored or not. The independent censoring condition holds automatically if the failure and censoring times are stochastically independent.

1.2 Counting processes and martingale theory in survival analysis

We will now start looking at failure time and its observing in terms of stochastic processes. Define the processes

$$\{N(t), t \geq 0\}, \quad \text{where } N(t) = \mathbb{I}(X \leq t, \delta = 1) = \mathbb{I}(T \leq t, \delta = 1),$$

and

$$\{Y(t), t \geq 0\}, \quad \text{where } Y(t) = \mathbb{I}(X \geq t).$$

The process $N(t)$ starts at value 0 and remains equal to 0 for the whole time $t \geq 0$ if the subject is censored, or jumps to 1 at a certain time t if and only if the failure occurs at t and we manage to observe it. The process $Y(t)$ starts at value 1 and jumps to 0 after the first of the events (failure or censoring) occurs. As long as $Y(t)$ is still equal to 1, we say that the subject is at risk. These two processes carry the same information as the pair of observations (X, δ). We assume to be working on a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ throughout the whole thesis. Let us define a filtration of this probability space,

$$\mathcal{F}_t = \sigma\{N(s), Y(s_+), 0 \leq s \leq t\} \quad \text{for all } \ t \geq 0.$$

For a fixed t, the σ-algebra $\mathcal{F}_t$ carries all information about the subject up to time t. Also, $N(t)$ is a counting process with respect to the filtration $\mathcal{F}_t$. Applying the Doob-Meyer decomposition gives us the following lemma.

Lemma 2. *The process $\int_0^t Y(s)\, d\Lambda(s)$ is a right-continuous $\mathcal{F}_t$-predictable process. The process*

$$M(t) = N(t) - \int_0^t Y(s)\, d\Lambda(s)$$

is an $\mathcal{F}_t$-martingale if and only if the independent censoring condition (Definition 2) holds.

Denote by $\langle M_1, M_2\rangle(t)$ the predictable covariation process of any martingales M_1, M_2 and by $\langle M_1, M_1\rangle(t) = \langle M_1\rangle(t)$ the predictable variation process of any martingale M_1. Let us remind that we have restricted ourselves to continuous T. Then

$$\langle M, M\rangle(t) = \int_0^t Y(s)\, d\Lambda(s),$$

where the martingale M is the one from Lemma 2. Also, if $M_1(0) = M_2(0) = 0$ a.s. (which in our case holds), then $\operatorname{var} M_1(t) = \mathsf{E}\,\langle M_1\rangle(t)$ and $\operatorname{cov}(M_1(t), M_2(t)) = \mathsf{E}\,\langle M_1, M_2\rangle(t)$.

Because we will be dealing a lot with martingale integrals, we are now introducing their properties in the following lemma.

Lemma 3. *Let $M_1(t), M_2(t)$ be $\mathcal{F}_t$-martingales as in Lemma 2. Let H_1, H_2 be bounded $\mathcal{F}_t$-predictable processes. Then*

$$L_j(t) = \int_0^t H_j(s)\, dM_j(s)$$

is an $\mathcal{F}_t$-martingale, $j = 1, 2$. Also, there exists a predictable covariation process $\langle L_1, L_2\rangle$ and it holds that

$$\langle L_1, L_2\rangle(t) = \int_0^t H_1 H_2\, d\langle M_1, M_2\rangle(s).$$

The last and crucial part of martingale theory that we will use later is the Central limit theorem for sums of martingale integrals. First, we are specifying the forms and properties of random processes used in this theorem.

Definition 3. *Let $N_i(t), i = 1, ..., n$ be counting processes adapted to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$. Then $\{N_1(t), ..., N_n(t)\}$ is called a multivariate counting process if $\mathsf{P}(\Delta N_i(t) = 1, \Delta N_j(t) = 1) = 0$ for all $i \neq j, i, j \in \{1, ..., n\}$ and all $t \geq 0$.*

Let $\{N_i^{(n)} \; : i = 1, ..., n\}$ be a multivariate counting process with respect to the stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathsf{P})$.
Let the compensators $A_i^{(n)}$ for $N_i^{(n)}$ (i.e. $M_i^{(n)} = N_i^{(n)} - A_i^{(n)}$ are $\mathcal{F}_t$-martingales) be continuous.
Let $H_{ki}^{(n)}$, $k = 1, ..., p$, $i = 1, ..., n$, be bounded $\mathcal{F}_t$-predictable processes on an interval $\langle 0, \tau\rangle$.
Denote

$$U_{ki}^{(n)}(t) = \int_0^t H_{ki}^{(n)}(s) dM_i^{(n)}(s) \qquad \text{and} \qquad U_k^{(n)}(t) = \sum_{i=1}^n U_{ki}^{(n)}(t).$$

Denote for any $\varepsilon > 0$

$$U_{ki,\varepsilon}^{(n)}(t) = \int_0^t H_{ki}^{(n)}(s)\, \mathbb{I}(|H_{ki}^{(n)}(s)| > \varepsilon) dM_i^{(n)}(s) \qquad \text{and} \qquad U_{k,\varepsilon}^{(n)}(t) = \sum_{i=1}^n U_{ki,\varepsilon}^{(n)}(t).$$

Then $U_{ki}^{(n)}(t), U_k^{(n)}(t), U_{ki,\varepsilon}^{(n)}(t)$ and $U_{k,\varepsilon}^{(n)}(t)$ are square integrable $\mathcal{F}_t$-martingales and

$$\langle U_k^{(n)}, U_l^{(n)}\rangle(t) = \sum_{i=1}^n \int_0^t H_{ki}^{(n)}(s) H_{li}^{(n)}(s) dA_i^{(n)}(s)$$

and

$$\langle U_{k,\varepsilon}^{(n)}, U_{l,\varepsilon}^{(n)}\rangle(t) = \sum_{i=1}^n \int_0^t H_{ki}^{(n)}(s) H_{li}^{(n)}(s)\, \mathbb{I}(|H_{ki}^{(n)}(s)| > \varepsilon)\, \mathbb{I}(|H_{li}^{(n)}(s)| > \varepsilon) dA_i^{(n)}(s).$$

Theorem 4. *Let for all $t \in \langle 0, \tau \rangle$ and for all $k, l = 1, ..., p$*

$$\langle U_k^{(n)}, U_l^{(n)} \rangle(t) \xrightarrow{P} c_{kl}(t) < \infty$$

as $n \longrightarrow \infty$, where $c_{kl}(t)$ are continuous functions. Let for all $\varepsilon > 0$ and all $k = 1, ..., p$

$$\langle U_{k,\varepsilon}^{(n)}, U_{k,\varepsilon}^{(n)} \rangle(t) \xrightarrow{P} 0$$

as $n \longrightarrow \infty$. Then

$$(U_1^{(n)}, U_2^{(n)}, ..., U_p^{(n)}) \Longrightarrow (W_1, ..., W_p) \quad \text{on } D^p\langle 0, \tau \rangle,$$

where $\Longrightarrow$ denotes weak convergence, $W_1, ..., W_p$ are (dependent) zero-mean Gaussian processes with independent increments, a.s. continuous sample paths and covariance function $cov(W_k(s), W_l(t)) = c_{kl}(s)$ for all $k, l = 1, ..., p$ and $0 \leq s \leq t \leq \tau$. $D\langle 0, \tau \rangle$ stands for the space of functions on $\langle 0, \tau \rangle$ which are right-continuous with finite left-hand limits.
As a consequence, the random vector $(U_1^{(n)}(t), U_2^{(n)}(t), ..., U_p^{(n)}(t))$ evaluated at a fixed $t \in \langle 0, \tau \rangle$ converges in distribution to a p-dimensional normal random vector with zero mean and covariance matrix with elements $c_{kl}(t), k, l = 1, ..., p$.

2. Regression on time-to-event data

Consider again a non-negative, continous random variable T (failure time, time to event) and a non-negative random variable C (censoring time). This time, we will suppose that both T and C can be affected by a p-variate vector of random processes $\{\mathbf{Z}(t), t \geq 0\}$ — a time-varying vector of covariates. We require the following condition (independent censoring condition in the presence of covariates) to hold:

$$\begin{aligned}\lambda(t|\mathbf{Z}) &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \leq T < t+h | T \geq t, \mathbf{Z}(t))}{h} \\ &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \leq T < t+h | T \geq t, C \geq t, \mathbf{Z}(t))}{h},\end{aligned} \tag{2.1}$$

where $\lambda(t|\mathbf{Z})$ is the conditional hazard function of T (the first equality is its definition). What we actually observe, are n individuals (subjects) with their values of covariates from a starting time $t = 0$ until either failure or censoring happens. Let us denote the independent and identically distributed observations as $(X_i, \delta_i, \mathbf{Z}_i)$, $i = 1, ..., n$, where $X_i = \min(T_i, C_i)$ is the time of failure or censoring (whichever happens first) and $\delta_i = \mathbb{I}(T_i \leq C_i)$ indicates whether we have observed the failure time or not. All the information given to us by X_i and δ_i can be also described by two processes $\{N_i(t), t \geq 0\}$ and $\{Y_i(t), t \geq 0\}$, as was mentioned in section 1.2.

Analyses of such data often aim to estimate the effect of covariates $\{\mathbf{Z}(t), t \geq 0\}$ on the failure time T. There are many regression models that are based on dividing the conditional hazard function $\lambda(t|\mathbf{Z})$ into a function of $\mathbf{Z}$ and a function which does not depend on $\mathbf{Z}$. In this chapter, we are going to describe two such models — the well-known Cox model (proportional hazards model) and the additive hazards model.

2.1 Cox proportional hazards model

We will now, based on Fleming and Harrington [1991], describe the basis of the Cox proportional hazards model for right-censored data and summarize the assumptions and asymptotic properties of its estimation via partial likelihood. The classical Cox proportional hazards model specifies the hazard function as

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) e^{\beta_0^\top \mathbf{Z}(t)},$$

where $\lambda_0(t)$ is some unknown hazard function that does not depend on $\mathbf{Z}$. While the relationship between T and $\mathbf{Z}$ is defined using a p-variate vector of parameters β_0, the function $\lambda_0(t)$, called the baseline hazard, is not further specified. In other words, we are dealing with a semiparametric model.

The estimation of β_0 starts with the partial likelihood function

$$L(\beta) = \prod_{i=1}^{n} \prod_{s>0} \left[\frac{e^{\beta^\top \mathbf{Z}_i(s)}}{\sum_{j=1}^{n} Y_j(s) e^{\beta^\top \mathbf{Z}_j(s)}} \right]^{\Delta N_i(s)}.$$

It is called *partial* likelihood because the model is not parametric and the likelihood is not a standard product of density computed at different values. Define

$$S_n^{(k)}(\beta, s) = \frac{1}{n}\sum_{i=1}^{n} Y_i(s) e^{\beta^\top \mathbf{Z}_i(s)} \mathbf{Z}_i^{\otimes k}(s) \quad \text{for} \quad k = 0, 1, 2,$$

where for a vector x, $x^{\otimes k}$ equals 1 for $k = 0$, x for $k = 1$ and $xx^\top$ for $k = 2$. We can notice that we get $S_n^{(1)}(\beta, s)$ (resp. $S_n^{(2)}(\beta, s)$) by differentiating $S_n^{(0)}(\beta, s)$ with respect to β once (resp. twice). After taking the logarithm of the partial likelihood, we obtain

$$l_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \left[\beta^\top \mathbf{Z}_i(s) - \log nS_n^{(0)}(\beta, s)\right] dN_i(s).$$

After differentiating this with respect to β we obtain the score statistic

$$U_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \left[\mathbf{Z}_i(s) - \frac{S_n^{(1)}(\beta, s)}{S_n^{(0)}(\beta, s)}\right] dN_i(s) \tag{2.2}$$

and (after multiplying by $-\frac{1}{n}$ and differentiating it with respect to β one more time) the observed information matrix

$$\mathcal{I}_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \int_0^\infty \left[\frac{S_n^{(2)}(\beta, s)}{S_n^{(0)}(\beta, s)} - \left(\frac{S_n^{(1)}(\beta, s)}{S_n^{(0)}(\beta, s)}\right)^{\otimes 2}\right] dN_i(s).$$

Let us also define statistics $U_n(\beta, t)$ and $\mathcal{I}_n(\beta, t)$ for $t \geq 0$ as modifications of $U_n(\beta)$ and $\mathcal{I}_n(\beta)$ where we take the integral from 0 to t (instead of 0 to ∞). Then $U_n(\beta) = U_n(\beta, \infty)$ and $\mathcal{I}_n(\beta) = \mathcal{I}_n(\beta, \infty)$. Before summarizing the theory, we will list the regularity conditions that are used for its derivation. When using a norm, we mean the maximum norm where for a p-variate $\mathbf{x}$, $\|\mathbf{x}\| = \max(|x_1|, ..., |x_p|)$.

(C.1) The data are observed on an interval $\langle 0, \tau\rangle$ (i.e. $X_i \in \langle 0, \tau\rangle$ a.s., $i = 1, ..., n$) where $\tau > 0$ is fixed, $\int_0^\tau \lambda_0(t)dt < \infty$ and there exists an $\varepsilon > 0$ such that $\mathsf{P}(Y_i(t) = 1) > \varepsilon$ for all $t \in \langle 0, \tau\rangle$ and all $i = 1, ..., n$.

(C.2) All components of the covariate process $\mathbf{Z}_i(t)$ are bounded by a constant on the interval $\langle 0, \tau\rangle$, i.e. there exists $M > 0$ such that $\|\mathbf{Z}_i(t)\| < M$ a.s. for all $t \in \langle 0, \tau\rangle$, $i = 1, ..., n$. This condition is not necessary, but it simplifies the proofs.

(C.3) There exists a neighbourhood $\mathcal{B}$ of β_0 and functions $s^{(0)}, s^{(1)}, s^{(2)}$ defined on $\mathcal{B} \times \langle 0, \tau\rangle$ such that

$$\sup_{t\in\langle 0,\tau\rangle, \beta\in\mathcal{B}} \|S_n^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \xrightarrow{P} 0$$

as $n \to \infty$, for $j = 0, 1, 2$.

(C.4) It holds for all $\beta \in \mathcal{B}$ and $t \in \langle 0, \tau\rangle$ that

$$\frac{\partial s^{(0)}(\beta, t)}{\partial \beta} = s^{(1)}(\beta, t) \;\text{ and }\; \frac{\partial s^{(1)}(\beta, t)}{\partial \beta} = s^{(2)}(\beta, t).$$

(C.5) The functions $s^{(j)}$ are bounded and $s^{(0)}$ is bounded away from 0 on $\mathcal{B} \times \langle 0, \tau\rangle$. For $j = 0, 1, 2$, the family of functions $\{s^{(j)}(\cdot, t), t \in \langle 0, \tau\rangle\}$ is equicontinuous at β_0.

(C.6) Define $\mathbf{e}(\beta, t) = \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)}$. The matrix

$$I(\beta_0, \tau) = \int_0^{\tau} \left[\frac{s^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} - \mathbf{e}(\beta, t)^{\otimes 2}\right] s^{(0)}(\beta_0, t)\lambda_0(t)dt$$

is positive definite.

Conditions (C.3) to (C.5) are automatically fulfilled if the observations are independent and identically distributed and all components of the covariate have bounded support. Notice that $U_n(\beta, \infty) = U_n(\beta, \tau)$ a.s. and $\mathcal{I}_n(\beta, \infty) = \mathcal{I}_n(\beta, \tau)$ a.s. due to (C.1). We will now summarize the most important theoretical results. Consider $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s_+), \mathbf{Z}_i(s), 0 \le s \le t, i = 1, ..., n\}$, a filtration of the probability space $(\Omega, \mathcal{F}, \mathsf{P})$, and assume that $\mathbf{Z}_i(t)$ is $\mathcal{F}_t$-predictable for all $t \ge 0, i = 1, ..., n$.

Theorem 5. *Consider process*

$$M_i(t) = N_i(t) - \int_0^t Y_i(s)e^{\beta_0^\top \boldsymbol{Z}_i(s)}\lambda_0(s)ds.$$

Assume that the independent censoring condition (2.1) *holds. Then* $\{M_i(t), t \ge 0\}$ *is an* $\mathcal{F}_t$*-martingale. Furthermore, at the true parameter* β_0 *and for all* $t \ge 0$

$$U_n(\beta_0, t) = \sum_{i=1}^{n} \int_0^t \left[\boldsymbol{Z}_i(s) - \frac{S_n^{(1)}(\beta, s)}{S_n^{(0)}(\beta, s)}\right] dM_i(s).$$

Thus, $U_n(\beta_0, t)$ *is an* $\mathcal{F}_t$*-martingale.*

The asymptotic results, mostly based on martingale theory (section 1.2), are summed up in the following theorems.

Theorem 6. *Let the conditions* (C.1) *to* (C.6) *hold. Then the process* $\frac{1}{\sqrt{n}}U_n(\beta_0, t)$ *converges weakly on* $D^p\langle 0, \tau\rangle$ *to a p-variate zero-mean Gaussian process with continuous sample paths, independent increments and variance function* $I(\beta_0, t)$*. Thus,*

$$\frac{1}{\sqrt{n}}U_n(\beta_0, \tau) \xrightarrow{D} N_p\Big(\boldsymbol{0}, I(\beta_0, \tau)\Big).$$

Let $\hat{\beta}$ *be any consistent estimator of* β_0*. Then*

$$\sup_{t \in \langle 0, \tau\rangle} \|\mathcal{I}_n(\hat{\beta}, t) - I(\beta_0, t)\| \xrightarrow{P} 0.$$

Now denote by $\hat{\beta}$ the partial likelihood estimator of β_0, i.e. $U_n(\hat{\beta}) = \mathbf{0}$. The matrix $\mathcal{I}_n(\beta, t)$ is positive semidefinite and if we assume that it is non-singular, then the unique $\hat{\beta}$ that maximizes the partial loglikelihood is the unique solution of $U_n(\hat{\beta}) = \mathbf{0}$. This system of equations (or one equation in case of $p = 1$) is in practice solved by the Newton-Raphson algorithm. The next theorem states consistency and asymptotic distribution of the partial likelihood estimator.

Theorem 7. *Let the conditions* (C.1) - (C.6) *hold. Then*

$$\hat{\beta} \xrightarrow{P} \beta_0$$

and

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N_p\left(\boldsymbol{0}, I^{-1}(\beta_0, \tau)\right).$$

2.2 Additive hazards regression

This section summarizes briefly, based on Lin and Ying [1994], the additive hazards model and the estimation of its parameters. As the name suggests, this model divides the conditional hazard function into two summands:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t) + \beta_0^\top \mathbf{Z}(t),$$

where $\lambda_0(t)$ is some unknown hazard function that does not depend on $\mathbf{Z}$. Thus, this is also a semiparametric model. Note that in contrast to the Cox model, here we need to constrain $\beta_0^\top \mathbf{Z}(t)$ to make sure that $\lambda(t|\mathbf{Z})$ is non-negative.
Similarly as in the Cox model, we get the score function

$$U_n(\beta) = \sum_{i=1}^{n} \int_0^\infty \{\mathbf{Z}_i(s) - \tilde{\mathbf{Z}}(s)\}\{dN_i(s) - Y_i(s)\beta^\top \mathbf{Z}_i(s)ds\},$$

where

$$\tilde{\mathbf{Z}}(s) = \frac{\sum_{j=1}^n Y_j(s)\mathbf{Z}_j(s)}{\sum_{j=1}^n Y_j(s)}.$$

The estimator of β_0 is such $\hat{\beta}$ that solves $U_n(\hat{\beta}) = 0$. This time, we do not have to use the Newton-Raphson algorithm and we can get an explicit form of the estimator:

$$\hat{\beta} = \left[\sum_{i=1}^{n} \int_0^\infty Y_i(s)\{\mathbf{Z}_i(s) - \tilde{\mathbf{Z}}(s)\}^{\otimes 2} dt\right]^{-1} \left[\sum_{i=1}^{n} \int_0^\infty \{\mathbf{Z}_i(s) - \tilde{\mathbf{Z}}(s)\} dN_i(s)\right].$$

This estimator has then, under certain conditions, the following asymptotic distribution:

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{D} N_p\left(\mathbf{0}, A^{-1}BA^{-1}\right),$$

where $A^{-1}BA^{-1}$ can be consistently estimated by $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$, where

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n} \int_0^\infty \{\mathbf{Z}_i(s) - \tilde{\mathbf{Z}}(s)\}^{\otimes 2} dN_i(t), \quad \hat{A} = \frac{1}{n}\sum_{i=1}^{n} \int_0^\infty Y_i(t)\{\mathbf{Z}_i(s) - \tilde{\mathbf{Z}}(s)\}^{\otimes 2} dt.$$

3. Regression on current status data

Consider non-negative random variables T (failure time, time to event) and C (now called monitoring time). With current status data, we do not directly observe the failure time. We only know the monitoring time and whether the failure has already occurred before the monitoring time or not. Current status data can be considered as a special case of interval censoring — a type of censoring where we only know that the failure time lies within a specific interval. In case of current status data, the interval would be either $(-\infty, C)$ (the failure time is left-censored) or (C, ∞) (right-censored).

Such data can be seen in many areas of research and their censoring can have different causes. A typical cause is that we are literally unable to observe the failure time. One example, which showed up in the works of Finkelstein [1986], Huang [1996] and Lin et al. [1998], are animal tumorigenicity experiments. These experiments aim to analyze the effect of a possible carcinogen on T — the time of tumor outbreak. Because T cannot be observed, the animals in the experiment (e.g. mice) are examined after their sacrifice (or death by another cause) and the presence or absence of a tumor is recorded. A different cause of current status data occurs when the authors of a study intentionally decide for this type of heavier censoring in order to avoid inaccuracy and bias of the data, which are mostly caused by the respondents' poor memory. Examples of such situations (discussed in Diamond and McDonald [1991]) are studies on the time of menopause or first menstrual bleeding. In these cases, the authors may decide to only ask whether the milestone has already happened, rather than risk the possibility of bias, caused for example by the respondents' tendency to state later age.

To proceed to the topic of regression, consider a p-dimensional vector of random processes $\mathbf{Z} = \{\mathbf{Z}(t), t \geq 0\}$ (vector of possibly time-varying covariates). Our aim is to express and estimate the potential effect of $\mathbf{Z}(t)$ on T. We need to make restrictions on the nature of the covariates. It is important that the covariates are external. In Kalbfleisch and Prentice [1980], external covariates are defined as covariates that satisfy the condition

$$\mathsf{P}(t \leq T < t + h | T \geq t; \mathbf{Z}(s), s \leq t) = \mathsf{P}(t \leq T < t + h | T \geq t; \mathbf{Z}(s), s \leq u) \quad (3.1)$$

for all $0 < t \leq u$, which says that adding information about the future values of covariates to the information up to present time does not change the conditional hazard at this time. An example of a covariate that violates this condition is blood pressure of an individual measured throughout a study which considers death as the failure. Knowing the blood pressure at some time in the future would guarantee that the individual stays alive at least until that future. Kalbfleisch and Prentice [1980] describe external covariates in more details by distinguishing three types. The first type of external covariate is a constant covariate. This can be for example gender or some health condition at the beginning of the study. The second type are time-varying covariates that are determined in advance. Examples of these could be age or some medicine whose dose changes in a predetermined way. The third type are literally external time-varying covariates that are not related to the individuals, such as outside temperature or the level of air pollution. In this thesis, we will need survival function conditioned by the covariates for further derivations in the models. While it does not generally make sense to talk about survival function for time-varying covariates, external covariates are an exception. This is one of the reasons why we restrict ourselves to external covariates, one of the technical parts of the theory where we need to have only external covariates to be able to proceed to the next step. However, this restriction also goes hand in hand with the nature of our data. The fact that we cannot observe the subject the whole time and wait until the failure happens and that we can only examine at a single monitoring time whether the failure happened or not, often means that we are not able to observe other individual characteristics until the monitoring time either.

3.1 Proportional hazards model with current status data

We have introduced the proportional hazards model in chapter 2.1 and specified the relationship between the failure time T and the covariates $\mathbf{Z}(t)$ through the hazard function of T:

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{\beta_0^\top \mathbf{Z}(t)},$$

where $\lambda_0(t)$ (called the *baseline hazard*) is some unknown hazard function that does not depend on $\mathbf{Z}$. As was stated at the beginning of this chapter, this time we have current status data — the extent of censoring is such that it prevents us from estimating β_0 as in chapter 2.1 (where we had right-censored data). Because the only thing that we know is whether the failure occured before the monitoring time or not, the probability of this event will, in some way, have to form the main part of any likelihood that we build to estimate β_0. Using Lemma 1, we can express the survival function as

$$S(t|\mathbf{Z}) = e^{-\int_0^t \lambda_0(s)e^{\beta_0^\top \mathbf{Z}(s)}ds},$$

which appears troublesome because of the integral of product of the unknown baseline hazard and the relevant $e^{\beta_0^\top \mathbf{Z}(s)}$. If we only have constant covariates, we can write

$$S(t|\mathbf{Z}) = e^{-e^{\beta_0^\top \mathbf{Z}}\int_0^t \lambda_0(s)ds} = \left(e^{-\Lambda_0(t)}\right)^{e^{\beta_0^\top \mathbf{Z}}} = S(t|\mathbf{Z}=0)^{e^{\beta_0^\top \mathbf{Z}}},$$

which is more convenient for further derivations. This is what was used in all the works that we are mentioning in the next paragraphs.

Finkelstein [1986] worked with the proportional hazards model, proposed a method to deal with interval-censored data and described current status data as a special case of interval-censored data. The assumptions used to develop this method included covariates being constant (w.r.t. time) and the monitoring time being independent of both the failure time and the covariates. She mentioned the data being treated as incomplete observations from grouped response time data and, concerning the asymptotic properties, pointed out problems when having continuous or arbitrarily interval-censored data.

Diamond and McDonald [1991] worked with the likelihood of a binary regression model and proposed two methods to deal with the unknown baseline hazard. The first method divides the continuous data into groups (intervals) and uses conditional probability of ending in an interval, conditional on survival up to the interval. This method requires the number of time-intervals to be small. The second method uses splines to specify the baseline hazard. The whole work considers only constant covariates and the dependence/independence of monitoring time (on covariates or time to event) or even the random character of monitoring time are not discussed.

Huang [1996] approached the problem through profile likelihood — the likelihood is first maximized with respect to the cumulative baseline hazard function and then used to estimate the desired parameter. Again, the covariates are assumed to be constant. As for the monitoring time, it is only assumed to be independent of time to event conditionally on the covariates. It is also required that the joint distribution of monitoring time and covariates does not involve β_0 and $\Lambda_0(t)$. He proved that his estimator is asymptotically normal with $\sqrt{n}$ convergence rate.

None of these works managed to avoid restriction to constant covariates and some estimation of the unknown distribution function or (cumulative) baseline hazard function. This is where lie the main advantages of the approach of Lin et al. [1998], introduced in the next section.

3.2 Additive hazards model with current status data

Let us have $\{T_i, C_i, \mathbf{Z}_i\}$, $i = 1, ..., n$, independent replicates of $\{T, C, \mathbf{Z}\}$. Assume that both T and C are continuous. We will analyze the current status data under the additive hazards regression model

$$\begin{aligned}\lambda_T(t|\mathbf{Z}_i) &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \le T_i < t+h | T_i \ge t; \mathbf{Z}_i(s), s \le t)}{h} \\ &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \le T_i < t+h | T_i \ge t, \mathbf{Z}_i(t))}{h} = \lambda_{T,0}(t) + \beta_0^\top \mathbf{Z}_i(t),\end{aligned} \tag{3.2}$$

$\lambda_{T,0}(t)$ is an unspecified baseline hazard function, $\beta_0 \in \mathbb{R}^p$. We have reached the main topic of this thesis, which is the method for analysing current status data described in the article of Lin et al. [1998]. In the vast majority of this chapter, we go through the theory of this method and supplement it with proofs (or parts of them) which were omitted or only briefly proposed in the article.

Apart from the additive hazards model for the failure time, the method of Lin et al. [1998] is built upon the relationship between monitoring time and the covariates, formulated through the proportional hazards model

$$\begin{aligned}\lambda_C(t|\mathbf{Z}_i) &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \le C_i < t+h | C_i \ge t; \mathbf{Z}_i(s), s \le t)}{h} \\ &= \lim_{h \searrow 0} \frac{\mathsf{P}(t \le C_i < t+h | C_i \ge t, \mathbf{Z}_i(t))}{h} = \lambda_{C,0}(t) e^{\gamma_0^\top \mathbf{Z}_i(t)},\end{aligned} \tag{3.3}$$

$\lambda_{C,0}(t)$ is again an unspecified baseline hazard function, $\gamma_0 \in \mathbb{R}^p$. We express the hazard functions at time t in both models with covariates only from time t,

but that does not mean that the hazard function cannot depend on the past of the covariates — we are free to define some components of $\mathbf{Z}_i(t)$ as functions of $\mathbf{Z}_i(s), s \leq t$.
An important assumption is that T_i and C_i are conditionally independent given $\mathbf{Z}_i$, i.e.

$$\mathsf{P}(T_i \leq a, C_i \leq b|\mathbf{Z}_i) = \mathsf{P}(T_i \leq a|\mathbf{Z}_i) \cdot \mathsf{P}(C_i \leq b|\mathbf{Z}_i) \quad \forall a, b \in \mathbb{R}.$$

Notice that in the special case of $\gamma_0 = \mathbf{0}$, the monitoring time C_i does not depend on $\mathbf{Z}_i$ and is therefore fully independent of the failure time T_i.

As mentioned at the beginning of this chapter, we only observe $\{C_i, \delta_i, \mathbf{Z}_i\}$, $i = 1, ..., n$, where this time $\delta_i = \mathbb{I}(C_i \leq T_i)$. Now let us define counting process $N_i(t) = \delta_i \, \mathbb{I}(C_i \leq t)$ for $t \geq 0$. It starts at 0 and then jumps to 1 at t if and only if the monitoring occurs at t and the subject is found to be without failure. If the other outcome is observed (failure occurs before the monitoring), we will regard the process $N_i(t)$ as censored. One has to keep in mind the almost opposite meanings of $N_i(t)$ and being censored in the Cox model and our model. In the former, the process $N_i(t)$ had a jump only if we observed the failure. In the latter, the process $N_i(t)$ has a jump only if we find (at time C_i) a failure-free subject. We also define process $Y_i(t) = \mathbb{I}(C_i \geq t)$. As long as it holds that $Y_i(t) = 1$, we will say that subject i is still at risk at time t (see Figure 3.1). This is quite similar to the Cox model — the jump from 1 to 0 happens as soon as we observe some event.

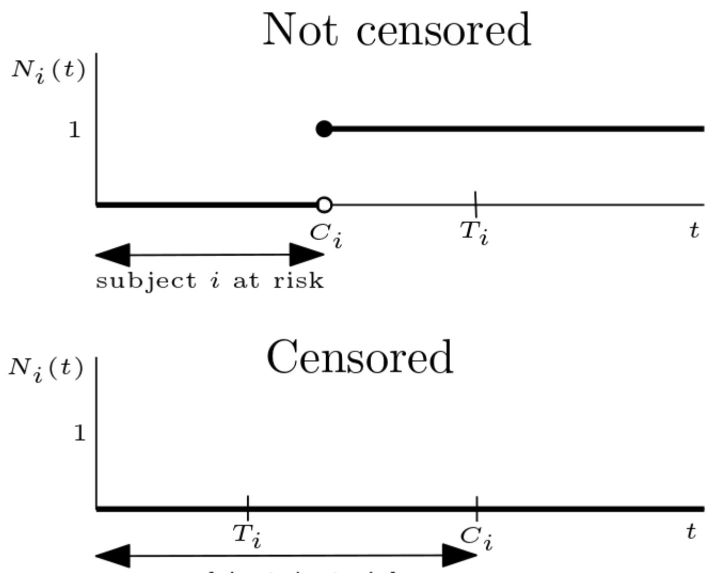

Figure 3.1: The terminology of censoring with current status data used in this thesis.

Suppose now that we define $\lambda(t|\mathbf{Z}_i)$ as the hazard of a jump of $N_i(t)$ for a subject that has not yet been monitored, as did Lin et al. [1998]. They explained that for the jump to happen at time t, we need a) the monitoring to happen at time t and b) the subject i to stay failure-free up to time t. Therefore, $\lambda(t|\mathbf{Z}_i)$ is equal to the product of hazard of C_i and the probability that the failure happens after t, i.e.

$$\lambda(t|\mathbf{Z}_i) = \lambda_C(t|\mathbf{Z}_i) \cdot \mathsf{P}(T_i \geq t|\mathbf{Z}_i(s), s \leq t). \tag{3.4}$$

This is an intuitive explanation, however, we will provide the reader with a more detailed definition of $\lambda(t|\mathbf{Z}_i)$ and a complete proof of the equality (3.4).

Definition 4. *Consider V_i, the time of jump of $N_i(t)$. We then define $\lambda(t|\boldsymbol{Z}_i)$ as*

$$\lim_{h\searrow 0} \frac{P(t \le V_i < t+h | V_i \ge t; C_i \ge t; \boldsymbol{Z}_i(s), s \le t)}{h}.$$

This is an appropriate moment to add that we assume the covariates to be external not just with respect to the failure time, as the definition in 3.1 indicates, but also with respect to the monitoring time. Due to that, the hazard in the definition above would not change if we conditioned it by the whole $\{\mathbf{Z}_i(t), t \ge 0\}$.

Theorem 8. *It holds for $\lambda(t|\boldsymbol{Z}_i)$, the hazard of a jump of $N_i(t)$ for a subject that has not yet been monitored, that*

$$\lambda(t|\boldsymbol{Z}_i) = \lambda_C(t|\boldsymbol{Z}_i)\cdot P(T_i \ge t|\boldsymbol{Z}_i(s), s\le t) = \lambda_0(t)\ e^{-\beta_0^\top \boldsymbol{Z}_i^*(t)+\gamma_o^\top \boldsymbol{Z}_i(t)},$$

where

$$\boldsymbol{Z}_i^*(t) = \int_0^t \boldsymbol{Z}_i(s)ds, \quad \lambda_0(t) = e^{-\Lambda_{T,0}(t)}\lambda_{C,0}(t) \quad and \quad \Lambda_{T,0}(t) = \int_0^t \lambda_{T,0}(s)ds.$$

Proof. The time of jump of $N_i(t)$ (i.e. V_i) equals either C_i (if the process is not censored) or ∞ (if the process is censored, i.e. the jump never happens). Therefore $C_i \ge t$ implies $V_i \ge t$. Using this, the law of total probability and the fact that $\mathsf{P}(T_i < V_i < \infty) = 0$, we get

$$\begin{aligned}\lambda(t|\mathbf{Z}_i) &= \lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le V_i < t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t)\\ &= \lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le V_i < t+h,\ \ T_i \ge t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t)\\ &\quad + \lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le V_i \le T_i < t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t).\end{aligned}$$

We can replace V_i by C_i in both of the summands because $T_i \ge V_i$ and $V_i < \infty$, i.e. $V_i = C_i$ in both of the events (whose probabilities we compute). The second summand then equals

$$\begin{aligned}&\lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le C_i < t+h,\ \ t\le T_i < t+h,\ \ C_i \le T_i|C_i \ge t; \mathbf{Z}_i(s), s\le t)\\ &\le \lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le C_i < t+h,\ \ t\le T_i < t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t)\\ &= \lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le C_i < t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t)\cdot \lim_{h\searrow 0}\mathsf{P}(t\le T_i < t+h|\mathbf{Z}_i(s), s\le t)\\ &= \lambda_C(t|\mathbf{Z}_i)\cdot 0 = 0.\end{aligned}$$

We have used (and will use again with the first summand) the assumption that T_i and C_i are independent given $\mathbf{Z}_i$. So the second summand goes to zero and the first summand equals

$$\begin{aligned}&\lim_{h\searrow 0}\frac{1}{h}\mathsf{P}(t\le C_i < t+h|C_i \ge t; \mathbf{Z}_i(s), s\le t)\cdot \lim_{h\searrow 0}\mathsf{P}(T_i \ge t+h|\mathbf{Z}_i(s), s\le t)\\ &= \lambda_C(t|\mathbf{Z}_i)\ e^{-\Lambda_{T,0}(t)-\beta_0^\top \mathbf{Z}_i^*(t)} = \lambda_0(t)\ e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_o^\top \mathbf{Z}_i(t)}.\end{aligned}$$

We have applied the well-known relationship between survival function and (cumulative) hazard function (Lemma 1). $\square$

Looking at the form of hazard $\lambda(t|\mathbf{Z}_i)$, one can see that we have ended up with a proportional hazards model.

3.2.1 Parameter estimation via one-step partial likelihood approach

We have started with an additive hazards model for the failure time and eventually ended up with a proportional hazards model which will be used for the estimation of parameters. In this section, we will discuss the similarities of interpretation of our model and interpretation of the classical Cox proportional hazards model with right-censored data. We will then show that the existing partial likelihood theory and standard software for fitting the proportional hazards model with right-censored data can be used for our matter.

The event on which we are focusing now (jump of $N_i(t)$) is analogous to the event (failure) that we focused on with the traditional Cox model and right-censored data. The time of event V_i is analogous to the former time of event T_i. In both of the models, the process $N_i(t)$ jumps to 1 if (and at the time) we observe the event (in our model it is actually the definition of the event). The two models have an analogous form of the hazard function (it is not important that we have $(-\beta_0^\top, \gamma_0^\top)^\top$ instead of β_0 and $(\mathbf{Z}_i^*(t)^\top, \mathbf{Z}_i(t)^\top)^\top$ instead of $\mathbf{Z}_i(t)$), but the properties of these hazard functions (and the relationships between the hazard function and the event) differ a little bit. In the Cox model, the independent censoring condition

$$\lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq T_i < t+h | T_i \geq t, \mathbf{Z}_i(t)) = \lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq T_i < t+h | T_i \geq t, C_i \geq t, \mathbf{Z}_i(t))$$

actually tells us that *the hazard of the event* equals *the hazard of the event for a subject that is still at risk.* This does not hold in our case, i.e.

$$\lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq V_i < t+h | V_i \geq t, \mathbf{Z}_i(s), s \leq t)$$
$$\neq \lim_{h \searrow 0} \frac{1}{h} \mathsf{P}(t \leq V_i < t+h | V_i \geq t, C_i \geq t, \mathbf{Z}_i(s), s \leq t).$$

This difference can be pointed out also in another way: in our case, once that the subject is censored, the event can not happen anymore, whereas in the Cox model, the event can happen even after the censoring — it is just not observed. However, this property of the classical Cox model is used only for showing that a certain process is a martingale. If we justify the martingale property also in our case, it will still be possible to use the rest of the theory for Cox model.

Let us denote

$$
\begin{aligned}
\mathbb{Z}_i(t) &= \begin{pmatrix} \mathbf{Z}_i^*(t) \\ \mathbf{Z}_i(t) \end{pmatrix}, \\
\mathbb{S}_n^{(0)}(\beta,\gamma,t) &= \frac{1}{n}\sum_{i=1}^n Y_i(t) e^{-\beta^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)}, \\
\mathbb{S}_n^{(1)}(\beta,\gamma,t) &= \frac{1}{n}\sum_{i=1}^n Y_i(t) e^{-\beta^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} \begin{pmatrix} \mathbf{Z}_i^*(t) \\ \mathbf{Z}_i(t) \end{pmatrix} = \begin{pmatrix} S_n^{*(1)}(\beta,\gamma,t) \\ S_n^{(1)}(\beta,\gamma,t) \end{pmatrix}, \\
\mathbb{S}_n^{(2)}(\beta,\gamma,t) &= \frac{1}{n}\sum_{i=1}^n Y_i(t) e^{-\beta^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} \begin{pmatrix} \mathbf{Z}_i^*(t)^{\otimes 2} & \mathbf{Z}_i^*(t)\mathbf{Z}_i(t)^\top \\ \mathbf{Z}_i(t)\mathbf{Z}_i^*(t)^\top & \mathbf{Z}_i(t)^{\otimes 2} \end{pmatrix} \\
&= \begin{pmatrix} S_n^{*(2)}(\beta,\gamma,t) & S_n^{+(2)}(\beta,\gamma,t) \\ (S_n^{+(2)}(\beta,\gamma,t))^\top & S_n^{(2)}(\beta,\gamma,t) \end{pmatrix}.
\end{aligned}
$$

First of all, we will state again the assumptions needed for the partial likelihood theory, this time modified for our model.

(C.1)* The data are observed on an interval $\langle 0,\tau\rangle$ (i.e. $C_i \in \langle 0,\tau\rangle$ a.s., $i = 1,...,n$) where $\tau > 0$ is fixed, $\int_0^\tau \lambda_0(t)dt < \infty$ and there exists an $\varepsilon > 0$ such that $\mathsf{P}(Y_i(t) = 1) > \varepsilon$ for all $t \in \langle 0,\tau\rangle$ and all $i = 1,...,n$.

(C.2)* All components of the covariate process $\mathbb{Z}_i(t)$ are bounded by a constant on the interval $\langle 0,\tau\rangle$, i.e. there exists $M > 0$ such that $\|\mathbb{Z}_i(t)\| < M$ a.s. for all $t \in \langle 0,\tau\rangle,\ i = 1,...,n$.

(C.3)* There exists a neighbourhood $\mathcal{B}$ of $\begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix}$ and functions $\mathbf{s}^{(0)}(\beta,\gamma,t)$,

$$
\begin{aligned}
\mathbf{s}^{(1)}(\beta,\gamma,t) &= \begin{pmatrix} s^{*(1)}(\beta,\gamma,t) \\ s^{(1)}(\beta,\gamma,t) \end{pmatrix}, \\
\mathbf{s}^{(2)}(\beta,\gamma,t) &= \begin{pmatrix} s^{*(2)}(\beta,\gamma,t) & s^{+(2)}(\beta,\gamma,t) \\ (s^{+(2)}(\beta,\gamma,t))^\top & s^{(2)}(\beta,\gamma,t) \end{pmatrix}
\end{aligned}
$$

defined on $\mathcal{B} \times \langle 0,\tau\rangle$ such that

$$
\sup_{t\in\langle 0,\tau\rangle, (\beta^\top,\gamma^\top)^\top \in \mathcal{B}} \|\mathbb{S}_n^{(j)}(\beta,\gamma,t) - \mathbf{s}^{(j)}(\beta,\gamma,t)\| \xrightarrow{P} 0
$$

as $n \to \infty$, for $j = 0, 1, 2$.

(C.4)* It holds for all $\begin{pmatrix} \beta \\ \gamma \end{pmatrix} \in \mathcal{B}$ and $t \in \langle 0,\tau\rangle$ that

$$
\frac{\partial \mathbf{s}^{(0)}(\beta,\gamma,t)}{\partial \begin{pmatrix} -\beta \\ \gamma \end{pmatrix}} = \mathbf{s}^{(1)}(\beta,\gamma,t) \ \text{ and } \ \frac{\partial \mathbf{s}^{(1)}(\beta,\gamma,t)}{\partial \begin{pmatrix} -\beta \\ \gamma \end{pmatrix}} = \mathbf{s}^{(2)}(\beta,\gamma,t).
$$

(C.5)* The functions $\mathbf{s}^{(j)}$ are bounded and $\mathbf{s}^{(0)}$ is bounded away from 0 on $\mathcal{B} \times \langle 0, \tau \rangle$. For $j = 0, 1, 2$, the family of functions $\{\mathbf{s}^{(j)}(\cdot, t), t \in \langle 0, \tau \rangle\}$ is equicontinuous at β_0.

(C.6)* Define $\mathbf{e}(\beta, \gamma, t) = \frac{\mathbf{s}^{(1)}(\beta,\gamma,t)}{\mathbf{s}^{(0)}(\beta,\gamma,t)}$. The $2p \times 2p$ matrix

$$I(\beta_0, \gamma_0, \tau) = \int_0^\tau \left[\frac{\mathbf{s}^{(2)}(\beta_0, \gamma_0, t)}{\mathbf{s}^{(0)}(\beta_0 \gamma_0, t)} - \mathbf{e}(\beta_0, \gamma_0, t)^{\otimes 2} \right] \mathbf{s}^{(0)}(\beta_0, \gamma_0, t) \lambda_0(t) dt$$

is positive definite.

The assumption (C.2)* could be replaced by the former (C.2) because boundedness of $\mathbf{Z}_i(t)$ implies boundedness of $\mathbf{Z}_i^*(t)$. We state here the conditions (C.3)* to (C.5)* merely to point out the existence and properties of $\mathbf{s}^{(j)}(\beta, \gamma, t),\ j = 0, 1, 2$. Because just like in chapter 2.1, these conditions are fulfilled if (C.2)* is fulfilled and the data are independent and identically distributed, which we already assume at the beginning of this chapter. Moreover, due to the data being i.i.d., $\mathsf{E}\ \mathbb{S}_n^{(j)}(\beta, \gamma, t) = \mathbf{s}^{(j)}(\beta, \gamma, t),\ j = 0, 1, 2$.

Theorem 9. *Let us define processes*

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) e^{-\beta_0^\top \mathbf{Z}_i^*(s) + \gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s), \quad i = 1, ..., n,$$

where $\Lambda_0(t) = \int_0^t \lambda_0(s) ds$. *Then* $M_i(t)$, $i = 1, ..., n$ *are martingales with respect to filtration* $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s_+), s \le t, \mathbf{Z}_i, i = 1, ..., n\}$.

The $\mathbf{Z}_i$ in the filtration stands for the whole process $\{\mathbf{Z}_i(s), s \ge 0\}$. This is another moment where we need the restriction to external covariates. Thanks to the covariates being external, the σ-algebra $\mathcal{F}_t$ can contain $\mathbf{Z}_i(s)$ over all $s \ge 0$ without carrying any information on what happens with $N_i(s)$ or $Y_i(s)$ after time t. Lin et al. [1998] included the covariates $\mathbf{Z}_i(s)$ in $\mathcal{F}_t$ only for $s \le t$. That, as we believe, makes it impossible to proceed with one step of the proof, specifically the equivalence between equalities (3.5) and (3.6) below. In the article, only some parts of the proof were shown, and this equivalence was not among them.

Proof. The random vectors $N_i(t), Y_i(s)$ for $s \le t$ and $(\mathbf{Z}_i^*(s), \mathbf{Z}_i(s))$ for $s \le t$ are clearly $\mathcal{F}_t$-measurable, therefore $M_i(t)$ is $\mathcal{F}_t$-measurable for all $t \ge 0$. Thus, the process $\{M_i(t), t \ge 0\}$ is $\mathcal{F}_t$-adapted.

The absolute moment of $M_i(t)$ is finite due to assumptions (C.1)* and (C.2)*.

Finally, we need to show that $\mathsf{E}\ [M_i(t+u)|\mathcal{F}_t] \overset{\text{a.s.}}{=} M_i(t)$ for any $t, u \ge 0$. This is equivalent to showing that

$$\mathsf{E}\ [N_i(t+u) - N_i(t)|\mathcal{F}_t] = \mathsf{E}\left[\int_t^{t+u} \mathbb{I}(C_i \ge s) e^{-\beta_0^\top \mathbf{Z}_i^*(s) + \gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s) | \mathcal{F}_t\right] \quad (3.5)$$

and that is equivalent to showing that

$$\mathsf{E}\ [N_i(t+u) - N_i(t)|\mathcal{F}_t] = \int_t^{t+u} \mathsf{P}(C_i \ge s|\mathcal{F}_t) e^{-\beta_0^\top \mathbf{Z}_i^*(s) + \gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s) \quad (3.6)$$

because

$$\mathsf{E}\left[\int_t^{t+u} \mathbb{I}(C_i \geq s) e^{-\beta_0^\top \mathbf{Z}_i^*(s)+\gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s) | \mathcal{F}_t\right]$$
$$= \int_t^{t+u} \mathsf{E}\left[\mathbb{I}(C_i \geq s) e^{-\beta_0^\top \mathbf{Z}_i^*(s)+\gamma_o^\top \mathbf{Z}_i(s)} | \mathcal{F}_t\right] d\Lambda_0(s)$$
$$= \int_t^{t+u} \mathsf{P}(C_i \geq s | \mathcal{F}_t) e^{-\beta_0^\top \mathbf{Z}_i^*(s)+\gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s)$$

due to $\mathbf{Z}_i^*(s), \mathbf{Z}_i(s)$ being $\mathcal{F}_t$-measurable for all $s \geq 0$. Because expected value of a random variable that can only be equal to 0 or 1 equals probability that this random variable is qual to 1,

$$\begin{aligned}
&\mathsf{E}\left[N_i(t+u) - N_i(t) | \mathcal{F}_t\right] = \mathsf{P}(N_i(t+u) - N_i(t) = 1 | \mathcal{F}_t) \\
&= \mathsf{P}(N_i(t+u) - N_i(t) = 1 | \mathcal{F}_t, C_i \leq t)\, \mathsf{P}(C_i \leq t | \mathcal{F}_t) \\
&\qquad + \mathsf{P}(N_i(t+u) - N_i(t) = 1 | \mathcal{F}_t, C_i > t)\, \mathsf{P}(C_i > t | \mathcal{F}_t) \\
&= 0 + \mathsf{P}(t < C_i \leq t+u, T_i \geq C_i | \mathcal{F}_t, C_i > t)\, \mathsf{P}(C_i > t | \mathcal{F}_t).
\end{aligned}$$

Now, the conditioning by $\mathcal{F}_t, C_i > t$ can be replaced by $\mathbf{Z}_i, C_i > t$ because once we know that $C_i > t$, the σ-algebra $\mathcal{F}_t$ doesn't bring any new information apart from $\mathbf{Z}_i$. Hence,

$$\mathsf{P}(t < C_i \leq t+u, T_i \geq C_i | \mathcal{F}_t, C_i > t) = \frac{\mathsf{P}(t < C_i \leq t+u, T_i \geq C_i | \mathbf{Z}_i)}{\mathsf{P}(C_i > t | \mathbf{Z}_i)}.$$

Denote by $f_{C|\mathbf{Z}}$ the conditional density of the monitoring time C_i given $\mathbf{Z}_i$. Then the numerator of the fraction above equals $\int_t^{t+u} f_{C|\mathbf{Z}}(s)\, \mathsf{P}(T_i \geq s | \mathbf{Z}_i) ds$. We can compute the density of C_i using the well-known relationship between density, hazard function and survival function (Lemma 1). The survival function of T_i can be computed using the already mentioned formula $S = e^{-\Lambda}$. We then get

$$\begin{aligned}
&\mathsf{E}\left[N_i(t+u) - N_i(t) | \mathcal{F}_t\right] = \frac{\int_t^{t+u} f_{C|\mathbf{Z}}(s)\, \mathsf{P}(T_i \geq s | \mathbf{Z}_i) ds}{\mathsf{P}(C_i > t | \mathbf{Z}_i)} \cdot \mathsf{P}(C_i > t | \mathcal{F}_t) \\
&= \frac{\int_t^{t+u} \mathsf{P}(C_i \geq s | \mathbf{Z}_i) \lambda_{C,0}(s) e^{\gamma_0^\top \mathbf{Z}_(s)} \cdot e^{-\Lambda_{T,0}(s) - \beta_0^\top \mathbf{Z}_i^*(s)} ds}{\mathsf{P}(C_i > t | \mathbf{Z}_i)} \cdot \mathsf{P}(C_i > t | \mathcal{F}_t) \\
&= \int_t^{t+u} \mathsf{P}(C_i \geq s | \mathcal{F}_t, C_i > t) e^{-\beta_0^\top \mathbf{Z}_i^*(s)+\gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s) \cdot \mathsf{P}(C_i > t | \mathcal{F}_t) \\
&= \int_t^{t+u} \mathsf{P}(C_i \geq s | \mathcal{F}_t) e^{-\beta_0^\top \mathbf{Z}_i^*(s)+\gamma_o^\top \mathbf{Z}_i(s)} d\Lambda_0(s).
\end{aligned}$$

□

Estimation with both parameters unknown

We have data $\{N_i(t), Y_i(t), \mathbb{Z}_i(t), t \geq 0\}$, a proportional hazards model $\lambda(t|\mathbf{Z}_i) = \lambda_0(t) e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_0^\top \mathbf{Z}_i(t)}$ and a martingale process $M_i(t)$ that are analogous to the data, model and martingale process from Cox proportional hazards model. We have already pointed out that the relationship between the hazard function and the process $N_i(t)$ differs a little bit, but the role of $N_i(t)$ in the score statistic

$$U_n(\beta, \gamma, \infty) = U_n(\beta, \gamma, \tau) = \sum_{i=1}^n \int_0^\tau \left[\mathbb{Z}_i(s) - \frac{\mathbb{S}_n^{(1)}(\beta, \gamma, s)}{\mathbb{S}_n^{(0)}(\beta, \gamma, s)}\right] dN_i(s) \qquad (3.7)$$

stays the same — it sorts out those subjects (out of n subjects) where we have seen the event. We can therefore use the standard software and the asymptotic properties for the Cox model with right-censored data. If we assume that the observed information matrix

$$I_n(\beta, \gamma, t) = -\frac{1}{n} \frac{\partial U_n(\beta, \gamma, t)}{\partial \begin{pmatrix} -\beta \\ \gamma \end{pmatrix}}$$

is non-singular, we obtain $\hat{\beta}$ and $\hat{\gamma}$, consistent estimates of β_0 and γ_0, with asymptotic distribution (by Theorem 7)

$$\sqrt{n} \left(\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} - \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} \right) \xrightarrow{D} N_{2p} \left(\mathbf{0}, I^{-1}(\beta_0, \gamma_0, \tau) \right).$$

Also, it holds for the observed information matrix that

$$\sup_{t \in \langle 0, \tau \rangle} \| I_n(\hat{\beta}, \hat{\gamma}, t) - I(\beta_0, \gamma_0, t) \| \xrightarrow{P} 0 \tag{3.8}$$

for any $\hat{\beta}$ and $\hat{\gamma}$ that are consistent estimates of β_0 and γ_0. Denote by I_β, $I_{\beta\gamma}$, $I_{\gamma\beta}$ and I_γ the topleft, topright, bottomleft and bottomright blocks of sizes $p \times p$ of the matrix $I(\beta_0, \gamma_0, \tau)$. Then the asymptotic variance of $\hat{\beta}$ is the topleft $p \times p$-sized block of the matrix $I^{-1}(\beta_0, \gamma_0, \tau)$, which equals $(I_\beta - I_{\beta\gamma} I_\gamma^{-1} I_{\gamma\beta})^{-1}$.

Estimation with a known parameter for monitoring

Suppose that we know the true value of γ_0. An example of this is the assumption of complete independence between T_i and C_i, i.e. $\gamma_0 = 0$. If we know the value of γ_0, the estimation reduces to finding such $\hat{\beta}$ that solves $U_{\beta,n}(\hat{\beta}, \gamma_0) = 0$, where $U_{\beta,n}(\beta, \gamma)$ denotes the first p components of $U_n(\beta, \gamma)$ (which correspond to the loglikelihood differentiated only with respect to $-\beta$), i.e.

$$U_{\beta,n}(\beta, \gamma) = \sum_{i=1}^{n} \int_0^\tau \left[\mathbf{Z}_i^*(s) - \frac{S_n^{*(1)}(\beta, \gamma, s)}{\mathbb{S}_n^{(0)}(\beta, \gamma, s)} \right] dN_i(s). \tag{3.9}$$

The asymptotic variance of $\hat{\beta}$ then equals I_β^{-1}, which is the inverse of the limit of $-\frac{1}{n} \frac{\partial U_{\beta,n}(\beta,\gamma)}{\partial (-\beta)^\top}$. Let us now go back and analyze the asymptotic variance of estimation of β_0 with unknown γ_0. The matrix I_γ is positive definite, therefore its inverse matrix is positive definite too and the Cholesky decomposition $I_\gamma^{-1} = LL^\top$ can be applied to it. Notice also that $I_{\gamma\beta} = I_{\beta\gamma}^\top$. We can then write $I_{\beta\gamma} I_\gamma^{-1} I_{\gamma\beta} = (I_{\beta\gamma} L)(I_{\beta\gamma} L)^\top$, which implies that this is a positive semidefinite matrix. Hence, I_β is greater than $I_\beta - I_{\beta\gamma} I_\gamma^{-1} I_{\gamma\beta}$ and I_β^{-1} is smaller than $(I_\beta - I_{\beta\gamma} I_\gamma^{-1} I_{\gamma\beta})^{-1}$. The estimation with known γ_0 is therefore more efficient than the estimation with both parameters unknown, which is not a surprising fact.

3.2.2 Parameter estimation via two-step partial likelihood approach

Let us now get back to the partial likelihood with both parameters unknown. In (3.7), $dN_i(s)$ causes that we only compute values of $\left[\mathbb{Z}_i(s) - \frac{\mathbb{S}_n^{(1)}(\beta,\gamma,s)}{\mathbb{S}_n^{(0)}(\beta,\gamma,s)} \right]$ for those subjects i where the failure happened after the monitoring and we throw away the rest of the summands (it does not mean that we throw away all information about subjects that were censored — these information are included in $\mathbb{S}_n^{(1)}(\beta,\gamma,s)$ and $\mathbb{S}_n^{(0)}(\beta,\gamma,s)$, only not at the times of the subjects' monitoring). This indicates that the partial likelihood applied as in the previous section might not be fully effective. We will now describe and analyze a two-step partial likelihood approach which was introduced by Lin et al. [1998]. This approach first computes $\tilde{\gamma}$, the estimate of γ_0, and then estimates β_0 using the score statistic $U_{\beta,n}(\beta,\tilde{\gamma})$.

Recall the proportional hazards model which was defined for the monitoring time in (3.3) — this is the classical Cox model with the monitoring time playing the role of the event that we are waiting for. Notice that in terms of this Cox model, all n subjects end up as uncensored. Assume that we apply the partial likelihood approach for the Cox model and get a score statistic

$$U_{\gamma,n}(\gamma)=\sum_{i=1}^{n}\int_0^\infty\left[\mathbf{Z}_i(s)-\frac{\frac{1}{n}\sum_{j=1}^n Y_j(s)e^{\gamma^\top \mathbf{Z}_j(s)}\mathbf{Z}_j(s)}{\frac{1}{n}\sum_{j=1}^n Y_j(s)e^{\gamma^\top \mathbf{Z}_j(s)}}\right]dN_i^C(s),$$

where, due to the lack of censoring, $Y_i(s)=\mathbb{I}(C_i\geq s)$ stays the same as before and $N_i^C(s)=\mathbb{I}(C_i\leq s)$. Let the regularity conditions (C.1) to (C.6) for the Cox model (adjusted to model (3.3) with monitoring time) hold. Notice that, again, this score statistic is almost surely equal to $U_{\gamma,n}(\gamma,\tau)$ (where the integral is taken from 0 to τ). We get from section 2.1 that

$$M_i^C(t)=N_i^C(t)-\int_0^t Y_i(s)e^{\gamma_0^\top \mathbf{Z}_i(s)}\lambda_{C,0}(s)ds$$

are martingales for $i=1,...,n$ with respect to the filtration

$$\mathcal{F}_t^C=\sigma\{N_i^C(s),\mathbf{Z}_i(s),s\leq t,i=1,...,n\}$$

and

$$\frac{1}{\sqrt{n}}U_{\gamma,n}(\gamma_0,\tau)\xrightarrow{D}N_p\big(\mathbf{0},J(\gamma_0,\tau)\big),\tag{3.10}$$

where

$$J(\gamma_0,\tau)=\int_0^\tau\left[\frac{s_\gamma^{(2)}(\gamma_0,t)}{s_\gamma^{(0)}(\gamma_0,t)}-e_\gamma(\gamma_0,s)^{\otimes 2}\right]s_\gamma^{(0)}(\gamma_0,t)\lambda_{C,0}(s)ds$$

is the limit (in probability) of $J_n(\gamma,t)=-\frac{1}{n}\frac{\partial U_{\gamma,n}(\gamma,t)}{\partial\gamma^\top}$ at $\gamma=\gamma_0$ and $t=\tau$. Denote by $\tilde{\gamma}$ the partial likelihood estimator of γ_0 that solves $U_{\gamma,n}(\tilde{\gamma})=0$. Then

$$\hat{\gamma}\xrightarrow{P}\gamma_0$$

and

$$\sup_{t\in\langle 0,\tau\rangle}\|J_n(\tilde{\gamma},t)-J(\gamma_0,t)\|\xrightarrow{P}0.\tag{3.11}$$

The convergence (3.11) actually holds for any $\tilde{\gamma}$ that is a consistent estimator of γ_0.

Consider the score statistic in (3.9) for estimating β_0 with γ_0 known. We will substitute the γ_0 "known" by the estimate $\tilde{\gamma}$ described above and then estimate β_0 by such $\tilde{\beta}$ that solves $U_{\beta,n}(\tilde{\beta},\tilde{\gamma})=0$. The following theorems and steps that we will take aim to derivation of asymptotic properties of such $\tilde{\beta}$.

Theorem 10. *Assume that $J_n(\gamma,\tau)$ is non-singular. Denote*

$$s^{(k)}_\gamma(\gamma,t) = \mathsf{E}\, Y_i(t)e^{\gamma_0^\top \boldsymbol{Z}_i(s)}\boldsymbol{Z}_i(t)^{\otimes k},\quad k=0,1,2 \ \ \textit{and}\ \ e_\gamma(\gamma_0,t) = \frac{s^{(1)}_\gamma(\gamma_0,t)}{s^{(0)}_\gamma(\gamma_0,t)}.$$

Then

$$\sqrt{n}(\tilde\gamma-\gamma_0) = J(\gamma_0,\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\{\boldsymbol{Z}_i(t)-e_\gamma(\gamma_0,t)\}dM_i^C(t)+o_P(1). \tag{3.12}$$

Also,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\{\boldsymbol{Z}_i(t)-e_\gamma(\gamma_0,t)\}dM_i^C(t)\quad \textit{is } O_P(1). \tag{3.13}$$

The proof of this theorem is not given in Lin et al. [1998].

Proof. The Taylor series expansion of the score statistic gives us

$$U_{\gamma,n}(\tilde\gamma,\tau) = U_{\gamma,n}(\gamma_0,\tau)+\frac{\partial U_{\gamma,n}(\gamma^*,\tau)}{\partial\gamma^\top}(\tilde\gamma-\gamma_0),$$

where γ^* lies between $\tilde\gamma$ and γ_0. Because the left hand side equals 0,

$$\sqrt{n}(\tilde\gamma-\gamma_0) = \frac{1}{\sqrt{n}}J_n(\gamma^*,\tau)^{-1}U_{\gamma,n}(\gamma_0,\tau). \tag{3.14}$$

Let us first deal with the information matrix. Because $\tilde\gamma \xrightarrow{P} \gamma_0$ and γ^* lies between $\tilde\gamma$ and γ_0, γ^* is also a consistent estimator of γ_0. Using (3.11) and the regularity of $J_n(\gamma,\tau)$ and $J(\gamma_0,\tau)$, we get

$$J_n(\gamma^*,\tau)^{-1} = J(\gamma_0,\tau)^{-1}+o_P(1). \tag{3.15}$$

We know from Theorem 5 that $\frac{1}{\sqrt{n}}U_{\gamma,n}(\gamma_0,\tau)$ equals

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\left[\mathbf{Z}_i(s)-\frac{S^{(1)}_{\gamma,n}(\gamma_0,s)}{S^{(0)}_{\gamma,n}(\gamma_0,s)}\right]dM_i^C(s)\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\left[\mathbf{Z}_i(s)-e_\gamma(\gamma_0,s)\right]dM_i^C(s)+\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_0^\tau\left[e_\gamma(\gamma_0,s)-\frac{S^{(1)}_{\gamma,n}(\gamma_0,s)}{S^{(0)}_{\gamma,n}(\gamma_0,s)}\right]dM_i^C(s),
\end{aligned}$$

where $S^{(k)}_{\gamma,n}(\gamma,t) = \frac{1}{n}\sum_{i=1}^n Y_i(t)e^{\gamma^\top\mathbf{Z}_i(s)}\mathbf{Z}_i(t)^{\otimes k}$ for $k=0,1,2$. Denote the second summand by $V_n^C(\tau)$. Then $\mathsf{E}\,V_n^C(\tau)=0$ (it is a martingale integral) and $\operatorname{var}V_n^C(\tau)=\mathsf{E}\,\langle V_n^C\rangle(\tau)$. By Lemma 3, $\langle V_n^C\rangle(\tau)$ equals

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\int_0^\tau\Bigg[e_\gamma(\gamma_0,s)^{\otimes 2}-e_\gamma(\gamma_0,s)\frac{S^{(1)}_{\gamma,n}(\gamma,t)^\top}{S^{(0)}_{\gamma,n}(\gamma,t)}-\frac{S^{(1)}_{\gamma,n}(\gamma,t)}{S^{(0)}_{\gamma,n}(\gamma,t)}e_\gamma(\gamma_0,s)^\top+\\
&\qquad\qquad+\left(\frac{S^{(1)}_{\gamma,n}(\gamma,t)}{S^{(0)}_{\gamma,n}(\gamma,t)}\right)^{\otimes 2}\Bigg]Y_i(s)e^{\gamma_0^\top\mathbf{Z}_i(s)}d\Lambda_{C,0}(s)\\
&=\sum_{i=1}^n\int_0^\tau\Bigg[e_\gamma(\gamma_0,s)^{\otimes 2}S^{(0)}_{\gamma,n}(\gamma,t)-e_\gamma(\gamma_0,s)S^{(1)}_{\gamma,n}(\gamma,t)^\top-S^{(1)}_{\gamma,n}(\gamma,t)e_\gamma(\gamma_0,s)^\top+\\
&\qquad\qquad+\frac{S^{(1)}_{\gamma,n}(\gamma,t)^{\otimes 2}}{S^{(0)}_{\gamma,n}(\gamma,t)}\Bigg]d\Lambda_{C,0},
\end{aligned}$$

which converges to 0 in probability due to condition (C.3). Hence,

$$\frac{1}{\sqrt{n}}U_{\gamma,n}(\gamma_0,\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[\mathbf{Z}_i(s) - e_{\gamma}(\gamma_0,s)\right]dM_i^C(s) + o_p(1). \tag{3.16}$$

Notice also that due to $\frac{1}{\sqrt{n}}U_{\gamma,n}(\gamma_0,\tau)$ being $O_P(1)$ (Theorem 6), the term $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left[\mathbf{Z}_i(s) - e_{\gamma}(\gamma_0,s)\right]dM_i^C(s)$ is $O_P(1)$ too. From this, (3.15) and (3.14) we get

$$\begin{aligned}\sqrt{n}(\tilde{\gamma}-\gamma_0) &= \left(J(\gamma_0,\tau)^{-1} + o_P(1)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t) - e_{\gamma}(\gamma_0,t)\}dM_i^C(t) + o_P(1)\right)\\ &= J(\gamma_0,\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t) - e_{\gamma}(\gamma_0,t)\}dM_i^C(t) + o_P(1).\end{aligned}$$

□

Lemma 11. *Denote by $e_\beta(\beta,\gamma,t)$ the first p components of $e(\beta,\gamma,t)$, i.e. $e_\beta(\beta,\gamma,t) = \frac{s^{*(1)}(\beta,\gamma,t)}{s^{(0)}(\beta,\gamma,t)}$. Then*

$$\begin{aligned}\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\tilde{\gamma}) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}dM_i(t)\\ &- I_{\beta\gamma}J(\gamma_0,\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\{\boldsymbol{Z}_i(t) - e_{\gamma}(\gamma_0,t)\}dM_i^C(t) + o_P(1).\end{aligned}$$

Proof. Consider the score statistic for β at points β_0,γ_0:

$$\begin{aligned}U_{\beta,n}(\beta_0,\gamma_0) =& \sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)}\right\}dN_i(t)\\ =& \sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)}\right\}dM_i(t)\\ &+ \sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)}\right\}Y_i(t)e^{-\beta_0^\top\mathbf{Z}_i^*(t)+\gamma_o^\top\mathbf{Z}_i(t)}d\Lambda_0(t)\\ =& \sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)}\right\}dM_i(t)\\ =& \sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\right\}dM_i(t)\\ &+ \sum_{i=1}^{n}\int_0^{\tau}\left\{e_\beta(\beta_0,\gamma_0,t) - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)}\right\}dM_i(t).\end{aligned}$$

After multiplying by $\frac{1}{\sqrt{n}}$, the second summand is $o_P(1)$, which can be proved completely analogically as (3.16) in the proof of Theorem 10. Hence,

$$\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\gamma_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\left\{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\right\}dM_i(t) + o_p(1). \tag{3.17}$$

By Taylor series expansion,

$$\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\tilde{\gamma}) = \frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\gamma_0) + \frac{1}{n}\frac{\partial U_{\beta,n}(\beta_0,\gamma^*)}{\partial\gamma^\top}\sqrt{n}(\tilde{\gamma}-\gamma_0),$$

where γ^* lies between γ_0 and $\tilde{\gamma}$. It follows from the consistency of $\tilde{\gamma}$ (and therefore also consistency of γ^*) and from (3.8) that

$$-\frac{1}{n}\frac{\partial U_{\beta,n}(\beta_0,\gamma^*)}{\partial \gamma^\top} \xrightarrow{P} I_{\beta\gamma} \quad \text{as} \quad n \longrightarrow \infty.$$

Combining this, (3.17) and Theorem 10 implies that $\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\tilde{\gamma})$ equals

$$\begin{aligned}
&\sum_{i=1}^{n}\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}\, dM_i(t) + o_p(1)\\
&-(I_{\beta\gamma}+o_P(1))\left(J(\gamma_0,\tau)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}dM_i^C(t) + o_P(1)\right).
\end{aligned}$$

Since we know from the proof of Theorem 10 that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}dM_i^C(t)$$

is $O_P(1)$, the proof of Lemma 11 is complete. □

Theorem 12. *Denote $J_\gamma = J(\gamma_0,\tau)$. Under conditions (C.1)* to (C.6)*,*

$$\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\tilde{\gamma}) \xrightarrow{D} N_p\left(\boldsymbol{0}, I_\beta - I_{\beta\gamma}J_\gamma I_{\beta\gamma}^\top\right).$$

Proof. Recall the equality that was proved in Lemma 11. Since we have martingale integrals, the expected value of the term

$$\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}dM_i(t)$$

is a zero vector and its variance is the expected value of predictable variation taken at time τ, which equals

$$\begin{aligned}
&\mathsf{E}\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}^{\otimes 2}Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_0^\top \mathbf{Z}_i(t)}d\Lambda_0(t)\\
&\quad= \mathsf{E}\,\frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}^{\otimes 2}Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_0^\top \mathbf{Z}_i(t)}d\Lambda_0(t)\\
&\quad= \mathsf{E}\int_0^\tau \Big\{S_n^{*(2)}(\beta_0,\gamma_0,t) - e_\beta(\beta_0,\gamma_0,t)S_n^{*(1)}(\beta_0,\gamma_0,t)^\top\\
&\qquad - S_n^{*(1)}(\beta_0,\gamma_0,t)e_\beta(\beta_0,\gamma_0,t)^\top + e_\beta(\beta_0,\gamma_0,t)^{\otimes 2}\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)\Big\}d\Lambda_0(t)\\
&\quad= \int_0^\tau \left\{\frac{s^{*(2)}(\beta_0,\gamma_0,t)}{\mathbf{s}^{(0)}(\beta_0,\gamma_0,t)} - e_\beta(\beta_0,\gamma_0,t)^{\otimes 2}\right\}\mathbf{s}^{(0)}(\beta_0,\gamma_0,t)d\Lambda_0(t) = I_\beta.
\end{aligned}$$

Analogously, the expected value of

$$\int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}dM_i^C(t)$$

is a zero vector and its variance is J_γ. We will continue with computing the covariance of these two terms. First, let us break down the martingale $M_i(t)$ with respect to which we compute the first integral:

$$dM_i(t) = dN_i(t)\cdot\{\mathbb{I}(T_i \geq t) + \mathbb{I}(T_i < t)\} - Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_0^\top \mathbf{Z}_i(t)}e^{-\Lambda_{T,0}(t)}d\Lambda_{C,0}(t).$$

After multiplying the terms in the brackets, the term $\mathbb{I}(T_i < t) \cdot dN_i(t)$ is clearly zero because after the original failure (with time T_i) happens, the jump of $N_i(t)$ cannot happen anymore. Also, $dN_i(t) \cdot \mathbb{I}(T_i \geq t) = dN_i^C(t) \cdot \mathbb{I}(T_i \geq t)$ because if we know that the original failure has not happened yet, then the jump of $N_i(t)$ at the present time is equivalent to the jump of $N_i^C(t)$. Moreover,

$$dN_i^C(t) \cdot \mathbb{I}(T_i \geq t) = (dM_i^C(t) + Y_i(t)e^{\gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_{C,0}(t)) \cdot \mathbb{I}(T_i \geq t).$$

Also notice that $e^{-\beta_0^\top \mathbf{Z}_i^*(t)+\gamma_0^\top \mathbf{Z}_i(t)} e^{-\Lambda_{T,0}(t)} = \mathsf{P}(T_i \geq t|\mathbf{Z}_i) \cdot e^{\gamma_0^\top \mathbf{Z}_i(t)}$. Altogether,

$$dM_i(t) = \mathbb{I}(T_i \geq t)dM_i^C(t) + \{\mathbb{I}(T_i \geq t) - \mathsf{P}(T_i \geq t|\mathbf{Z}_i)\}Y_i(t)e^{\gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_{C,0}(t).$$

We get

$$\begin{aligned}
&\mathrm{cov}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}dM_i(t), \int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}dM_i^C(t)\right] \\
&= \mathrm{cov}\left[\int_0^\tau \mathbb{I}(T_i \geq t)\{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}dM_i^C(t), \int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}dM_i^C(t)\right] \\
&\quad + \mathrm{cov}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}\{\mathbb{I}(T_i \geq t) - \mathsf{P}(T_i \geq t|\mathbf{Z}_i)\}Y_i(t)e^{\gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_{C,0}(t), \right. \\
&\qquad\qquad\qquad \left. \int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}dM_i^C(t)\right].
\end{aligned}$$

Let us, to simplify the matter, label the first summand by $\mathrm{cov}[A, B^\top]$ and the second summand by $\mathrm{cov}[C, D^\top]$. We will first focus on the second summand. Expected value of the integral D is a zero vector because it is a martingale integral. Thus, the covariance of the two integrals is equal to the expected value of their product. Let us compute this expected value using conditioning by $\mathbf{Z}_i$ and C_i in the following way: $\mathsf{E}\,[C, D^\top] = \mathsf{E}\,[\mathsf{E}\,[C, D^\top|\mathbf{Z}_i, C_i]]$. The integral D is $\sigma\{\mathbf{Z}_i, C_i\}$-measurable and can be pulled out from the conditioned expected value. The summand $\mathrm{cov}[C, D^\top]$ then equals

$$\begin{aligned}
&\mathsf{E}\left[\int_0^\tau \mathsf{E}\left[\{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}\{\mathbb{I}(T_i \geq t) - \mathsf{P}(T_i \geq t|\mathbf{Z}_i)\}Y_i(t)e^{\gamma_0^\top \mathbf{Z}_i(t)}|\mathbf{Z}_i, C_i\right]d\Lambda_{C,0}(t)\right. \\
&\qquad\qquad\qquad \left. \cdot \left(\int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}dM_i^C(t)\right)^\top\right].
\end{aligned}$$

Everything except $\{\mathbb{I}(T_i \geq t) - \mathsf{P}(T_i \geq t|\mathbf{Z}_i)\}$ can be, again, pulled out from the conditioned expected value and $\mathsf{E}\,[\mathbb{I}(T_i \geq t) - \mathsf{P}(T_i \geq t|\mathbf{Z}_i)|\mathbf{Z}_i, C_i] = 0$. Thus, $\mathrm{cov}[C, D^\top] = 0$. Now we need to compute $\mathrm{cov}[A, B^\top]$. Because both A and B are martingale integrals, $\mathrm{cov}[A, B^\top] = \mathsf{E}\; AB^\top$ equals

$$\mathsf{E}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}\{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}^\top\, \mathbb{I}(T_i \geq t)d\langle M_i^C, M_i^C\rangle(t)\right].$$

Furthermore, $\langle M_i^C, M_i^C\rangle(t) = N_i^C(t) - M_i^C(t)$, which implies that

$$\mathrm{cov}[A, B^\top] = \mathsf{E}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0, \gamma_0, t)\}\{\mathbf{Z}_i(t) - e_\gamma(\gamma_0, t)\}^\top\, \mathbb{I}(T_i \geq t)dN_i^C(t)\right]$$

because the integral with respect to $M_i^C(t)$ has a zero expected value. At last, processes $\mathbb{I}(T_i \geq t)dN_i^C(t)$ and $dN_i(t)$ are equal for $t \in \langle 0, \tau\rangle$, from which follows that $\operatorname{cov}[A, B^\top]$ is equal to

$$\begin{aligned}
&\mathsf{E}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}\{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}^\top dN_i(t)\right] \\
&\quad = \mathsf{E}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}\{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}^\top \right. \\
&\qquad\qquad \left. \left(dM_i(t) + Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_0(t)\right)\right] \\
&\quad = \mathsf{E}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\}\{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\}^\top \right. \\
&\qquad\qquad \left. Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_0(t)\right] \\
&\quad = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i^*(t)\mathbf{Z}_i(t)^\top - e_\beta(\beta_0,\gamma_0,t)\mathbf{Z}_i(t)^\top - \mathbf{Z}_i^*(t)e_\gamma(\gamma_0,t)^\top \right. \\
&\qquad\qquad \left. + e_\beta(\beta_0,\gamma_0,t)e_\gamma(\gamma_0,t)^\top\} Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_0(t)\right] \\
&\quad = \int_0^\tau \left\{\frac{s^{+(2)}(\beta_0,\gamma_0,t)}{\mathbf{s}^{(0)}(\beta_0,\gamma_0,t)} - e_\beta(\beta_0,\gamma_0,t)\left(\frac{s^{(1)}(\beta_0,\gamma_0,t)}{\mathbf{s}^{(0)}(\beta_0,\gamma_0,t)}\right)^\top\right\}\mathbf{s}^{(0)}(\beta_0,\gamma_0,t) d\Lambda_0(t).
\end{aligned}$$

Now, let us remind and calculate $I_{\beta\gamma}$, the topright block of matrix $I(\beta_0,\gamma_0,\tau)$. It is equal to the limit (in probability) of

$$\begin{aligned}
-\frac{1}{n}\frac{\partial U_{\beta,n}(\beta_0,\gamma_0)}{\partial \gamma^\top} = \frac{1}{n}\sum_{i=1}^n \int_0^\tau \left\{\frac{S_n^{+(2)}(\beta_0,\gamma_0,t)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)} - \frac{S_n^{*(1)}(\beta_0,\gamma_0,t)S_n^{(1)}(\beta_0,\gamma_0,t)^\top}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,t)^2}\right\}\cdot \\
\cdot Y_i(t)e^{-\beta_0^\top \mathbf{Z}_i^*(t) + \gamma_0^\top \mathbf{Z}_i(t)} d\Lambda_0(t).
\end{aligned}$$

We can now see that

$$\operatorname{cov}\left[\int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\} dM_i(t), \int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\} dM_i^C(t)\right] = I_{\beta\gamma}.$$

Let us now, using Lemma 11, rewrite the statistic whose asymptotic distribution we want to prove:

$$\frac{1}{\sqrt{n}}U_{\beta,n}(\beta_0,\tilde{\gamma}) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n U_i - 0\right) + o_P(1),$$

$$U_i = \int_0^\tau \{\mathbf{Z}_i^*(t) - e_\beta(\beta_0,\gamma_0,t)\} dM_i(t) - I_{\beta\gamma}J(\gamma_0,\tau)^{-1}\int_0^\tau \{\mathbf{Z}_i(t) - e_\gamma(\gamma_0,t)\} dM_i^C(t),$$

where U_i are i.i.d. random vectors with zero mean and a finite covariance matrix equal to

$$I_\beta - I_{\beta\gamma}\left(I_{\beta\gamma}J_\gamma^{-1}\right)^\top - I_{\beta\gamma}J_\gamma^{-1}I_{\beta\gamma}^\top + I_{\beta\gamma}J_\gamma^{-1}J_\gamma\left(I_{\beta\gamma}J_\gamma^{-1}\right)^\top = I_\beta - I_{\beta\gamma}J_\gamma^{-1}I_{\beta\gamma}^\top.$$

After applying Central limit theorem to $\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n U_i - 0\Big)$, the proof is complete. $\square$

The last thing that we need before we show the asymptotic distribution of $\tilde{\beta}$ is the proof of consistency of this two-step estimator. Lin et al. [1998] do not present this proof and suggest that it is similar to that of the ordinary maximum partial likelihood from the Cox model. However, it is not completely analogical (the consistency from Cox model cannot be applied to it like we did in the case of the one-step estimator). That is why we present the proof here. We will not build it from the beginning, but use some facts that have already been proved for the case of the ordinary maximum partial likelihood from the Cox model (and that can be applied to our case). These facts are summed up in the following two lemmata.

Lemma 13. *Let E be an open convex subset of $\mathbb{R}^q, q \in \mathbb{N}$ and let $F_1, F_2, ...,$ be a sequence of random concave functions on E and f a real-valued function on E such that, for all $x \in E$,*

$$F_n(x) \xrightarrow{P} f(x) \quad as \quad n \to \infty.$$

Let F_n have a unique maximum at X_n for all $n \in \mathbb{N}$ and let f have one at x. Then $X_n \xrightarrow{P} x$ as $n \to \infty$.

The proof of this lemma can be found in Andersen and Gill [1982].

Lemma 14. *Let the conditions (C.1)* to (C.6)* hold. Recall the partial log-likelihood*

$$l_n(\beta,\gamma) = \sum_{i=1}^{n} \int_0^{\tau} \Big[-\beta^\top \boldsymbol{Z}_i^*(s) + \gamma^\top \boldsymbol{Z}_i(s) - \log n\mathbb{S}_n^{(0)}(\beta,\gamma,s)\Big] dN_i(s)$$

and denote

$$X_n(\beta,\gamma) = \frac{1}{n}\left(l_n(\beta,\gamma) - l_n(\beta_0,\gamma_0)\right)$$

and

$$A(\beta,\gamma) = \int_0^{\tau} \Bigg[\left(\begin{pmatrix}-\beta\\ \gamma\end{pmatrix} - \begin{pmatrix}-\beta_0\\ \gamma_0\end{pmatrix}\right)^\top \mathbb{S}_n^{(1)}(\beta_0,\gamma_0,s) \\ - \log\left(\frac{\mathbb{S}_n^{(0)}(\beta,\gamma,s)}{\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,s)}\right)\mathbb{S}_n^{(0)}(\beta_0,\gamma_0,s)\Bigg]\lambda_0(s)ds.$$

Then $A(\beta,\gamma)$ has a unique maximum at $\begin{pmatrix}\beta\\ \gamma\end{pmatrix} = \begin{pmatrix}\beta_0\\ \gamma_0\end{pmatrix}$ and $X_n(\beta,\gamma)$ converges to $A(\beta,\gamma)$ in probability as $n \to \infty$ for all $\begin{pmatrix}\beta\\ \gamma\end{pmatrix} \in \mathcal{B}$.

The proof of this lemma can be found in Fleming and Harrington [1991] as a part of proof of consistency of the ordinary maximum partial likelihood estimator from the Cox model.

Theorem 15. *Assume that $-\frac{1}{n}\frac{\partial U_{\beta,n}(\beta,\gamma)}{\partial(-\beta)^\top}$ is non-singular. Under conditions (C.1)* to (C.6)* and under conditions (C.1) to (C.6) modified for the estimation of γ_0,*

$$\tilde{\beta} \xrightarrow{P} \beta_0.$$

Proof. Denote $\mathcal{X}_n(\beta) = X_n(\beta,\tilde{\gamma})$ and $\mathcal{A}(\beta) = A(\beta,\gamma_0)$. We are going to use these two functions together with Lemma 13 to prove the consistency of $\tilde{\beta}$. Note that $\frac{\partial \mathcal{X}_n(\beta)}{\partial \beta} = -\frac{1}{n}U_{\beta,n}(\beta,\tilde{\gamma})$ and $\frac{\partial \mathcal{X}_n(\beta)}{\partial\beta\partial\beta^\top} = \frac{1}{n}\frac{\partial U_{\beta,n}(\beta,\gamma)}{\partial(-\beta)^\top}$, which is a negative definite matrix. Thus, $\mathcal{X}_n(\beta)$ is a concave function and has a unique maximum at $\beta = \tilde{\beta}$ for all $n \in \mathbb{N}$. We get from Lemma 14 that $\mathcal{A}(\beta)$ has a unique maximum at $\beta = \beta_0$ and that

$$X_n(\beta, \gamma_0) \xrightarrow{P} \mathcal{A}(\beta)$$

as $n \to \infty$ for all $\beta \in \mathcal{B}^*$, where $\mathcal{B}^* = \{\beta : (\beta, \gamma_0)^\top \in \mathcal{B}\}$. Thus, we only need to show that

$$(\mathcal{X}_n(\beta) - X_n(\beta, \gamma_0)) \xrightarrow{P} 0, \quad n \to \infty, \quad \beta \in \mathcal{B}^*.$$

By Taylor series expansion,

$$\mathcal{X}_n(\beta) - X_n(\beta, \gamma_0) = X_n(\beta, \tilde{\gamma}) - X_n(\beta, \gamma_0) = \frac{\partial X_n(\beta, \gamma^*)}{\partial \gamma}(\tilde{\gamma} - \gamma_0),$$

where γ^* lies between $\tilde{\gamma}$ and γ_0. We know that $(\tilde{\gamma} - \gamma_0)$ is $o_P(1)$ and therefore it remains to show that $\frac{\partial X_n(\beta,\gamma^*)}{\partial \gamma}$ is $O_P(1)$. We have

$$\begin{aligned}
\left\| \frac{\partial X_n(\beta, \gamma^*)}{\partial \gamma} \right\| &= \left\| \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left[\mathbf{Z}_i(s) - \frac{S_n^{(1)}(\beta, \gamma^*, s)}{\mathbb{S}_n^{(0)}(\beta, \gamma^*, s)} \right] dN_i(s) \right\| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{s \in \langle 0, \tau \rangle} \left\| \mathbf{Z}_i(s) - \frac{S_n^{(1)}(\beta, \gamma^*, s)}{\mathbb{S}_n^{(0)}(\beta, \gamma^*, s)} \right\| \\
&\leq \frac{1}{n} \sum_{i=1}^{n} \sup_{s \in \langle 0, \tau \rangle} \left\| \mathbf{Z}_i(s) \right\| + \sup_{s \in \langle 0, \tau \rangle} \left\| \frac{S_n^{(1)}(\beta, \gamma^*, s)}{\mathbb{S}_n^{(0)}(\beta, \gamma^*, s)} \right\|.
\end{aligned}$$

The first summand is $\leq M$ a.s. due to condition (C.2)*. The second summand is $\leq \sup_{s \in \langle 0, \tau \rangle, (\beta^\top, \gamma^\top)^\top \in \mathcal{B}} \left\| \frac{S_n^{(1)}(\beta,\gamma,s)}{\mathbb{S}_n^{(0)}(\beta,\gamma,s)} \right\|$ and that is $O_P(1)$ due to conditions (C.3)* and (C.5)*. Thus, $\frac{\partial X_n(\beta,\gamma^*)}{\partial \gamma}$ is $O_P(1)$ and therefore $\mathcal{X}_n(\beta) \xrightarrow{P} \mathcal{A}(\beta)$ as $n \to \infty$ for all $\beta \in \mathcal{B}^*$. Lemma 13 now implies the consistency of $\tilde{\beta}$. □

Theorem 16. *Assume that $-\frac{1}{n}\frac{\partial U_{\beta,n}(\beta,\gamma)}{\partial(-\beta)^\top}$ is non-singular. Let the conditions (C.1)* to (C.6)* and the conditions (C.1) to (C.6) modified for the estimation of γ_0 hold. Let $\tilde{\beta}$ be the estimate of β_0 that solves $U_{\beta,n}(\tilde{\beta}, \tilde{\gamma}) = 0$. Then*

$$\sqrt{n}(\tilde{\beta} - \beta_0) \xrightarrow{D} N_p\left(\boldsymbol{0}, I_\beta^{-1} - I_\beta^{-1} I_{\beta\gamma} J_\gamma^{-1} I_{\beta\gamma}^\top I_\beta^{-1}\right).$$

Proof. By Taylor series expansion,

$$0 = U_{\beta,n}(\tilde{\beta}, \tilde{\gamma}) = U_{\beta,n}(\beta_0, \tilde{\gamma}) + \frac{\partial U_{\beta,n}(\beta^*, \tilde{\gamma})}{\partial \beta^\top}(\tilde{\beta} - \beta_0),$$

where β^* lies between $\tilde{\beta}$ and β_0. This implies that

$$\sqrt{n}(\tilde{\beta} - \beta_0) = -\left(\frac{1}{n}\frac{\partial U_{\beta,n}(\beta^*, \tilde{\gamma})}{\partial \beta^\top}\right)^{-1} \frac{1}{\sqrt{n}} U_{\beta,n}(\beta_0, \tilde{\gamma}).$$

Because $\tilde{\beta}$ is a consistent estimator of β_0, so is β^*. The matrix $\frac{1}{n}\frac{\partial U_{\beta,n}(\beta^*,\tilde{\gamma})}{\partial \beta^\top}$ is positive definite and its inverse converges to I_β^{-1}. This completes, together with Theorem 12, the proof. □

The following derivations will be similar to those at the end of section 3.2.1. The matrix J_γ is positive definite, therefore its inverse matrix is positive definite too and the Cholesky decomposition $J_\gamma^{-1} = LL^\top$ can be applied to it. We can then write $I_\beta^{-1} I_{\beta\gamma} J_\gamma^{-1} I_{\beta\gamma}^\top I_\beta^{-1} = (I_\beta^{-1} I_{\beta\gamma} L)(I_\beta^{-1} I_{\beta\gamma} L)^\top$, which implies that this is a positive semidefinite matrix. Therefore, the asymptotic variance of the two-step estimator $I_\beta^{-1} - I_\beta^{-1} I_{\beta\gamma} J_\gamma^{-1} I_{\beta\gamma}^\top I_\beta^{-1}$ is smaller than I_β^{-1}, which is the asymptotic variance of estimator with γ_0 known, and that is smaller than $(I_\beta - I_{\beta\gamma} I_\gamma^{-1} I_{\gamma\beta})^{-1}$, which is the asymptotic variance of the one-step estimator. This implies that the two-step approach is the most efficient of the three mentioned approaches. It is more efficient for the estimation of β_0 to estimate γ_0 from the data first, even if we know the true value of γ_0.

3.2.3 Estimation with semiparametric efficient score function

Martinussen and Scheike [2002] developed a method to analyze current status data under the additive hazards model without the need to assume proportional hazards (or any other) model for monitoring time. In this section, we will briefly sum up the description of this method and its comparison with the method of Lin et al. [1998] from the article.

Assume that the additive hazards model (3.2) holds for time to event and that we only obtain the triplets of $\{C_i, \delta_i, \mathbf{Z}_i\}$, as described in the previous section. The covariates can be time-dependent, though it is not properly highlited in the article (nor the necessity to restrict ourselves to external covariates). The theory begins with expressing the loglikelihood function

$$l(\beta, \Lambda_{T,0}) = \sum_{i=1}^{n} \delta_i \log(e^{-\Lambda_{T,0}(C_i) - \beta^\top \mathbf{Z}_i^*(C_i)}) + (1-\delta_i)\log(1 - e^{-\Lambda_{T,0}(C_i) - \beta^\top \mathbf{Z}_i^*(C_i)}),$$

this leads to $U_n(\beta, \Lambda_{T,0})$, the empirical version of the efficient score for β_0, which includes the processes

$$N_{1i}(t) = \delta_i \, \mathbb{I}(C_i \le t) \quad \text{and} \quad N_{2i}(t) = (1-\delta_i)\, \mathbb{I}(C_i \le t).$$

This is compared with (3.7), where we only use $N_{1i}(t)$. The estimator $\hat{\beta}_{MS}$ is defined as the solution to

$$U_n(\beta, \widehat{\Lambda_{T,0}}) = 0,$$

where $\widehat{\Lambda_{T,0}}$ is an estimator of $\Lambda_{T,0}$. If it is a uniformly consistent estimator of $\Lambda_{T,0}$, then $\hat{\beta}_{MS}$ converges in distribution to a zero-mean normal random vector and it is also efficient — which is an important advantage of this method, compared to $\hat{\beta}_{LOY}$, the estimator by Lin et al. [1998]. In the article, the theory is discussed separately for the cases of monitoring time independent of/dependent on the covariates.

Martinussen and Scheike [2002] performed two simulations studies, both compared the estimators $\hat{\beta}_{MS}$ and $\hat{\beta}_{LOY}$. In the first study, the monitoring time is generated independently of the covariates (and, of course, also of time to event). It comes out from this study that $\hat{\beta}_{MS}$ is indeed more efficient — its standard error is 25% (resp. 17%) lower than the standard error of $\hat{\beta}_{LOY}$ for $n = 100$ (resp.

$n = 200$). On the other hand, in the case of $\hat{\beta}_{MS}$, nonconvergence occured in 12% (resp. 5%) of the 10000 replications for $n = 100$ (resp. $n = 200$). The second study was built to demonstrate the other advantage of the method proposed by Martinussen and Scheike [2002] — we do not need to assume proportional hazards model for the monitoring time. This is, without any doubt, an advantage. The consequences of violating this assumption in the method of Lin et al. [1998] are unknown. However, we believe that this study, specifically the part with estimates $\hat{\beta}_{LOY}$, was not done entirely correctly. This will be described in more detail in chapter 4.

To sum up, the advantages of the method of Martinussen and Scheike [2002] lie undoubtedly in its efficiency and lighter assumptions on the monitoring time. However, the method of Lin et al. [1998] can still be sometimes prefered, especially because of its elegance and simplicity (we do not have to estimate any form of baseline hazard) and also lower number of cases with nonconvergence. Also, the assumption on proportional hazards for monitoring time is not that restrictive when it comes to controlled experiments. Moreover, it is always possible to test whether this assumption is fulfilled, because we have all the observations of monitoring time C_i.

4. Simulation study

In the last chapter, we would like to describe a simulation study that we performed to demonstrate the properties of the estimation method proposed by Lin et al. [1998]. We also intend to show the differences between using three different approaches to this method — the one-step approach with both parameters unknown (3.2.1), the approach using a known parameter for monitoring (3.2.1) and, at last, the two-step approach (3.2.2). Before providing the results, we will describe the generating procedure used for our data.

4.1 Generating time to event

For the simulation study, we needed to generate observations of time to event (and monitoring time/censoring time) from a piecewise exponential distribution (probability distribution with a piecewise constant hazard function) and also from a distribution with hazard function $\lambda(t) = \lambda_0(t) + \beta^\top \mathbf{Z}(t)$, where $\lambda_0(t)$ is the hazard function of Weibull distribution. For the former, we have used a function in R from the package `cpsurvsim` (Hochheimer [2019]) which, according to the words of the author, draws on the work from Walke [2010]. For the latter, we have built a function which works analogously as the functions from package `cpsurvsim`. We will describe the generating process and add a proof showing that the generated observations are from the desired distribution.

4.1.1 Piecewise exponential distribution

Suppose that we want to generate an observation from a distribution with piecewise hazard function

$$\lambda(t) = \lambda_i, \quad \tau_{i-1} \leq t < \tau_i, \;\; i = 1, ..., m,$$

where $\tau_0 = 0, \tau_m = \infty$ and $\lambda_1, ..., \lambda_m \in (0, \infty)$ are the values and $\tau_1, ..., \tau_{m-1} \in (0, \infty)$ the change points defining the distribution. Note that $\Lambda(0) = 0,\; \Lambda(\infty) = \infty$ and compute values of the cumulative hazard function at the change points:

$$\Lambda(\tau_i) = \Lambda(\tau_{i-1}) + \lambda_i(\tau_i - \tau_{i-1}) = \sum_{j=1}^{i} \lambda_j(\tau_j - \tau_{j-1}), \quad i = 1, ..., m-1.$$

Consider now a random variable $X \sim \text{Exp}(1)$. We will obtain T as a transformation of X that satisfies $\Lambda(T) = X$. Taking into consideration that $\Lambda(t)$ has different forms on different intervals $\langle \tau_{i-1}, \tau_i)$, we get

$$\begin{aligned} \Lambda(\tau_{i-1}) + \lambda_i(T - \tau_{i-1}) = X, & \qquad X \in \langle \Lambda(\tau_{i-1}), \Lambda(\tau_i)) \\ T = \frac{X - \Lambda(\tau_{i-1})}{\lambda_i} + \tau_{i-1}, & \qquad X \in \langle \Lambda(\tau_{i-1}), \Lambda(\tau_i)), i = 1, ..., m. \end{aligned}$$

We will show that if we create observations by generating from Exp(1) and then applying the transformation above, those observations will follow the desired distribution with piecewise constant hazard function $\lambda(t)$. Because the cumulative distribution function $\Lambda(t)$ is increasing and therefore has an inverse function, we can write

$$\lim_{h\searrow 0}\frac{\mathsf{P}(t\leq T<t+h|T\geq t)}{h}=\lim_{h\searrow 0}\frac{\mathsf{P}(t\leq \Lambda^{-1}(X)<t+h\mid \Lambda^{-1}(X)\geq t)}{h}$$
$$=\lim_{h\searrow 0}\frac{\mathsf{P}(\Lambda(t)\leq X<\Lambda(t)+\int_t^{t+h}\lambda(s)ds\mid X\geq \Lambda(t))}{h},$$

which equals, if h is small enough,

$$=\lim_{h\searrow 0}\frac{\mathsf{P}(\Lambda(t)\leq X<\Lambda(t)+\lambda_i h\mid X\geq \Lambda(t))}{h},\quad t\in\langle\tau_{i-1},\tau_i)$$
$$=\lim_{h\searrow 0}\frac{\mathsf{P}(\Lambda(t)\leq X<\Lambda(t)+\lambda_i h\mid X\geq \Lambda(t))}{\lambda_i h}\cdot\lambda_i=\lambda_i,\quad t\in\langle\tau_{i-1},\tau_i).$$

This holds for any $i=1,...,m$.

4.1.2 Baseline hazard from Weibull distribution

Suppose that we want to generate an observation from a distribution with hazard function of the form

$$\lambda(t)=\frac{a}{b^a}t^{a-1}+\theta_i,\quad \tau_{i-1}\leq t<\tau_i,\ \ i=1,...,m,$$

where $a,b>0,\tau_0=0,\tau_m=\infty$ and $\theta_1,...,\theta_m\in\langle 0,\infty)$ are the values and $\tau_1,...,\tau_{m-1}\in(0,\infty)$ the change points defining the distribution in different time intervals. The values of cumulative hazard function at change points are

$$\Lambda(\tau_i)=\frac{\tau_i^a}{b^a}+\sum_{j=1}^{i}\theta_j(\tau_j-\tau_{j-1}),\ \ i=1,...,m-1.$$

Just like in section 4.1.1, we will consider a random variable $X\sim\text{Exp}(1)$ and look for such T, a transformation of X, that fulfills $\Lambda(T)=X$, in this case

$$\Lambda(\tau_{i-1})+\frac{T^a}{b^a}+\theta_i T-\frac{\tau_{i-1}^a}{b^a}-\theta_i\tau_{i-1}=X,\ \ X\in\langle\Lambda(\tau_{i-1}),\Lambda(\tau_i))$$
$$\frac{T^a}{b^a}+\theta_i T+\sum_{j=1}^{i-1}\tau_j(\theta_j-\theta_{j+1})=X,\ \ X\in\langle\Lambda(\tau_{i-1}),\Lambda(\tau_i)),\ \ i=1,...,m.$$

For $a=1$, this represents the case from previous section. For $a=2$, we get

$$T=\frac{b^2}{2}\left(-\theta_i+\sqrt{\theta_i^2-\frac{4}{b^2}(\sum_{j=1}^{i-1}\tau_j(\theta_j-\theta_{j+1})-X)}\right),\ \ X\in\langle\Lambda(\tau_{i-1}),\Lambda(\tau_i)).$$

Let us verify that this transformation gives us a random variable with the desired hazard function. As in the previous section,

$$\lim_{h\searrow 0}\frac{\mathsf{P}(t\leq T<t+h|T\geq t)}{h}=\lim_{h\searrow 0}\frac{\mathsf{P}(\Lambda(t)\leq X<\Lambda(t)+\int_t^{t+h}\lambda(s)ds\mid X\geq \Lambda(t))}{h}$$

and if h is small enough, that equals

$$
\begin{aligned}
&\lim_{h\searrow 0} \frac{\mathsf{P}(\Lambda(t) \leq X < \Lambda(t) + \frac{h^2}{b^2} + h(\frac{2t}{b^2} + \theta_i) \mid X \geq \Lambda(t))}{h} \\
&= \lim_{h\searrow 0} \frac{\mathsf{P}(\Lambda(t) \leq X < \Lambda(t) + \frac{h^2}{b^2} + h(\frac{2t}{b^2} + \theta_i) \mid X \geq \Lambda(t))}{\frac{h^2}{b^2} + h(\frac{2t}{b^2} + \theta_i)} \cdot \left(\frac{h}{b^2} + \frac{2t}{b^2} + \theta_i\right) \\
&= \frac{2t}{b^2} + \theta_i, \;\; t \in \langle \tau_{i-1}, \tau_i).
\end{aligned}
$$

4.2 Details and results of the study

The simulations were performed in the program R using the Monte Carlo principle with $B = 10000$ repetitions. The summaries of results in tables were computed from those repetitions that converged. Among all the studies, the maximum of nonconverging cases out of 10000 simulations was 3. We use two methods to analyze the generated data. The first (and the main) one is the method for current status data proposed by Lin et al. [1998] and we perform it in three versions — one-step approach, approach with γ_0 known and two-step approach (sections 3.2.1 and 3.2.2). Their implementation in R included adjustments of data (mostly splitting observations into time intervals and computing the integrals of covariates at different time points), construction of the correct formula (based on the number of covariates and the chosen approach) and application of existing functions `coxph` and `Surv` (from package `survival`, Therneau [2020]), designed to analyze time-to-event data under the Cox model. The second method is a method for right-censored data with additive hazards assumption proposed by Lin and Ying [1994] and described in section 2.2. We implemented this method similarly as was described in Zavřelová [2020], but adjusted it for time-dependent covariate.

For the majority of this study, we consider the following two covariates:

$$Z_1(t) = Z_1 \sim \text{Unif}(0, 3), \;\; t \geq 0$$

and

$$Z_2(t) = W \cdot \mathbb{I}(t \geq 0.5), \;\; t \geq 0, \;\; W \sim \text{Bernoulli}\left(\frac{1}{2}\right).$$

The first covariate does not change over time and the second covariate starts at value 0 and either stays equal to 0 or jumps to 1 at $t = 0.5$. The fact whether $Z_2(t)$ jumps to 1 or not differs among the generated observations (subjects) and is known from the beginning $t = 0$ (otherwise the condition of external covariates would not be met). If not specified differently, these two covariates are independent. The parameters defining their effect on time to event and monitoring time (or censoring time) are set to $\beta_0 = (0.5, 0.5)^\top$ and $\gamma_0 = (1, 1)^\top$. Because the generating procedure of time to event and its censoring was done in more ways, we are describing it together with the results in smaller sections below.

4.2.1 Current status data

In this section, time to event was censored to the form of current status data. We have tried more alternatives by changing the baseline hazard, fulfilment of conditions and relationships between the covariates. A simple summary of the character of datasets generated according to these alternatives can be found in Table 4.1. All simulated data in this section were analyzed using each of the three approaches (one-step, γ_0 known, two-step).

Data	failures	failures (jump in $Z_2(t)$)	avg. Z_1 (failure)	avg. Z_1 (no failure)
$\lambda_{T,0} = \frac{3}{2}$	122.7	19.4	1.5	1.3
$\lambda_{T,0} = \frac{8}{9}t$	121.8	18.6	1.7	1.5
$\lambda_{T,0} = \frac{3}{2}$, cov. correlated	121.7	12.1	1.6	1.3
$\lambda_{T,0} = \frac{3}{2}$, addhaz monitoring	124.5	18.2	1.5	1.5
$\lambda_{T,0} = \frac{8}{9}t$, addhaz monitoring	124.9	20.2	1.5	1.5

Table 4.1: Summary of 1000 generations of data with $n = 200$. The average number of failures, average number of failures among observations with a jump in covariate $Z_2(t)$, average value of Z_1 among observations with a failure, average value of Z_1 among observations without a failure.

Constant baseline hazard of time to event

The distribution of covariates is described above. In this simulation, time to event T follows distribution with the hazard function

$$\lambda_T(t|\mathbf{Z}) = \lambda_{T,0}(t) + \beta_0^\top \mathbf{Z}_(t) = \frac{1}{3} + 0.5Z_1 + 0.5Z_2(t), \quad t \geq 0$$

and monitoring time C follows distribution with the hazard function

$$\lambda_C(t|\mathbf{Z}) = \lambda_{C,0}(t) \cdot e^{\gamma_0^\top \mathbf{Z}_(t)} = \frac{1}{3} \cdot e^{Z_1 + Z_2(t)}, \quad t \geq 0.$$

With our specific covariates, both of these conditional hazard functions are piecewise constant and we could use the generating procedure described in section 4.1.1. So in each of $B = 10000$ repetitions, we generated n independent copies of $(Z_1, Z_2(t), T, C)$ and censored them to the form of current status data. We can see an example of such data in the Figure 4.1. The time to event and monitoring time are joined with a line for each individual observation. Remember that we regard the observations as censored if $T < C$, this is marked by red colour. The plot indicates that the covariate Z_1 has a greater effect on the monitoring time than on the time to event. This agrees with the fact that $\gamma_0 > \beta_0$ and the fact that the monitoring time has proportional hazards model (and the time to event only additive hazards).

We have performed these simulations for the cases of $n = 200, 400, 800$, the results are recorded in the Table 4.2. For each option of estimates $\hat{\beta}_1, \hat{\beta}_2$ (options of n and chosen approach), we have computed the mean and standard error of $\hat{\beta}_1, \hat{\beta}_2$ over the 10000 repetitions. We have also computed the means of $\hat{\sigma}_1, \hat{\sigma}_2$, which are the estimates of $\sqrt{\mathrm{var}(\hat{\beta}_j)}$, $j = 1, 2$ that were computed using the results from the very end of section 3.2.2 and the standard observed information matrix from the Cox model partial likelihood method. We also added, for every option of estimation, the coverage of the true parameter by confidence intervals.

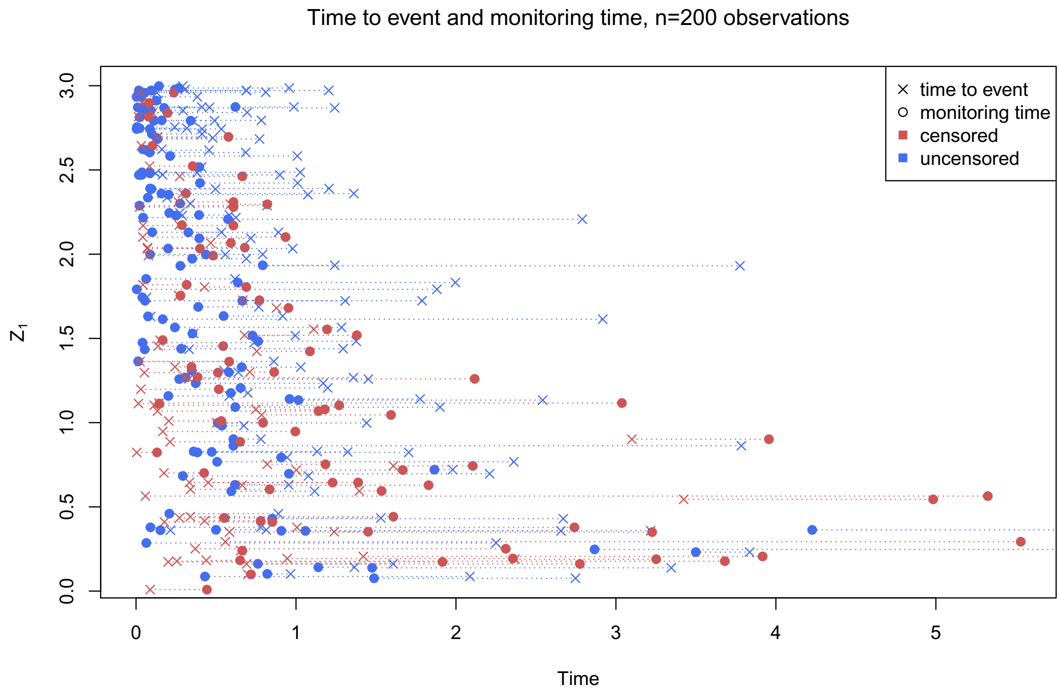

Figure 4.1: An example of dataset generated in the simulation study. The current status censoring is marked using colours. The so-called censored observations are those where time to event comes before the monitoring time.

For all options of n, the two-step approach has the lowest variability and the one-step approach the largest variability, just as was concluded at the end of section 3.2.2. For example, for $n = 200$, the standard error of $\hat{\beta}_1$ is 0.42 with the one-step approach, 0.29 with the approach with γ_0 known and 0.24 with the two-step approach. The approach with γ_0 known is the least biased, which seems reasonable, given that we provide the computation with the real value of γ_0, while in the other approaches, it has to be estimated. For $n = 200$, the mean of $\hat{\beta}_2$ is 0.68 with the one-step approach, 0.54 with the approach with γ_0 known and 0.57 with the two-step approach. The estimation of β_2 is naturally more problematic because the monitoring often occurs before $t = 0.5$, the time of possible jump of $Z_2(t)$. This is even more noticeable with the one-step approach, which converges very slowly — the mean and standard error of $\hat{\beta}_2$ are 0.68 and 1.32, 0.59 and 0.70, 0.55 and 0.44 for n equal to $200, 400, 800$, respectively.

Linear baseline hazard of time to event

Assume now that the time to event of a subject with all covariates equal to 0 has Weibull$(2, 1.5)$ distribution, that is,

$$\lambda_T(t|\mathbf{Z}) = \lambda_{T,0}(t) + \beta_0^\top \mathbf{Z}(t) = \frac{8}{9}t + 0.5Z_1 + 0.5Z_2(t), \quad t \geq 0.$$

The monitoring time stays with the baseline hazard $\lambda_{C,0}(t) = \frac{1}{3}$. Such simulations were performed for $n = 200$ and $n = 400$. The results of these simulations can be found in Table 4.3. The estimates have greater variability than in the case with

Data	Parameter		One-step	γ_0 known	Two-step
Current status, $\lambda_{T,0} = \frac{1}{3}$, $n = 200$	β_1	Mean($\hat{\beta}_1$)	0.51	0.50	0.51
		SE($\hat{\beta}_1$)	0.42	0.29	0.24
		Mean($\hat{\sigma}_1$)	0.41	0.29	0.24
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	0.68	0.54	0.57
		SE($\hat{\beta}_2$)	1.32	0.79	0.70
		Mean($\hat{\sigma}_2$)	1.07	0.74	0.66
		Coverage by CI	97 %	96 %	97 %
Current status, $\lambda_{T,0} = \frac{1}{3}$, $n = 400$	β_1	Mean($\hat{\beta}_1$)	0.50	0.50	0.50
		SE($\hat{\beta}_1$)	0.28	0.20	0.17
		Mean($\hat{\sigma}_1$)	0.28	0.20	0.16
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	0.59	0.53	0.54
		SE($\hat{\beta}_2$)	0.70	0.48	0.43
		Mean($\hat{\sigma}_2$)	0.65	0.47	0.42
		Coverage by CI	95 %	95 %	95 %
Current status, $\lambda_{T,0} = \frac{1}{3}$, $n = 800$	β_1	Mean($\hat{\beta}_1$)	0.50	0.50	0.50
		SE($\hat{\beta}_1$)	0.19	0.14	0.11
		Mean($\hat{\sigma}_1$)	0.19	0.14	0.11
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	0.55	0.52	0.52
		SE($\hat{\beta}_2$)	0.44	0.32	0.29
		Mean($\hat{\sigma}_2$)	0.43	0.31	0.28
		Coverage by CI	95 %	95 %	95 %

Table 4.2: Current status data. Time to event follows additive hazards model with constant baseline hazard.

constant baseline hazard, especially concerning the estimate of β_2. For example, standard error of estimates of β_2 with the two-step method and $n = 200$ is 0.98 in this model, compared to 0.70 in the model with constant baseline hazard. This is not because we would generate data with too little failures — as we can see in Table 4.1, the number of overall failures and the number of failures among observations with jump in $Z_2(t)$ are comparable in the two models.

Correlated covariates

In this part, we keep $\lambda_{T,0}(t) = \frac{1}{3}$ and $\lambda_{C,0}(t) = \frac{1}{3}$, but we create correlation between the covariates. Specifically, Z_1 stays with the distribution Unif$(0, 3)$ and

$$Z_2(t) = W \cdot \mathbb{I}(t \geq 0.5),\ \ t \geq 0,\ \ W \sim \text{Bernoulli}(Q),$$
$$Q = 0.1 \cdot \mathbb{I}(Z_1 \in (0, 1\rangle) + 0.5 \cdot \mathbb{I}(Z_1 \in (1, 2\rangle) + 0.9 \cdot \mathbb{I}(Z_1 \in (2, 3)).$$

The results of this simulation for $n = 200$ are in Table 4.4. The estimates of β_1 have similar properties as the ones from datasets with uncorrelated covariates. The estimates of β_2 are more biased and have much greater variability. For example, the mean and standard error of estimates of β_2 with the two-step

Data	Parameter		One-step	γ_0 known	Two-step
Current status, $\lambda_{T,0} = \frac{8}{9}t$, $n = 200$	β_1	Mean($\hat{\beta}_1$)	0.50	0.49	0.51
		SE($\hat{\beta}_1$)	0.49	0.33	0.26
		Mean($\hat{\sigma}_1$)	0.48	0.32	0.25
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	0.70	0.53	0.56
		SE($\hat{\beta}_2$)	1.86	1.10	0.98
		Mean($\hat{\sigma}_2$)	1.58	1.05	0.94
		Coverage by CI	97 %	97 %	98 %
Current status, $\lambda_{T,0} = \frac{8}{9}t$, $n = 400$	β_1	Mean($\hat{\beta}_1$)	0.50	0.49	0.50
		SE($\hat{\beta}_1$)	0.33	0.22	0.18
		Mean($\hat{\sigma}_1$)	0.33	0.22	0.18
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	0.61	0.52	0.54
		SE($\hat{\beta}_2$)	1.04	0.68	0.62
		Mean($\hat{\sigma}_2$)	0.98	0.67	0.61
		Coverage by CI	96 %	96 %	97 %

Table 4.3: Current status data. Time to event follows additive hazards model with baseline hazard from Weibull distribution.

Data	Parameter		One-step	γ_0 known	Two-step
Current status, $\lambda_{T,0} = \frac{1}{3}$, covar. correlated, $n = 200$	β_1	Mean($\hat{\beta}_1$)	0.51	0.50	0.51
		SE($\hat{\beta}_1$)	0.41	0.30	0.25
		Mean($\hat{\sigma}_1$)	0.40	0.29	0.24
		Coverage by CI	95 %	95 %	95 %
	β_2	Mean($\hat{\beta}_2$)	1.02	0.61	0.73
		SE($\hat{\beta}_2$)	3.19	1.65	1.46
		Mean($\hat{\sigma}_2$)	2.14	1.40	1.25
		Coverage by CI	95 %	94 %	94 %

Table 4.4: Current status data. Time to event follows additive hazards model with constant baseline hazard. Covariates are correlated.

method are 0.73 and 1.46. It is possible that the lower number of failures among observations with jump in $Z_2(t)$ in case of these data (Table 4.1) also contributed to these results. In these data with correlated covariates, the number of failures among observations with jump in $Z_2(t)$ was 12.1 on average, whereas in data with uncorrelated covariates it was 19.4 on average.

Additive hazards for monitoring time

The last simulations with current status data were designed to try the method of Lin et al. [1998] in the case of violated assumptions. We generated the monitoring time from a distribution with additive hazard instead of proportional hazard. That is,

$$\lambda_C(t) = \lambda_{C,0}(t) + \gamma_0^\top \mathbf{Z}(t), \;\; t \geq 0.$$

We have tried constant and linear baseline hazard for the monitoring time, $\lambda_{C,0}(t) = \frac{1}{3}$ and $\lambda_{C,0}(t) = \frac{8}{9}t$. The baseline hazard for time to event was set to $\lambda_{T,0}(t) = \frac{1}{3}$. The results of these simulations for $n = 200$ are in the Table 4.5. In all cases, the approach with γ_0 known is the most (and quite heavily)

Data	Parameter		One-step	γ_0 known	Two-step
Current status, add. haz. for monitoring, exp, $n = 200$	β_1	Mean($\hat{\beta}_1$)	0.47	1.16	0.46
		SE($\hat{\beta}_1$)	0.37	0.29	0.20
		Mean($\hat{\sigma}_1$)	0.37	0.25	0.20
		Coverage by CI	95 %	25 %	94 %
	β_2	Mean($\hat{\beta}_2$)	0.39	0.94	0.37
		SE($\hat{\beta}_2$)	1.46	0.94	0.77
		Mean($\hat{\sigma}_2$)	1.20	0.81	0.72
		Coverage by CI	95 %	93 %	94 %
Current status, add. haz. for monitoring, weibull, $n = 200$	β_1	Mean($\hat{\beta}_1$)	0.81	1.27	0.54
		SE($\hat{\beta}_1$)	0.38	0.27	0.18
		Mean($\hat{\sigma}_1$)	0.38	0.24	0.20
		Coverage by CI	89 %	12 %	96 %
	β_2	Mean($\hat{\beta}_2$)	0.67	1.42	0.48
		SE($\hat{\beta}_2$)	1.38	0.94	0.74
		Mean($\hat{\sigma}_2$)	1.27	0.82	0.72
		Coverage by CI	96 %	80 %	96 %

Table 4.5: Current status data, where the assumption of proportional hazards for monitoring time is violated.

biased. The means of estimates of β_1, β_2 with this approach are 1.16, 0.94 and 1.27, 1.42 in the case of constant and linear $\lambda_{C,0}(t)$, respectively. This is reasonable — adding the specific value of γ_0 to the wrong form of $\lambda_C(t)$ makes the assumptions used for computing $\hat{\beta}_j, j = 1,2$ even more wrong. Let us concentrate on the option with two-step approach. The results here are quite peculiar. In the case of constant $\lambda_{C,0}(t)$, the estimate of β_1 is a bit more biased than in the situation with proportional hazard for monitoring, but has lower variability (mean($\hat{\beta}_1$) = 0.46, SE($\hat{\beta}_1$) = 0.20). The estimate of β_2 is much more biased and has a bit greater variability in comparison with proportional hazard for monitoring (mean($\hat{\beta}_2$) = 0.37, SE($\hat{\beta}_2$) = 0.77). In the case of linear $\lambda_{C,0}(t)$, the estimates seem almost more successful than with the data with proportional hazards for monitoring. The estimate of β_1 is slightly more biased with lower variability (mean($\hat{\beta}_1$) = 0.54, SE($\hat{\beta}_1$) = 0.18) and the estimate of β_2 is less biased and has lower variability (mean($\hat{\beta}_2$) = 0.48, SE($\hat{\beta}_2$) = 0.74). We would need many more simulations with different baseline hazards and also different values of β_0, γ_0 to make a better picture about how the method works when the assumption of proportional hazards for monitoring is violated.

Reproduction of a study by Martinussen and Scheike [2002]

As was mentioned in 3.2.3, Martinussen and Scheike [2002] made a small study to demonstrate that their estimator $\hat{\beta}_{MS}$ (unlike the one from Lin et al. [1998],

$\hat{\beta}_{LOY}$) works also when the monitoring time does not follow proportional hazards model. We repeat again that this is true according to the theory — we need this assumption for deriving assymptotic properties of $\hat{\beta}_{LOY}$, while $\hat{\beta}_{MS}$ does not need any restriction of this kind. However, we believe that the estimate $\hat{\beta}_{LOY}$ in this particular study was done incorrectly and that the results are actually not that much worse than the results of $\hat{\beta}_{MS}$. According to their study with the true value of parameter $\beta_0 = 0.5$, mean($\hat{\beta}_{MS}$) = 0.51 and SE($\hat{\beta}_{MS}$) = 0.21, while mean($\hat{\beta}_{LOY}$) = 0.18 (a considerable bias); SE($\hat{\beta}_{LOY}$) is not stated in the article.

Just like in the study from Martinussen and Scheike [2002], we have generated one constant covariate Z from distribution Unif$(0, \sqrt{12})$, the failure time was generated from distribution with hazard function $\lambda_T(t|Z) = 1 + 0.5Z$ (i.e. $\beta_0 = 0.5$) and the monitoring time was generated from distribution with hazard function $\lambda_C(t|Z) = 0.5 + 0.5\sqrt{Z}$ (i.e. the assumption of proportional hazards was violated). This was done in $B = 10000$ repetitions with a sample size of $n = 200$. The data were censored to the form of current status data. We have analyzed these data with two approaches. The first is the approach with γ_0 known, but we use intentionally the wrong assumption that $\gamma_0 = 0$ (i.e. monitoring is independent of covariates). The second approach is the two-step approach. The results can be found in Table 4.6. In the case of approach with the wrong assumption that $\gamma_0 = 0$, mean($\hat{\beta}_{LOY}$) = 0.18 and SE($\hat{\beta}_{LOY}$) = 0.24. In the case of the two-step approach, mean($\hat{\beta}_{LOY}$) = 0.49 and SE($\hat{\beta}_{LOY}$) = 0.23. We deduce from this that Martinussen and Scheike [2002] used the first of these two approaches. In our opinion, there was no reason for that and the two-step approach should have been used instead, because it was stated in the article Lin et al. [1998] that it is more efficient even if we (think that we) know the true value of γ_0.

Data	Parameter		γ_0 "known"	Two-step
Current status, n=200	β_0	Mean($\hat{\beta}_0$)	0.18	0.49
		SE($\hat{\beta}_0$)	0.24	0.23
		Mean($\hat{\sigma}_0$)	0.25	0.23
		Coverage by CI	74 %	95 %

Table 4.6: Current status data. The assumptions on monitoring time are violated. The approach with γ_0 "known" assumes that monitoring time is independent of covariates, which is not true.

4.2.2 Right-censored data

In the end, we would like to describe a small simulation study with right-censored data to show the impact of lighter/heavier censoring on estimation. We generated exactly the same observations of covariates, time to event and monitoring time as in the first section, i.e.

$$\lambda_T(t|\mathbf{Z}) = \frac{1}{3} + 0.5Z_1 + 0.5Z_2(t), \quad t \geq 0,$$
$$\lambda_C(t|\mathbf{Z}) = \frac{1}{3} \cdot e^{Z_1 + Z_2(t)}, \quad t \geq 0,$$

and then applied right-censoring instead of censoring to current status data. The results of estimating with the method by Lin and Ying [1994] for $n = 200$ are in Table 4.7. As expected, the estimates are better than with the current status

Data		β_1	β_2
Right-censored, $\lambda_{T,0} = \frac{1}{3}$, $n = 200$	Mean($\hat{\beta}_j$)	0.51	0.52
	SE($\hat{\beta}_j$)	0.16	0.44
	Mean($\hat{\sigma}_j$)	0.16	0.42
	Coverage by CI	95 %	95 %
Current status, $\lambda_{T,0} = \frac{1}{3}$, $n = 200$	Mean($\hat{\beta}_j$)	0.51	0.57
	SE($\hat{\beta}_j$)	0.24	0.70
	Mean($\hat{\sigma}_j$)	0.24	0.66
	Coverage by CI	95 %	97 %

Table 4.7: The comparison between estimating from right-censored data and current status data. In the case of current status data, the results are from the two-step method.

data. The mean and standard error of estimates are 0.51 and 0.16 for β_1, 0.52 and 0.44 for β_2.

Conclusion

This thesis described how to deal with current status data if the goal is to estimate the effect of (possibly time-varying) covariates on time to event. In the first chapter, we introduced the basics of survival analysis in general and also important results in martingale theory. The second chapter is dedicated to regression in survival analysis in the case of right-censored data, specifically to the Cox proportional hazards model and the additive hazards model.

We began the third chapter with the description of current status data and described the term of external covariates — covariates whose future values do not add any information on the status of the subject in the present. Restricting ourselves to external covariates turns out to be important in many steps of dealing with current status data. The third chapter continues with comparing the works of Finkelstein [1986], Diamond and McDonald [1991] and Huang [1996], who proposed methods for regression on current status data using the proportional hazards model. These methods are quite complicated and their usage is limited to constant covariates.

In the vast majority of the third chapter, we analyzed the method proposed by Lin et al. [1998], which was the main topic of this thesis. We studied the theoretical aspects of this method in detail and supplemented it with proofs (or parts of proofs) which were omitted in the article. This method assumes the additive hazards model for the time to event and the proportional hazards model for the monitoring time. We proved that these assumptions let us define an event whose hazard can be expressed in such way, that it eventually takes the form of proportional hazards model. Because the idea is to use the asymptotic properties and software for the Cox model, it is necessary to specify the regularity assumptions needed for its theory modified for the model with current status data. We extended and specified the proof of martingale property of a process given in the article, which makes it possible to use the theory of the Cox model. This way, we get two estimators: a one-step estimator and an estimator which uses a known parameter for the monitoring time (if it is known).

One can take three approaches to the method proposed by Lin et al. [1998]. The third and the most effective one is the two-step approach. In the first step, it uses all observations to estimate the effect of covariates on the monitoring time. This estimate is then used in the second step to estimate the desired parameter, which describes the effect of covariates on the time to event. We needed additional proofs to derive the asymptotic properties of this two-step approach. The final result is that the two-step estimator is also asymptotically normal and has the lowest variability (in comparison with the one-step estimator and the estimator which uses a known parameter for monitoring).

In the fourth chapter, we present a simulation study which analyzes the estimator (with all three possible approaches) in different situations with current status data. We try the exponential and Weibull distribution for the time to event, a situation with correlated covariates and also a situation where the assumptions on the monitoring time are violated. The results of the study agree with the theory and show that the two-step estimator has the lowest variability.

Bibliography

P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.

I. Diamond and J. McDonald. Analysis of current-status data. In J. Trussell, R. Hankinson, and J. Tilton, editors, *Demographic Applications of Event History Analysis*, pages 231–252. Oxford University Press, Oxford, 1991.

Dianne M. Finkelstein. A proportional hazards model for interval-censored failure time data. *Biometrika*, 42:845–854, 1986.

T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. John Wiley & Sons, New York, 1991.

Camille Hochheimer. *cpsurvsim: Simulating Survival Data from Change-Point Hazard Distributions*, 2019. URL `http://github.com/camillejo/cpsurvsim`.

J. Huang. Efficient estimation for the proportional hazards model with interval censoring. *The Annals of Statistics*, 24(2):540–568, 1996.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, New York, 1980.

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

D.Y. Lin, David Oakes, and Zhilian Ying. Additive hazards regression with current status data. *Biometrika*, 85(2):289–298, 1998.

T. Martinussen and T. H. Scheike. Efficient estimation in additive hazards regression with current status data. *Biometrika*, 89(3):649–658, 2002.

Terry M Therneau. *A Package for Survival Analysis in R*, 2020. URL `https://CRAN.R-project.org/package=survival`. R package version 3.2-7.

Rainer Walke. *Example for a Piecewise Constant Hazard Data Simulation in R,* Max Planck Institute for Demographic Research, Rostock. 2010.

Adéla Zavřelová. Semiparametric additive risk model. Master's thesis, Faculty of Mathematics and Physics, Charles University, 2020. URL `https://is.cuni.cz/webapps/zzp/detail/208842/49593411/`.