Regession models in survival analysis and reliability
I would like to express thanks to my supervisor Doc. Petr Volf, CSc. for valuable advice, recommendations, patience and providing the literature and data.

Also, I would like to thank my parents for their unconditional support during the whole time of my studies.
I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources.

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

In ............. date ............... signature of the author
Název práce: Regresní modely v analýze přežití a spolehlivosti
Autor: Petr Novák
Katedra: Katedra pravděpodobnosti a matematické statistiky
Vedoucí disertační práce: Doc. Petr Volf, CSc., ÚTIA AV ČR
Abstrakt: V předložené práci studujeme metody pro modelování závislosti dat z oblasti analýzy přežití a spolehlivosti na dostupných vysvětlujících proměnných. V první části práce studujeme základní modely analýzy přežití, porovnáváme vlastnosti Coxova modelu proporcionálního rizika, Aalenova aditivního modelu a modelu zrychleného času. Úvádíme metody pro testování dobré shody modelu s daty, založené na teorii čítacích procesů a martingalů, umožňující rozpoznat, který model popisuje data nejlépe. Druhá část se věnuje modelování opravitelných systémů. Studujeme způsoby, jak do modelu zahrnout informace o historii zařízení, včetně vlivu oprav a preventivní údržby. Užití představených metod předvádíme na příkladech z praxe a na simulovaných datech zkoumáme jejich chování v různých situacích.
Klíčová slova: analýza přežití, regresní modely, modely oprav

Title: Regression models in survival analysis and reliability
Author: Petr Novák
Department: Department of Probability and Mathematical Statistics
Supervisor: Doc. Petr Volf, CSc., ÚTIA AV ČR
Abstract: In present work we study methods for modeling the dependence of data from survival and reliability setting on available explanatory variables. The first part of the work compares the properties of the Cox proportional hazards model, Aalen additive model and the Accelerated failure model for survival data. We present methods for testing goodness-of-fit based on counting processes and martingale theory, allowing to identify which model fits the data best. The second part focuses on modeling the lifetime of repairable systems. We study the means of incorporating the history of studied devices into the models, including the influence of corrective repairs and preventive maintenance actions. We demonstrate the introduced methods on real applications and study their properties in various situations on simulated data.
Keywords: survival analysis, regression models, repairs
Contents

List of tables 2
List of figures 3
Introduction 5
1 Survival analysis 8
  1.1 Inference in survival analysis 9
  1.2 Data as counting processes 10
  1.3 Martingale representation 11
  1.4 Hazard function estimator 14
2 Regression models for survival data 15
  2.1 Cox proportional hazards model 16
    2.1.1 Inference 16
    2.1.2 Model checking 18
  2.2 Aalen additive hazards model 22
    2.2.1 Inference 22
    2.2.2 Goodness-of-fit testing procedures 23
  2.3 Accelerated failure time model 26
    2.3.1 Notable cases of time-dependent covariates 26
    2.3.2 Inference 27
  2.4 Goodness-of-fit testing of the AFT model 32
    2.4.1 Graphic test 32
    2.4.2 Classic regression tests 33
    2.4.3 Test based on martingale residuals 35
  2.5 Further models 46
  2.6 Conclusion 47
3 Empirical properties of the goodness-of-fit tests 48
  3.1 Properties of the AFT test based on martingale residuals 48
    3.1.1 Constant covariates 49
    3.1.2 Time-varying covariates 52
  3.2 Comparison of the basic models 56
    3.2.1 Time-invariant covariates 56
    3.2.2 Time-varying covariates 58
    3.2.3 Gamma baseline distribution 60
  3.3 Modeling of the impact of a strike 61
# List of tables

1. Testing the AFT model - weight comparison ........................................... 49
2. Testing the AFT model - sample size comparison .................................. 52
3. Testing the AFT model - data from AFT with one jump ......................... 53
4. Testing the AFT model - data from Cox with one jump .......................... 53
5. Testing the AFT model - AFT with a misspecified jump .......................... 54
6. Testing the basic models - constant covariates ..................................... 57
7. Testing the basic models - one jump ................................................. 59
8. Testing the basic models - constant covariates ..................................... 60
9. Smelting cells - model fitting ......................................................... 62
10. Repair models - baseline hazard testing ............................................. 90
11. Repair data - estimation with one pump ........................................... 92
12. Repair data - estimation with five pumps .......................................... 93
## List of figures

1. Cumulative baseline hazard divided by quartiles - Cox .......................... 19
2. Resampling of the Cox model ................................................................. 21
3. Notable time transformations - AFT model ............................................ 28
4. Cumulative baseline hazard divided by quartiles - AFT .......................... 33
5. Residuals of the AFT model ................................................................. 34
6. Replications of $W(t)$ under the AFT model ....................................... 51
7. Replications of $W(t)$ - AFT with time-varying covariates ..................... 55
8. Comparison of cumulative baseline hazard estimates .............................. 88
Introduction

In present work we study the means of modeling the dependence of survival and reliability data on available additional information. Survival analysis deals with data representing the time until some predefined event - a failure. It has many applications mainly in medical and technical research. It can be used for example for predicting the time from the onset of a certain disease until the death of the patient, or for modeling of the lifetime of a device in a technological circuit. To interpret the nature of the dependence of the time-to-failure distribution on the explanatory data, it is necessary to choose an appropriate model. And to determine whether the selected model fits the data properly, it is advisable to check whether the assumptions of the used model hold. Thus we study and further develop goodness-of-fit testing procedures for survival regression data. We demonstrate the properties of discussed methods on both simulated and real data.

This work is divided into two main parts. The first part (chapters 1-3) deals with the data in classic survival setting where the failure for each subject is a final state. In the second part (chapters 4-5) we work with reliability data in technical setting, where the studied subjects can be repaired and put to use again.

In the first chapter we introduce the basics of survival analysis. We deal with the problem of censoring, which arises when some subjects are removed from observation prior to reaching the failure. We also describe, how the data can be viewed as counting processes, which can then be utilized in statistical inference. The second chapter deals with basic regression models available for the survival data. We study the Cox proportional hazards model, Aalen additive hazards model and the Accelerated failure time model with the estimation procedures for the components of these models. We also present both existing and new diagnostic methods and goodness-of-fit tests for these models, which can be used to see which model fits the studied data best or if the model fits the data at all. In the third chapter we study the properties of presented testing procedures on simulated examples in various settings and compare the results.

The fourth chapter introduces the concept of repairable systems and possible ways to model the lifetime of repairable devices depending on available information. Both past repairs and preventive maintenance actions are incor-
incorporated in the models. We develop inference for these models and establish their large-sample properties. In the fifth chapter we show the applications of the repair models on real data and we also study their properties on simulated examples.

For utilizing the models in practice we developed a package for the R software covering the newly introduced approaches. The documentation is available in the Appendix.

State of the art

Various means for dealing with survival regression data have been suggested and widely used. Perhaps the best known is the proportional hazards model, introduced by Cox (1972). An alternative is presented by the additive hazards model (Aalen, 1980), with both methods focusing on the modeling and interpreting the hazard function. The Accelerated failure time model (AFT), presented also by Cox (1972) and further precised by Buckley and James (1979) is based on modeling the virtual age of the studied subjects. Semiparametric inference for the proportional hazards model was given by Cox (1975). The introduction of the counting process theory for survival data by Andersen and Gill (1982) allowed to develop large sample results based on martingale approach and has many uses in establishing the properties of model estimates. Semiparametric inference for the AFT model with recurrent events and constant covariates was studied by Lin et al. (1998) and for the AFT model with time-varying covariates by Lin and Ying (1995).

The counting process approach was used to develop goodness-of-fit tests for the fully parametric variants of the models (Lin and Spiekerman, 1996), for the semiparametric Cox model (Lin et al., 1993) and the Aalen model (Aalen, 1993). As an addition, we present model checking techniques for the AFT model based on linear regression techniques and introduce a goodness-of-fit test based on martingale residuals (Novák, 2013). In the first three chapters we provide an overview of this methods, give proofs of the new approaches and study the properties of the model checking techniques in simulation experiments.

In the second part of the work we focus on methods of modeling the lifetime of repairable systems, with the subjects returning into observation after a failure. The Cox and AFT models were adjusted for recurrent events by Andersen
and Gill (1982) and Lin et al. (1998) respectively. Percy and Kobbacy (1998) and Percy and Alkali (2005) introduced a parametric generalized proportional intensities model based on the proportional intensities model by Cox (1972a) as an extension the Cox proportional hazards model, including the number of repairs and maintenances as time-varying covariates. We introduce the parametric generalization of the the Accelerated failure time model as well as semiparametric variants of both models, the respective inference procedures and the large sample properties of the estimates.

The estimation and model checking of the Cox and Aalen models for survival data are implemented in the R package timereg (Martinussen and Scheike, 2006). We provide an R package for dealing with the Accelerated failure time model in survival setting and also the implementation of the presented repair models.

The main contribution of this work thus lies in the studying and development of the model checking techniques for survival data and their properties, the introduction of new models for repairable systems and also in the implementation of the presented methods for use in R software.
1. Survival analysis

The scope of survival analysis is studying of non-negative random variables representing the time which passes from the beginning of the observation until some predefined event is reached. In medical setting, we can study the progress of a disease on clinical patients, monitoring the time passing from the onset of a disease until a further stage is reached or the death of the patient occurs. In technical research, we can observe the lifetime of some device under workload until its breakdown, either in real service or laboratory testing. The event at the end of observation is often called a failure for often being a negative occurrence, but the methods can be extended to any area requiring the study of time-to-event data, for example the length of unemployment after losing a job.

Suppose that auxiliary information is given in form of a vector of covariates, which can possibly also be time-varying. In medical studies we may have the basic biological data as height, weight of age of patients available, as well as other characteristics, for example if the patient does smoke, is diagnosed with diabetes or any other relevant conditions. In technical studies we can take into account the material of an examined part or the level of stress (pressure, voltage) applied. The aim is then to estimate the distribution of the time to failure and its dependence on possible explanatory variables.

An important concept present in survival analysis is the presence of censoring, meaning we do not have the complete data of the whole observation for each study subjects. In this work we deal with right censoring, when the subjects may be removed from the study prior to reaching the failure. For example, a patient can leave the hospital before the end of study, some technical part can cease functioning for reasons unrelated to the study or the researcher may close the study before all studied subjects have reached failure. In this case, we only know that the subjects were alive at the time of censoring, but do not know when the failure occurred afterwards.
1.1 Inference in survival analysis

Suppose we have \( n \) subjects in the study. Denote \( T_1^*, ..., T_n^* \) the real failure times and \( C_1, ..., C_n \) the censoring times. We observe the censored events

\[
T_i = \min(T_i^*, C_i) \quad i = 1, ..., n
\]

along with the information whether the event was a failure or a censored event

\[
\Delta_i = I(T_i^* \leq C_i).
\]

We suppose that the censoring mechanism is independent of the distribution of the time to failure. This assumption is however not possible to verify mathematically, as we observe only the minimum of the failure time and censoring time.

Throughout this work, the time-to-failure distribution is considered to be continuous. Denote the failure time distribution functions as \( F_i(t) = P(T_i^* \leq t) \) and the corresponding densities as \( f_i(t) \). In survival analysis we work with the survival function \( S_i(t) = 1 - F_i(t) \) and hazard function

\[
\alpha_i(t) = \lim_{h \to 0^+} \frac{P(t \leq T_i^* \leq t + h | T_i^* \geq t)}{h}
\]

which represents the measure of the probability of immediate failure given that the subject has survived until the time \( t \). It follows that

\[
\alpha_i(t) = \frac{f_i(t)}{S_i(t)}.
\]  

(1)

The cumulative hazard function

\[
A_i(t) = \int_0^t \alpha_i(s) ds
\]

can be interpreted as the cumulative wear of the subject up to time \( t \) and due to (1) can be written as

\[
A_i(t) = - \log S_i(t).
\]

Denote the vector of covariates as

\[
Z_i = (Z_{i1}, ..., Z_{ip})^T
\]

or

\[
Z_i(t) = (Z_{i1}(t), ..., Z_{ip}(t))^T
\]
if they are time-varying. The observed data then has the form of

$$(T_i, \Delta_i, Z_i(t)), \quad i = 1, \ldots, n$$

which we assume to be independent for $i \neq j$.

Most often when dealing with explanatory variables, we want to measure their influence by introducing a model with appropriate regression parameters and a link, such as $Z_i^T \beta$, which is then incorporated in the hazard function. The estimation of the regression parameters can then be done by using maximum likelihood approach. The maximum likelihood for censored survival data is obtained by inserting the density values for the subjects where a failure occurred and the survival function at the censored times (Kalbfleisch and Prentice, 2002, Ch.6):

$$L = \prod_{\Delta_i=1} f_i(T_i) \prod_{\Delta_i=0} S_i(T_i) = \prod_{i=1}^n (f_i(T_i))^\Delta_i (S_i(T_i))^{1-\Delta_i} = \prod_{i=1}^n (\alpha_i(T_i))^\Delta_i S_i(T_i).$$

The log-likelihood then has the form

$$l = \sum_{i=1}^n \left( \Delta_i \log(\alpha_i(T_i)) - A_i(T_i) \right).$$

### 1.2 Data as counting processes

The data can be interpreted as counting processes of failures. Denote

$$N_i(t) = I(T_i \leq t, \Delta_i = 1)$$

the indicator process whether up to time $t$ a noncensored failure occurred on the i-th subject. Let

$$Y_i(t) = I(T_i \geq t)$$

be the at-risk process, indicating whether the i-th subject is still alive at time $t$. The processes are observed on an interval $[0, \tau]$, with $0 < \tau < \infty$ being larger than the highest observed time value but taken as finite to avoid problems at infinity.

The likelihood then can be rewritten as

$$l = \sum_{i=1}^n \int_0^\tau \left( \log(\alpha_i(t)) dN_i(t) - Y_i(t) \alpha_i(t) dt \right).$$

(2)
The stochastic integrals with respect to $dN_i(t)$ equal in fact to the respective integrated functions at the time point $T_i$ if $\Delta_i = 1$ and zero otherwise. This is because $dN_i(t) = 1$ at a time of an uncensored event of the i-th subject and zero elsewhere. At this time it is possible to parametrize the hazard, insert it in the likelihood and compute the score for the parameters as

$$U(\beta) = \frac{d}{d\beta} \sum_{i=1}^{n} \int_{0}^{T} \left( \log(\alpha_i(t))dN_i(t) - Y_i(t)\alpha_i(t)dt \right).$$

### 1.3 Martingale representation

Denote the history of events up to time $t$ as the filtration $\mathcal{F}_t = \sigma\{N_i(s), Y_i(s), Z_i(s), i = 1, ..., n, 0 \leq s \leq t\}$ and the history just prior the time $t$ as $\mathcal{F}_{t-} = \sigma\{N_i(s), Y_i(s), Z_i(s), i = 1, ..., n, 0 \leq s < t\}$.

A random process $X(t)$ is called $\mathcal{F}_t$-adapted, if $X(t)$ is $\mathcal{F}_t$-measurable for each $t$. A process is predictable, if, considered as a mapping from $\Omega \times \mathbb{R}_+$, is measurable with respect to the $\sigma$-algebra generated by all left-continuous adapted processes (Flemming and Harrington, 1991). We work with cadlag processes - right-continuous processes with left limits. An important tool in survival analysis is the martingale representation of the data.

**Definition 1 (Martingale).**
A cadlag process $M(t)$ is called a martingale (submartingale) with respect to the filtration $\mathcal{F}_t$, if

(i) $M(t)$ is $\mathcal{F}_t$-adapted

(ii) $E|M(t)| < \infty$ for all $0 \leq t < \infty$

(iii) $E(M(t)|\mathcal{F}_s) = M(s)$ (or $E(M(t)|\mathcal{F}_s) \geq M(s)$) almost surely $\forall t \geq s \geq 0$.

**Theorem 2 (Doob-Meyer decomposition).**
Let $N(t)$ be a right-continuous nonnegative $\mathcal{F}_t$-submartingale. Then there exists a right-continuous martingale $M(t)$ and a non-decreasing right-continuous predictable process $\Lambda(t)$ called a compensator, for which $E(\Lambda(t)) < \infty$ and

$$N(t) = M(t) + \Lambda(t) \quad \text{almost surely for all} \ t \geq 0.$$
If the distribution of the data is continuous, each of the processes $N_i(t)$ is a right-continuous $\mathcal{F}_t$-submartingale. As such, we can find a compensator $\Lambda_i(t)$ with respect to $\mathcal{F}_t$, so that

$$M_i(t) = N_i(t) - \Lambda_i(t)$$

are zero mean $\mathcal{F}_t$-martingales. If there is a function $\lambda_i(t)$ such that $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$, it is denoted as the intensity of $N_i(t)$ and $\Lambda_i(t)$ as the cumulative intensity. It then can be proven (Flemming and Harrington, 1991), that the intensity has the form

$$\lambda_i(t) = Y_i(t)\alpha_i(t)$$

with $\alpha_i(t)$ being the hazard function for the $i$-th subject.

Similarly as in classic linear regression models $Y_i = X_i^T\beta + \epsilon_i$ with zero mean i.i.d. errors $\epsilon_i$, the data in survival setting can be interpreted as

$$N_i(t) = \Lambda_i(t) + M_i(t)$$

data = model + noise.

When $M(t)$ is a square-integrable martingale, then the process $M^2(t)$ is a square-integrable submartingale. By the Doob-Meyer decomposition, it has a compensator, called the predictable variation process of $M$ and is denoted as $\langle M, M \rangle$ or just as $\langle M \rangle$.

It can be shown (Flemming and Harrington, 1991), that for absolutely continuous survival data with independent right censoring

$$\langle M_i(t) \rangle = \Lambda_i(t) = \int_0^t Y_i(s)\alpha_i(s)ds.$$ 

The martingale representation is useful for establishing the properties of survival models, as many of expressions studied in further chapters have the form of martingale integrals

$$M^{(n)}(t) = \sum_{i=1}^{n} \int_0^t H_i(s)dM_i(s).$$

with various processes $H_i$. Martingales have desirable properties (zero mean and independent increments), and can be therefore used for establishing large
sample theory. Rebolledo (1980) introduced the central limit theorem for martingale integrals. Denote

\[ M^{(n)}(t) = \sum_{i=1}^{n} \int_{0}^{t} H_{i}(s) I_{|H_{i}(s)| \geq \epsilon} dM_{i}(s). \]

**Theorem 3 (Rebolledo, 1980).**

Suppose that for each \( n = 1, 2, \ldots \), there exists a multivariate counting process \((N_{1}(t), \ldots, N_{n}(t))^{T}\), on \( t \in [0, \tau]^{n} \) for some \( \tau < \infty \).

Further suppose that

- \((m-a)\) \( \Lambda_{i}(t) = \int_{0}^{t} \lambda_{i}(s) ds = \int_{0}^{t} Y_{i}(s) \alpha_{i}(s) ds \) are the compensators of processes \( N_{i}(t) \) on \( t \in [0, \tau] \) and \( M_{i}(t) = N_{i}(t) - \int_{0}^{t} Y_{i}(s) \alpha_{i}(s) ds \).
- \((m-b)\) Let \( H_{i}(t) i = 1, \ldots, n, \) be \( p \)-dimensional locally bounded predictable processes on \( t \in [0, \tau]^{p} \).
- \((m-c)\) There exists a positive definite \( p \times p \) matrix \( \Sigma(t) \) of functions continuous on \( [0, \tau] \), such that

\[ \langle M^{(n)}(t) \rangle = \sum_{i=1}^{n} \int_{0}^{t} H_{i}(s) H_{i}(s)^{T} Y_{i}(s) \alpha_{i}(s) ds \overset{P}{\to} \Sigma(t) \]

for each \( t \in [0, \tau] \) as \( n \to \infty \).
- \((m-d)\) (Lindeberg condition) For each \( \epsilon > 0 \) and \( i = 1, \ldots, n \)

\[ \langle M_{i}^{(n)}(t) \rangle = \sum_{i=1}^{n} \int_{0}^{t} H_{i}(s) H_{i}(s)^{T} I_{|H_{i}(s)| \geq \epsilon} Y_{i}(s) \alpha_{i}(s) ds \overset{P}{\to} 0 \]

for each \( t \in [0, \tau] \) as \( n \to \infty \).

Then

\[ M^{(n)}(t) = \sum_{i=1}^{n} \int_{0}^{t} H_{i}(s) dM_{i}(s) \]

converges weakly on \( D[0, \tau]^{p} \) to a zero-mean Gaussian process with a covariance function \( \Sigma(\min(s, t)) \).

**Proof.** See Rebolledo (1980) or Flemming and Harrington (1991). \( \square \)

Condition \((m-b)\) states that the key processes \( H_{i}(t) \) are required to be predictable and locally bounded. A property is said to hold locally for a process \( X(t), t \geq 0 \), if there exists an increasing sequence of stopping times \( \tau_{n} \to \infty \) with respect to \( \mathcal{F}_{t} \), such that the condition holds for the process \( X_{n} = \{ X(\min(t, \tau_{n})) : t \geq 0 \} \).
1.4 Hazard function estimator

Assume that the data are i.i.d. observations \((T_i, \Delta_i) i = 1, ..., n\) with common hazard function \(\alpha(t)\) and no additional information. Consider the counting processes of all failures and subjects at risk

\[
N_\bullet(t) = \sum_{i=1}^n N_i(t), \quad Y_\bullet(t) = \sum_{i=1}^n Y_i(t).
\]

Then \(\Lambda(t) = \int_0^t Y_\bullet(s)\alpha(s)ds\) can be taken as a compensator for \(N_\bullet(t)\), with

\[
M_\bullet(t) = \sum_{i=1}^n M_i(t) = N_\bullet(t) - \int_0^t Y_\bullet(s)\alpha(s)ds
\]

being zero mean \(\mathcal{F}_t\)-martingales. If we take the increment processes \(dM_i(t)\) and \(dN_i(t)\) then

\[
dM_\bullet(t) = dN_\bullet(t) - Y_\bullet(t)dA_0(t)
\]

has zero expectation (using \(A(t) = \int_0^t \alpha(s)ds\)) and we can find the estimator for the cumulative hazard as

\[
\hat{A}(t) = \int_0^t \frac{I(Y_\bullet(s) > 0)}{Y_\bullet(s)}dN_\bullet(s).
\]

This is known as the Nelson-Aalen estimator (Aalen, 1975) and serves as a motivation for the cumulative hazard estimates in more complex models described further.
2. Regression models for survival data

The relation between the time-to-failure distribution and available covariates can be explained with the help of a suitable regression model. There are several models which are used in practice and have been described well in the literature. The most commonly used is the Cox proportional hazards model (Cox, 1972), which assumes that the covariates increase or decrease the hazard function. The Aalen additive hazard model on the other hand suggests that the hazard is composed as a sum of different parts, with each part corresponding to one covariate. The Accelerated failure time model (AFT, Buckley and James, 1979) takes a different approach by supposing that the covariate values either accelerate or decelerate the flow of the internal time of each subject, causing it to age faster or slower.

The Cox model is widely used in medical research due to its straightforward interpretation and is implemented in many software packages. The Aalen model can be seen as its extension, with each covariate adding hazard of a possibly different shape to the total. In this work we focus mostly on the Accelerated failure time model and its generalizations. Its natural interpretation, that additional influences on the subject cause its internal time to run faster or slower and therefore to be more likely for the failure to occur sooner or later than without said influence, makes it easy to explain and can be used in many fields. In this part we introduce the properties of these models and the estimating procedures of their parts and parameters.

When studying experimental data, it is important to choose a model which fits the data well, for an accurate interpretation of the relation between the failure times and the covariates. To see if a selected model describes the data well, we may perform a goodness-of-fit test. Graphical methods have been suggested to give a quick overview of the overall fit for the Cox and AFT models. For the Cox and Aalen model, goodness-of-fit procedures based on martingale residuals and resampling have been developed. For the AFT model, we discuss a method based on classic linear regression and introduce a test based on martingale residuals.
2.1 Cox proportional hazards model

The Cox proportional hazards model (Cox, 1972) assumes that each covariate influences the hazard function multiplicatively. The hazard is assumed to have the form

\[ \alpha_i(t) = \alpha_0(t) \exp(Z_i^T(t)\beta), \quad t \in [0, \tau], \]

(4)

where \( \alpha_0(t) \) is a baseline hazard function and \( \beta = (\beta_1, \ldots, \beta_p)^T \) is a vector of regression coefficients. The measure of influence of the covariates can be interpreted by the coefficients as

\[ \exp(\beta_1) = \frac{\alpha(t, Z_1 + 1)}{\alpha(t, Z_1)} \]

with \( \exp(\beta_1) \) being the multiplicative effect of the hazard when increasing \( Z_1 \) by one while keeping the other covariates fixed. By fitting the Cox model we can determine whether the influence of any given covariate is statistically significant and see the basic trend of the influence. More complex dependence can be modeled by using \( \alpha_i(t) = \alpha_0(t) r(Z_i(t)) \).

2.1.1 Inference

For estimating the model parameters we use the likelihood approach from section 1.1. The baseline hazard function can be either parametrized with its parameters estimated by maximum likelihood approach as well, or as shown here, we can use a semiparametric approach to avoid imposing a shape on the baseline and focusing on the model parameters \( \beta \). Denote

\[ S_0(s, \beta) = \sum Y_i(s)e^{\beta^T Z_i(s)}, \quad S_1(s, \beta) = \sum Z_i(s)Y_i(s)e^{\beta^T Z_i(s)}, \]

\[ S_2(s, \beta) = \sum Z_i(s)(Z_i(s))^TY_i(s)e^{\beta^T Z_i(s)}, \quad E(s, \beta) = \frac{S_1(s, \beta)}{S_0(s, \beta)}. \]

By inserting the hazard function (4) into the log-likelihood (2) we get

\[ l = \sum_{i=1}^n \int_0^\tau \left( (Z_i^T(t)\beta + \log \alpha_0(t))dN_i(t) - Y_i(t)e^{Z_i^T(t)\beta}\alpha_0(t)dt \right). \]

Taking the derivative with respect to \( \beta \) we obtain the score as

\[ U(\beta) = \sum_{i=1}^n \int_0^\tau \left( Z_i^T(t)dN_i(t) - Y_i(t)Z_i^T(t)e^{Z_i^T(t)\beta}dA_0(t) \right). \]
The score depends on the unknown baseline hazard, which can be replaced with a Nelson-Aalen type estimator. We use the martingale decomposition \( N_i(t) = M_i(t) + \Lambda_i(t) \). Then the increment process

\[
dM_i(t) = dN_i(t) - \sum_i Y_i(t) e^{Z_i^T(t)\beta} dA_0(t)
\]

has zero mean and the estimate can be taken as

\[
\hat{A}_0(t, \beta) = \int_0^t \frac{1}{S_0(s, \beta)} dN_i(s).
\]

This is called the Breslow estimate. Inserting into the score we obtain the approximate score in form

\[
\tilde{U}(\beta) = \sum_{i=1}^n \int_0^\tau \left( Z_i^T(t) dN_i(t) - Y_i(t) Z_i^T(t) e^{Z_i^T(t)\beta} \frac{dN_i(t)}{S_0(t, \beta)} \right) = \sum_{i=1}^n \int_0^\tau \left( Z_i^T(t) - \sum_k Y_k(t) Z_k^T(t) e^{Z_k^T(t)\beta} \frac{dN_i(t)}{S_0(t, \beta)} \right) = \sum_{i=1}^n \int_0^\tau \left( Z_i^T(t) - E(t, \beta) \right) dN_i(t)
\]

and the parameter estimates can be obtained by solving

\[
\tilde{U}(\beta) = 0.
\]

Taking the second derivative of the negative of the score with respect to \( \beta \) we get the sample information matrix:

\[
\bar{I}(\beta) = \sum_{i=1}^n \int_0^\tau \left( \frac{S_2(s, \beta)}{S_0(s, \beta)} - \left( \frac{S_1(s, \beta)}{S_0(s, \beta)} \right)^2 \right) dN_i(s) = \int_0^\tau V(s, \beta) dN_i(s).
\]

The asymptotic properties of the estimates can be obtained using regularity conditions

**Theorem 4.**

Let \( \beta_0 \) be the real value of the parameters. Suppose there exists a set \( B \) surrounding \( \beta_0 \) such that:

(a) \( E \left( \sup_{t \in [0,\tau], \beta \in B} Y_i(t) | X_{ij}(t)X_{ik}(t) \exp(Z_i^T(t)\beta) \right) < \infty \forall j, k = 1, ..., p \)

(b) \( P(Y_i(t) = 1 \forall t \in [0, \tau]) > 0 \)

(c) There exists a positive semidefinite matrix \( \Sigma \), such that

\[
n^{-1} \int_0^\tau V(t, \beta_0) S_0(t, \beta_0) dA_0(t) \xrightarrow{P} \Sigma.
\]
Then for \( n \to \infty \):
\[
\begin{align*}
    n^{-1/2} \tilde{U}(\beta_0) & \xrightarrow{D} N(0, \Sigma), \\
    n^{1/2} (\hat{\beta} - \beta_0) & \xrightarrow{D} N(0, \Sigma^{-1}), \\
    n^{-1} \tilde{I}(\hat{\beta}) & \xrightarrow{P} \Sigma, \\
    n^{1/2} (\hat{A}_0(t, \hat{\beta}) - A_0(t)) & \xrightarrow{D} W(t),
\end{align*}
\]
where \( W(t) \) is a zero mean Gaussian process with a finite variance function.

**Proof.** See Andersen and Gill (1982), Martinussen and Scheike (2006), ch.6, p.184.

The consistency and asymptotic normality of the estimates can be used for testing hypotheses \( H_0 : \beta = \beta_0 \) with the score, Wald and likelihood ratio statistics
\[
\begin{align*}
    \tilde{U}(\beta_0)^T \tilde{I}(\beta_0)^{-1} \tilde{U}(\beta_0),
    \quad (\hat{\beta} - \beta_0)^T \tilde{I}(\hat{\beta})(\hat{\beta} - \beta_0),
    \quad -2 \log \left( \frac{\tilde{L}(\beta_0)}{\tilde{L}(\hat{\beta})} \right),
\end{align*}
\]
respectively, all having asymptotic \( \chi^2_p \) distribution under \( H_0 \).

### 2.1.2 Model checking

In some cases it is possible that the Cox model does not fit the data well. For some data the covariates do not influence the hazard multiplicatively or their influence changes through time. Goodness-of-fit testing procedures have been developed to check if the assumptions are met.

**Graphic tests - stratified data**

Assume that the covariates are time-invariant. We stratify the data into \( K \) groups according to the covariate values. Thus in each group, \( \mathbf{Z}_i^T \beta \) are relatively similar. Because \( \alpha_i(t) = e^{\mathbf{Z}_i^T \beta} \alpha_0(t) \), we get \( A_i(t) = e^{\mathbf{Z}_i^T \beta} A_0(t) \) and therefore
\[
\log A_i(t) = \mathbf{Z}_i^T \beta + \log A_0(t).
\]
We compute the Nelson-Aalen estimate of the cumulative baseline hazard in each group separately, not taking the effects of the covariates into account. Then we plot simultaneously 

\[(t, \log A_k(t)), \quad k = 1, ..., K.\]

If the hazards are indeed proportional, the estimates should be approximately parallel. This can give us a basic information of the validity of the model.

On Fig. 1 we see the Nelson-Aalen estimates for data generated from the Cox model with baseline distribution \(\Gamma(20, 1/100)\), \(\beta = 1\) and \(Z_i\) generated from \(N(0, 1)\). The data has been split into four groups along the quartiles of \(Z_i\).

**Martingale residuals**

If the Cox model holds, the processes

\[M_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(Z_i^T(s)\beta_0)\alpha_0(s)ds,\]
are zero-mean \( F_t \) martingales. By inserting \( \hat{\beta} \) and the Breslow estimate we get their estimates

\[
\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(Z_i^T(s)\hat{\beta}) d\hat{A}_0(s) ds =
\]

\[
= N_i(t) - \int_0^t Y_i(s) \exp(Z_i^T(s)\hat{\beta}) \frac{1}{S_0(s, \beta)} d\hat{N}_i(s).
\]

For the increment processes we get

\[
d\hat{M}_i(s) = dN_i(s) \frac{Y_i(s) \exp(Z_i^T(s)\hat{\beta})}{S_0(s, \beta)} d\hat{N}_i(s).
\]

If we multiply the equations by \( Z_i(t) \), integrate up to time \( t \) and sum over all subjects, we obtain:

\[
\sum_{i=1}^n \int_0^t Z_i(s) d\hat{M}_i(s) = \sum_{i=1}^n \int_0^t \left( Z_i(s) - E(s, \hat{\beta}) \right) dN_i(s) = \tilde{U}(\hat{\beta}, t)
\]

which is called the score process, i.e. the score integrated up to the time \( t \) instead the end point \( \tau \), computed at the point \( \hat{\beta} \).

It is possible to replicate the score process by finding a process which has the same asymptotic distribution:

**Theorem 5.**

Suppose that \( E(t, \beta_0) \overset{w}{\to} e(t, \beta_0) \) on \( t \in [0, \tau] \) and the assumptions of the Cox model (Theorem 4) are met. Then the process

\[
n^{-1/2} \tilde{U}(\hat{\beta}, t) = n^{-1/2}(M_G(t) - \tilde{I}(t, \hat{\beta})\tilde{I}^{-1}(\tau, \hat{\beta})M_G(\tau)),
\]

where \( \tilde{I}(t, \beta) \) is the sample information matrix integrated up to \( t \) instead of \( \tau \) and

\[
M_G(t) = \sum_{i=1}^n \int_0^t (Z_i(u) - E(u, \hat{\beta}))d\hat{M}_i(u)G_i,
\]

with \( G_i \overset{i.i.d.}{\sim} N(0, 1) \).

**Proof.** See Lin et al. (1993). \( \Box \)
Figure 2: The score process and its 50 replications under the Cox model

Now we can replicate the score process \( n^{-1/2} \tilde{U}_G(\hat{\beta}, t) \) many times and compare the replications to that of the original process \( n^{-1/2} \tilde{U}(\hat{\beta}, t) \). On Fig.2 we see the score process and its 50 replications. As the test statistics we can use

\[
\sup_{t \in [0, \tau]} | \tilde{U}_j(\hat{\beta}, t) | \quad \text{or} \quad \sup_{t \in [\delta, \tau - \delta]} \left| \frac{\tilde{U}_j(\hat{\beta}, t)}{\sqrt{\text{var} \tilde{U}_j(\hat{\beta}, t)}} \right|, \quad j = 1, ..., p,
\]

with \( \text{var} \tilde{U}_j(\hat{\beta}, t) \) being an estimate of the variance of the score process and a small \( \delta > 0 \) to avoid problems at the endpoints of \([0, \tau]\).

In the case of time-invariant covariates the score process equals to a weighted sum of martingale residuals

\[
n^{-1/2} \tilde{U}(\hat{\beta}, t) = \sum_{i=1}^{n} Z_i \tilde{M}_i(t).
\]

Lin et al. (1993) shows that it is possible to use a relatively general class of weights allowing to compare the results of the test.
2.2 Aalen additive hazards model

The additive hazards model (Aalen, 1980) assumes that each covariate influences an additive part of the hazard function. The hazard is taken as a sum
\[ \alpha_i(t) = Z_i^T(t)\beta(t), \quad t \in [0, \tau], \]
with \( Z_i(t) = (Z_{i1}(t), ..., Z_{ip}(t))^T \) being the covariate values for the i-th subject. While the model is quite broad, the disadvantage is that the parts of the hazard function \( \beta(t) = (\beta_1(t), ..., \beta_p(t))^T \) have to be estimated nonparametrically.

The additive model is useful when modeling a serial system of independent components of a technological circuit, where a failure of a single part means the failure of the whole circuit. Suppose that the survival function of the j-th component is \( S_j(t) \) and the corresponding hazard function \( \beta_j(t) \), for the whole system we get:
\[ S(t) = \prod_{j=1}^{p} S_j(t), \]
\[ \alpha(t) = -\frac{\partial}{\partial t} \log S(t) = \sum_{j=1}^{p} -\frac{\partial}{\partial t} \log S_i(t) = \sum_{j=1}^{p} \beta_j(t). \]

The additive shape of the hazard can be interpreted as a first order Taylor expansion of the hazard with respect to the covariates.
\[ \alpha(t, Z(t)) = \alpha(t, 0) + Z(t)^T \alpha'(t, Z(t)^*). \]

2.2.1 Inference

Suppose \( \int_0^\tau |\beta_j(t)|dt < \infty, \ j = 1, ..., p \). Using the least squares method it is possible to estimate the cumulative parts of the hazard function
\[ B_j(t) = \int_0^t \beta_j(s)ds. \]
Denote \( Z_Y(t) = (Y_1(t)Z_1(t), ..., Y_n(t)Z_n(t))^T \) and its pseudoinverse matrix
\[ Z_{Y}^{-1}(t) = (Z_Y^T(t)Z_Y(t))^{-1}Z_Y^T(t). \]
The vector of estimates is taken as
\[
\hat{B}(t) = \int_0^t \hat{Z}_Y(s)dN(s),
\]
where \(dN(s) = (dN_1(t), ..., dN_n(t))^T\). The asymptotic properties can be summarized in the following theorem:

**Theorem 6.**

Let

(a) \(\sup_{t \in [0, \tau]} E(Y_i(t)Z_{ij}(t)Z_{ik}(t)Z_{il}(t)) < \infty \forall j, k, l = 1, ..., p\)

(b) \(r_2(t) = E(Y_i(t)Z_i^{\otimes 2}(t))\) is regular \(\forall t \in [0, \tau]\)

then for \(n \to \infty\) follows:

\[
n^{1/2}(\hat{B}(t) - B(t)) \xrightarrow{D} W(t),
\]

where \(W(t)\) is a Gaussian martingale with variance function

\[
\Phi(t) = \int_0^t \phi(s)ds,
\]

with

\[
\Phi(t) = r_2^{-1}(t)E(Y_i(t)Z_i^{\otimes 2}(t)Z_i^T(t)\beta(t))r_2^{-1}(t).
\]

Moreover,

\[
\hat{\Phi}(t) = n \int_0^t \hat{Z}_Y(s)\text{diag}(dN(s))(\hat{Z}_Y(s))^T
\]

is a consistent estimate of \(\Phi(t) \forall t\).

**Proof.** See Martinussen and Scheike (2006), chapter 5, p.110. \(\square\)

The pointwise \(1 - \alpha\) confidence interval for each \(B_j(t)\) can be taken as

\[
\hat{B}_j(t) \pm n^{-1/2}u_{1-\alpha/2}(\hat{\Phi}_j^{1/2}(t)),
\]

with \(u_{1-\alpha/2}\) being the corresponding quantile of \(N(0,1)\).

### 2.2.2 Goodness-of-fit testing procedures

The Aalen model is very flexible because of its nonparametric nature. It can however happen that it does not fit the data well, for example if the influence of the covariates is not linear or if there are important interactions between covariates missing.
Martingale residuals

Denote $N(t) = (N_1(t), ..., N_n(t))^T$ and $\hat{M}(t) = (\hat{M}_1(t), ..., \hat{M}_n(t))^T$. The estimated martingale residual processes have the form

$$\hat{M}(t) = N(t) - \int_0^t Z_Y(s)d\hat{B}(s) = \int_0^t I - Z_Y(t)Z_Y(t)J(s).$$

Choose a $n \times m$-dimensional matrix process $K(t) = (K_{i,j}(t))$, $i = 1, ..., n$, $j = 1, ..., m$ and define the K-cumulative residual process as

$$M_K(t) = \int_0^t K^T(s)d\hat{M}(s).$$

The asymptotic distribution of $M_K(t)$ can be again estimated with the help of simulation methods

**Theorem 7.**

Suppose the assumptions of Theorem 6 are met. Then the process $M_K(t)$ converges weakly to a zero-mean Gaussian process with a finite variance. Moreover, the replicated processes

$$M_K^G(t) = \sum_{i=1}^n G_i \int_0^t (K_i(s)^T - K^T(s)Z_Y(s)(Z_Y^T(s)Z_Y(s))^{-1}Z_i(s))d\hat{M}(s),$$

with $G_i$, $i = 1, ..., n$ i.i.d. $N(0,1)$ has the same limiting distribution as $M_K(t)$


The goodness of fit can be verified by comparing the statistics

$$\sup_{t \in [0,\tau]} |M_K_j(t)|, \quad \text{or} \quad \sup_{t \in [\delta,\tau-\delta]} \frac{|M_K_j(t)|}{\sqrt{\text{var}M_K_j(t)}}, \quad j = 1, ..., m.$$

computed from the original and replicated processes, with a small $\delta > 0$ to avoid problems at endpoints of the interval $[0, \tau]$.

It is useful if large values of the test statistics indicate the departure from the model assumptions to some certain alternative, for example to indicate that the dependence of the hazard on k-th covariate is not well described by the
model. For such a test it is possible to take \( K \) time-invariant as a matrix of indicators that the values of the tested covariate belong to one of its quartiles:

\[
K_i^k = (I(Z_{ik} < 1.\text{q}(Z_k)), I(1.\text{q}(Z_k) < Z_{ik} < 2.\text{q}(Z_k)), \\
I(2.\text{q}(Z_k) < Z_{ik} < 3.\text{q}(Z_k)), I(3.\text{q}(Z_k) < Z_{ik})), \ i = 1, \ldots, n.
\]

Then

\[
M_{K_j}^k(t) = \int_0^t \sum_{i=1}^n K_i^k d\tilde{M}(s) = \sum_{i:Z_{ik} \in j.\text{q}(Z_k)} Z_{ik}M_i(t), j = 1, \ldots, 4
\]

is a sum of martingale residuals for the subjects with the values of the tested covariate belonging to a given quartile. If the test statistics \( \sup_{t \in [0,T]} |M_K^k(t)| \) or its variance-standardised version computed from the data exceed the \((1 - \alpha)\)-quantile of the simulated statistics, the dependence on the k-th covariate is not described well by the model.
2.3 Accelerated failure time model

The Accelerated failure time model (AFT, see Buckley and James, 1979) provides an alternative to the most widely used and well described Cox model. In the AFT model, we assume that the covariate values influence the internal time of the device, causing it to age faster or slower. With time-invariant covariates, we assume the log-linear dependence

$$\log T_i^* = -Z_i^T \beta_0 + \epsilon_i,$$

where $\epsilon_i$ are i.i.d. random variables with an unknown distribution. The hazard function has the form

$$\alpha_i(t) = \alpha_0(\exp(Z_i^T \beta_0) t) \exp(Z_i^T \beta_0),$$

where the baseline hazard $\alpha_0(t)$ is the hazard rate of $\exp(\epsilon_i)$.

We can also work with time-dependent covariates $Z_i(t)$. Cox and Oakes (1984) and Lin and Ying (1995) proposed representing the failure times via following time transformation:

$$e^{\epsilon_i} = \int_0^{T_i^*} e^{Z_i^T(s) \beta_0} ds.$$

This can be interpreted as the time transformation from observed to virtual age

$$t \rightarrow \int_0^t e^{Z_i^T(s) \beta_0} ds =: h_i(t, \beta).$$

For constant covariates we get the original model $t \rightarrow t e^{Z_i^T \beta_0}$. It follows that the hazard function has the form

$$\alpha_i(t) = \alpha_0(h_i(t, \beta_0)) \exp(Z_i^T(t) \beta_0).$$

Time-varying covariates may be of many kinds, we introduce some important cases below.

2.3.1 Notable cases of time-dependent covariates

While the transformation $h_i(t, \beta)$ used for time-dependent covariates may not seem intuitive, there are useful applications with straightforward interpretation. If some constant influence is added for each individual at a different
time point, say \( s_i \), we can take:

\[
Z_i(t) = \begin{cases} 
1 & t > s_i \\
0 & t \leq s_i.
\end{cases}
\]

Then \( h_i(t, \beta) = \min(t, s_i) + e^\beta (t - s_i)^+ \), which means that at the time \( s_i \), the individual \( i \) starts to age internally faster or slower by the factor of \( e^\beta \).

If the influence added at time \( s_i \) is different for each individual, we can take:

\[
Z_i(t) = \begin{cases} 
Z_i & t > s_i \\
0 & t \leq s_i.
\end{cases}
\]

Then \( h_i(t, \beta) = \min(t, s_i) + e^{Z_i \beta} (t - s_i)^+ \). At the time \( s_i \), the individual starts to age internally faster or slower by the factor of \( e^{Z_i \beta} \).

If the effect of one covariate increases gradually, it can be modeled as:

\[
Z_i(t) = g(t) Z_i
\]

with \( g(t) \) being a strictly increasing non-negative function, e.g. \( \log(1 + t) \).

On Fig.3 we see the comparison of the observed and virtual age of three subjects in each case of time-dependent covariates.

### 2.3.2 Inference

As with the Cox model, we assume that the baseline hazard function is unknown. To estimate the parameters, we use the counting process approach, using the time transformed versions of counting processes. Denote \( h_i^{-1}(t, \beta) \) the inverse of the function \( h_i(t, \beta) \) with respect to \( t \). This is the time transformation of the real observed data to the virtual i.i.d. data. We observe the processes on the virtual age scale:

\[
N^*_i(t, \beta) = N_i(h_i^{-1}(t, \beta)) = \Delta_i \{ h_i(T_i, \beta) \leq t \}
\]

\[
Y^*_i(t, \beta) = Y_i(h_i^{-1}(t, \beta)), \quad Z^*_i(t, \beta) = Z_i(h_i^{-1}(t, \beta)).
\]

Denote

\[
S^*_0(t, \beta) = \sum_{i=1}^{n} Y^*_i(t, \beta), \quad S^*_1(t, \beta) = \sum_{i=1}^{n} Y^*_i(t, \beta) Z^*_i(t, \beta),
\]
Figure 3: The time transformations from virtual to observed age for three cases of time-dependent covariates. Dashed, dotted and dash-dot curves represent the virtual data, full line is the non-transformed state.
\[ S_2^*(t, \beta) = \sum_{i=1}^{n} Y_i^*(t, \beta)Z_i^*(t, \beta)Z_i^T(t, \beta) \]
\[ E^*(t, \beta) = \frac{S_i^*(t, \beta)}{S_0^*(t, \beta)}. \]

By inserting the hazard function \( h_i(t, \beta) \) into the log-likelihood function \( (2) \), we obtain
\[ l = \sum_{i=1}^{n} \int_{0}^{T} \left( \log(\alpha_0(h_i(t, \beta))) + Z_i^T(t, \beta)dN_i(t) - Y_i(t)\alpha_0(h_i(t, \beta)) \exp(Z_i^T(t, \beta)dt \right). \]

Taking the derivative with respect to \( \beta \) and working on the virtual time scale we get the score
\[ U(\beta) = \sum_{i=1}^{n} \int_{0}^{T} Q_i(t, \beta)(dN_i^*(t, \beta) - Y_i^*(t, \beta)dA_0(t)) \]
with \( Q_i(t, \beta) = Z_i^T(t) + \frac{\alpha'_0(t)}{\alpha_0(t)} \int_{0}^{t} h_i^{-1}(s, \beta)Z_i^T(s)\exp(Z_i^T(s, \beta)ds. \)

Again, we want to replace the cumulative baseline hazard by its estimate. It can be shown that
\[ M_i^*(t, \beta) = N_i^*(t, \beta) - \int_{0}^{t} Y_i^*(s)dA_0(s) \]
are zero mean martingales, thus the increment processes of all data
\[ dM_i^*(t, \beta) = dN_i^*(t, \beta) - Y_i^*(s)dA_0(s) \]
are also zero mean, motivating the Nelson-Aalen type estimator of the cumulative baseline hazard
\[ \hat{A}(t, \beta) = \int_{0}^{t} \frac{dN_i^*(s, \beta)}{S_0^*(s, \beta)}. \]

Inserting the estimate into the score we get the approximate score in the form
\[ \tilde{U}(\beta) = \sum_{i=1}^{n} \int_{0}^{T} \left( Q_i(s, \beta) - \frac{\sum_{k} Q_k(s, \beta)Y_k^*(s, \beta)}{\sum_{k} Y_k^*(s, \beta)} \right)dN_i^*(s, \beta), \]
with \( Q_i(s, \beta) \) as above. It can be shown, that the estimated parameters are consistent when working with an approximate score \( \tilde{U}(\beta) \) obtained by replacing \( Q_i(s, \beta) \) with \( Q_{1i}(s, \beta) = Z_i^*(s, \beta): \)
\[ \tilde{U}(\beta) = \sum_{i=1}^{n} \int_{0}^{T} (Z_i^{T}(s, \beta) - E^*(s, \beta))dN_i^*(s, \beta), \]
which we shall use in further applications. The estimated parameters \( \hat{\beta} \) are taken as those minimizing \( \| \tilde{U}(\beta) \| \), because the score is not continuous in \( \beta. \)
Theorem 8.
Suppose that following regularity conditions hold:

(C1) $\forall i = 1, \ldots, n, k = 1, \ldots, p$: $Z_{ik}(t)$ have uniformly bounded total variation, i.e. $\exists D : Z_{ik}(0) + \int_0^T |dZ_{ik}(s)| \leq D$.

Because of (C1), $Z_{ik}$ can be decomposed into $Z_{ik}(t) = Z_{ik}(0) + Z^+_{ik}(t) - Z^-_{ik}(t)$, where $Z^\pm_{ik}(\cdot)$ are increasing functions with $Z^\pm_{ik}(0) = 0$.

(C2) There exist $\eta_0 > 0$ and $\kappa_0 > 0$, such that $\forall k = 1, \ldots, p$:

$$\sup_{|t-s| + \|\beta_1 - \beta_2\| \leq \eta_0} \frac{1}{n} \sum_{i=1}^n |Z^+_{ik}(h^{-1}(t, \beta_1)) - Z^+_{ik}(h^{-1}(s, \beta_2))| = O_P(n^{-\frac{1}{2} - \eta_0})$$

and for $d_n > 0$, $d_n \to 0$ exists $\epsilon_0$, such that $\forall k = 1, \ldots, p$:

$$\sup_{|t-s| + \|\beta_1 - \beta_2\| \leq d_n} \frac{1}{n} \sum_{i=1}^n |Z^+_{ik}(h^{-1}(t, \beta_1)) - Z^+_{ik}(h^{-1}(s, \beta_2))| = o_P(\max(d_n^{\epsilon_0}, n^{-\epsilon_0}))$$

(C3) The baseline density $f_0$ and its derivative $f_0’$ are bounded, $\int_0^T \left( f_0’(s) \right)^2 f(s) ds$ is finite and $\int_0^T x^\alpha f(x) dx < \infty$ for some $\alpha > 0$.

(C4) The densities $g_i$ of the censoring distribution of $C_i$ are uniformly bounded, i.e. $\sup_{t, \beta} g_i(t) < \infty$.

Suppose there exist positive semidefinite matrices $\Sigma$ and $A$, such that

$$n^{-1} I(\beta) := n^{-1} \int_0^T \left( \frac{S_2(t, \beta)}{S_0(t, \beta)} - E^*(t, \beta)^{\otimes 2} \right) \to \Sigma.$$  

$$n^{-1} \int_0^T (Z_i^T(t, \beta) - E^*(t, \beta)) Q_i(t, \beta) L_i(t, \beta) dF_0(t) \to A$$

where $L_i$ is the survival function of $h_i(C_i, \beta)$. Then for $n \to \infty$:

$$n^{-1/2} \bar{U}(\beta_0) \overset{D}{\to} N(0, \Sigma),$$  

$$n^{1/2} (\hat{\beta} - \beta_0) \overset{D}{\to} N(0, A\Sigma^{-1}A).$$

Proof. See Lin and Ying (1995) \hfill \square

For testing hypotheses $H_0 : \beta = \beta_0$ we can use the score statistics:

$$\bar{U}^T(\beta_0) I^{-1}(\beta_0) \bar{U}(\beta_0)$$

30
which has asymptotically \( \chi^2_p \) distribution. The Wald statistics cannot be used outright, because the matrix \( A \) contains unknown terms \( \alpha' \) and \( \alpha \). There have been however developed methods to obtain the asymptotic variance \( n^{1/2}(\hat{\beta} - \beta_0) \) by resampling (see section 2.4.3).

To see whether the estimates and their interpretation is reliable, we may want to employ goodness-of-fit testing procedures. For the AFT model we introduce new approaches along with proofs and therefore we devote a whole section to them.
2.4 Goodness-of-fit testing of the AFT model

The Accelerated failure time model presents a way to easily describe and interpret data, stating that the covariates influence the distribution of the time to failure by accelerating or slowing the internal aging of the studied subjects. It may be that the model is not accurate in some cases, for example when the influence is not proportional or when the time-varying nature of the influence is not introduced in the model. In this section we explore possible methods for model checking. First, a basic idea on the fit of the model can be obtained from a graphical representation of the data. If the covariates are time-invariant, it is possible to employ classic regression methods, adjusted to accommodate censored data. Finally, we construct a formal test based on martingale residuals.

2.4.1 Graphic test

We can stratify the data into K groups with the individuals having similar covariate values. In each group, we construct the Nelson-Aalen estimate of the cumulative hazard based on the values $\log T_i$, denoted as $\hat{A}_k^{(\text{log})}(\log t)$. Because

$$F_i^{(\text{log})}(\log t) = P(\log T_i^* < \log t) = P(-Z_i^T \beta + \epsilon_i < \log t) = P(\epsilon_i < \log t + Z_i^T \beta) = F_0^{(\text{log})}(\log t + Z_i^T \beta),$$

we get

$$A_i^{(\text{log})}(\log t) = A_0^{(\text{log})}(\log t + Z_i^T \beta).$$

In each group the part $Z_i^T \beta$ should have similar values. Thus if we insert the values

$$\left(\log t, \hat{A}_k^{(\text{log})}(\log t)\right), \quad k = 1, ..., K$$

for each group into one plot. If the AFT models holds, the estimates should be approximately parallel, shifted along the horizontal axis.

On Fig.4 we see the estimates of the cumulative hazard for data generated from the AFT model with baseline distribution $\Gamma(5, 1/100)$, $\beta = 1$ and $Z_i \sim N(0, 1)$. Data has been divided into groups along the quartiles of $Z_i$. 

32
2.4.2 Classic regression tests

In the case of fixed covariates, it is possible to employ methods based on classic linear regression. Consider first the model with no censoring

\[ \log T_i = -Z_i^T \beta + \epsilon_i. \]

When inserting the maximum likelihood estimates \( \hat{\beta} \), we obtain the residuals

\[ r_i = \log T_i + Z_i^T \hat{\beta}. \]

If we assume that the second order moments of \( \epsilon_i \) are finite, the residuals have the same expectation and asymptotic variance. We can divide the data into subsets e.g. by the values of one of the covariates and check the equality of the means of these subsets with a t-test or Wilcoxon test.

If we deal with censored data, the residuals cannot be used outright. It is however possible to estimate the expected real failure times \( T_i^* \) by creating synthetic data points (Buckley and James, 1979):

\[ \log \hat{T}_i^* = \Delta_i \log T_i + (1 - \Delta_i)E(\log T_i^*|Z_i, T_i^* > T_i). \]

The residuals are then estimated by inserting \( \log T_i^* \) and subtracting \(-Z_i^T \beta\):

\[ \hat{r}_i = \Delta_i r_i + (1 - \Delta_i)E(\epsilon|\epsilon > r_i). \]
The mean $E(\epsilon|\epsilon > r_i)$ is then estimated by taking

$$\hat{E}(\epsilon|\epsilon > r_i) = \frac{\sum_{j: r_j > r_i} r_j d\hat{F}_0(r_j)}{1 - \hat{F}_0(r_i)}$$

where $\hat{F}_0$ is the Kaplan-Meier estimator of the baseline distribution function:

$$\hat{F}_0(t) = 1 - \prod_{j: r_j \leq t} \left( \frac{n - j}{n - j + 1} \right)^{\Delta_j}.$$

On Fig. 5 we see the estimated residuals $r_i$ at the ends of the dashed lines. The observed data at the ends of full lines are smaller, indicating that the covariates caused the subject to age faster than they would without their influence. The data was divided into two groups along the values of the covariate. Because the residuals of the two groups do not seem to differ significantly in mean value, one could conclude that the two sample tests would indicate that the AFT model fits the data well.

This method is very straightforward and one can easily get an idea whether the dependence on each covariate is well described by the AFT model. The downside is that the residuals are neither independent nor identically distributed, and therefore the mentioned two-sample tests do not yield exact results. Also it cannot be adapted outright to accommodate time-varying covariates.
2.4.3 Test based on martingale residuals

Goodness-of-fit procedures based on sums of martingale residuals were proposed by Lin and Spiekerman (1996) and Bagdonavičius and Nikulin (2002), p.252 for the AFT model with parametric baseline hazard and by Lin et al. (1993) for the Cox proportional hazards model. Here we propose a similar testing procedure also for the AFT model.

Denote the martingale residuals
\[ M_i^*(t, \beta) = N_i^*(t, \beta) - \int_0^t Y_i^*(s, \beta) dA_0(s) \]
and their empirical counterparts
\[ \hat{M}_i^*(t, \beta) = N_i^*(t, \beta) - \int_0^t Y_i^*(s, \beta) d\hat{A}_0(s, \beta). \]

When the model holds, the martingale residuals should fluctuate around zero, otherwise they would deviate from zero systematically. The proposed test process is
\[ W(t) = n^{-\frac{1}{2}} \sum_{i=1}^n w_i \hat{M}_i^*(t, \hat{\beta}), \]
where \( w_i := f(Z_i(t))I(Z_i(t) \leq z) \) are weights with a bounded function \( f \) and a vector of constants \( z \). There are many possibilities how to choose the weights, most simple choice is to set \( f(Z_i(t)) = Z_i(t) \) or \( f(Z_i(t)) \equiv 1 \) and the elements \( z_k \) of the vector \( z \) as quantiles of corresponding covariates or use no truncation \( (z = \infty) \). One can also try using the test with various weights and compare the results, see Chapter 3, section 3.1.1.

The idea of the test is to measure the distance of the process from zero, which can be done by computing
\[ \sup_{t \in [0, \tau]} |W(t)| \quad \text{or} \quad \sup_{t \in [\delta, \tau-\delta]} \left| \frac{W(t)}{\sqrt{\text{var}W(t)}} \right| \]
with a suitable variance estimator and some small positive number \( \delta \) to avoid possible problems at the edges.

As we show later, under the null hypothesis, the asymptotic distribution of \( W(t) \) is a zero mean Gaussian process with a covariance function which is
difficult to obtain. To assess whether the difference from zero is significant for given data, it is possible to devise a process $\hat{W}(t)$ which has the same limiting distribution under the null hypothesis and is easy to replicate. First we work with time-invariant covariates. Denote

$$S_w(t, \beta) = \sum_i w_i Y_i^*(s, \beta), \quad E_w^*(t, \beta) = \frac{S_w^*(t, \beta)}{S_w^0(t, \beta)}$$

$$f_N(t) = \frac{1}{n} \sum_i \Delta_i w_i f_0(t) t Z_i, \quad f_Y(t) = \frac{1}{n} \sum_i w_i g_0(t) t Z_i,$$

where $f_0(t)$ and $g_0(t)$ are the baseline densities of $e^{\epsilon_i}$ and $T^T_i e^{-Z_i \beta_0}$, respectively. Let $\hat{f}_N$ and $\hat{f}_Y$ be their empirical counterparts with kernel estimates $\hat{f}_0(t)$ and $\hat{g}_0(t)$. Estimates for $T^T_i e^{-Z_i \beta_0}$ for obtaining $\hat{g}_0(t)$ can be obtained by inserting the estimated parameters $\hat{\beta}$. The quantities $\exp(\epsilon_i)$ for estimating $f_0(t)$ can be to some extent approximated by $\exp(\hat{\epsilon}_i)$, with $\hat{\epsilon}_i$ being the modified regression residuals of Buckley and James (1979) used in section 2.4.2. We will use a different approach introduced by Diehl and Stute (1988), which produces a kernel estimate of censored data using the residuals $r_i$ and the Kaplan-Meier estimate by taking

$$\hat{f}_0(t) = \frac{1}{a_n} \sum_{i=1}^n K\left(\frac{t - e^{r_i}}{a_n}\right) d\hat{F}_0(e^{r_i}).$$

For further applications it is sufficient to use the Gaussian kernel with commonly used Silverman’s bandwidth $a_n = 1.06 \hat{\sigma} n^{-1/5}$ (Silverman, 1986).

The score process is taken as the score integrated only up to time $t$ instead of $\tau$. For the case with constant covariates it has the form

$$U(t, \beta_0) = \sum_{i=1}^n \int_0^t Q(s, \beta_0)(Z_i - E^*(s, \beta_0)) dN_i^*(s, \beta_0)$$

with $Q(s, \beta) = \frac{\alpha_0'(s) s}{\alpha(s)} + 1$. It can be shown, that also with other choices of $Q(s, \beta)$, such as $Q_1 \equiv 1$ or $Q_2(s, \beta) = \frac{1}{n} S_0^*(s, \beta)$, the estimated parameters are consistent. With some algebra, we get

$$U(t, \beta_0) = \sum_{i=1}^n \int_0^t Q(s, \beta_0)(Z_i - E^*(s, \beta_0)) dM_i^*(s, \beta_0)$$

Take $G_i, i = 1, \ldots, n$ as i.i.d. standard normals, let

$$U^G_w(t, \beta) = \sum_{i=1}^n \int_0^t Q(s, \beta)(w_i - E_w^*(s, \beta)) d\hat{M}_i^*(s, \beta)G_i,$$
\[ U^G(t, \beta) = \sum_{i=1}^{n} \int_{t_0}^{t} Q(s, \beta)(Z_i - E^*(s, \beta))d\hat{M}_s^*(s, \beta)G_i. \]

Take \( \hat{\beta}^* \) as the solution of the equation
\[ U(\beta) = U^G(\hat{\beta}). \]

It is of note, that \( n^{1/2}(\hat{\beta} - \hat{\beta}^*) \) has the same limiting distribution as \( n^{1/2}(\beta - \beta_0) \) (see Lin et al., 1998), which is useful for approximating the distribution of \( \beta \). Now we have all the components needed to introduce the result:

**Theorem 9.**

We will treat the covariates \( Z_i \) as random variables. Suppose, that:

(i) \( Z_i \) are bounded,
(ii) \((N_i^*, C_i, Z_i)\) are i.i.d.,
(iii) \( Q, E^*, E_w^* \) and \( \frac{1}{n}S_w \) have bounded variation and converge almost surely to continuous functions \( q, e, e_w \) and \( s_w \), respectively,
(iv) \( C_i^* = C_i e^{T_\beta}Z_i \) have a uniformly bounded density and \( A_0(t) \) has a bounded second derivative,
(v) \( f_N(t) = \frac{1}{n} \sum_i \Delta_i w_i f_0(t) t Z_i \) and \( f_Y(t) = \frac{1}{n} \sum_i w_i g_0(t) t Z_i \) have bounded variation and converge almost surely to \( f^*_N(t) \) and \( f^*_Y(t) \), respectively,
(vi) The kernel estimates \( \hat{f}_0 \) and \( \hat{g}_0 \) have a bounded variation and converge in probability, uniformly in \( t \in [0, \tau] \), to \( f_0 \) and \( g_0 \), respectively.

Under the assumptions (i)-(vi) given the observed data \((N_i(t), Y_i(t), Z_i), i = 1, ..., n\), the process \( W(t) \) from above has asymptotically a zero-mean Gaussian distribution with a finite variance function. Moreover, the limiting distribution is the same the limiting distribution of
\[
\hat{W}(t) = \frac{1}{\sqrt{n}} U^G_w(t, \hat{\beta}) - \sqrt{n} \left( \int_{t_0}^{t} \hat{f}_N(s) d\hat{A}_0(s, \hat{\beta}) \right)^T (\hat{\beta} - \hat{\beta}^*) \\
- \frac{1}{\sqrt{n}} \int_{t_0}^{t} S_w(s, \hat{\beta}) d(\hat{A}_0(s, \hat{\beta}) - \hat{A}_0(s, \hat{\beta}^*)). 
\]

We now prove the asymptotic equivalency of \( W(t) \) and \( \hat{W}(t) \). First we work with fixed covariates and then we generalize the theorem also for time-dependent covariates.

Lin et al. (1998) shows, that under (i)-(iv) for \( d_n \to 0 \):
\[
\sup_{\|\beta - \beta_0\| < d_n} \|U(\beta) - U(\beta_0) + nA(\beta - \beta_0)\|/(n^{1/2} + n\|\beta - \beta_0\|) = o_P(1), \quad (10)
\]
\[ \sup_{t \in [0, \tau], \|\beta - \beta_0\| < \delta_n} |n^{\frac{1}{2}}(\hat{A}_0(t, \beta) - \hat{A}_0(t, \beta_0)) - b^T(t)n^{\frac{1}{2}}(\beta - \beta_0)| = o_p(1), \quad (11) \]

where \( A = \int_0^T q(t)E[Y^*_i(t, \beta_0)(Z_1 - c(t))^{(2)}]d(\alpha_0(t)t + b(t) = - \int_0^T e(s)d(\alpha_0(s)t). \)

### Convergence for sums of \( N_i^* \) and \( Y_i^* \)

First, we need to show the asymptotic properties of \( f_N \) and \( f_Y \):

**Lemma 10.**
Conditional on \( Z_i \), under (i)-(vi) for \( d_n \to 0 \):

\[ \sup_{t \in [0, \tau], \|\beta - \beta_0\| < \delta_n} |n^{\frac{1}{2}} \sum w_i(N_i^*(t, \beta) - N_i^*(t, \beta_0)) + f_N^T(t)n^{\frac{1}{2}}(\beta - \beta_0)| = o_p(1), \]

\[ \sup_{t \in [0, \tau], \|\beta - \beta_0\| < \delta_n} |n^{\frac{1}{2}} \sum w_i(Y_i^*(t, \beta) - Y_i^*(t, \beta_0)) - f_Y^T(t)n^{\frac{1}{2}}(\beta - \beta_0)| = o_p(1), \]

with \( f_N \) and \( f_Y \) defined in (5).

**Proof.** We treat \( Z_i \) as fixed values. We have

\[ n^{-\frac{1}{2}} \sum w_i(N_i^*(t, \beta) - N_i^*(t, \beta_0)) \]
\[ = n^{-\frac{1}{2}} \sum w_i \Delta_i [I(T_i^* e Z_i^T \beta \leq t) - I(T_i^* e Z_i^T \beta_0 \leq t)] \]
\[ = n^{-\frac{1}{2}} \sum w_i \Delta_i [I(T_i^* \leq 0) - I(T_i^* \leq te^{-Z_i^T \beta})] \]
\[ = n^{-\frac{1}{2}} \sum w_i \Delta_i [I(te^{-Z_i^T \beta} \leq T_i^* \leq te^{-Z_i^T \beta}) - I(te^{-Z_i^T \beta} \leq T_i^*)]. \]

From Lemma 1 of Lin and Ying (1995) it follows that uniformly in \( t \in [0, \tau] \):

\[ \sup_{\|\beta - \beta_0\| < \delta_n} |n^{-\frac{1}{2}} \sum w_i(N_i^*(t, \beta) - N_i^*(t, \beta_0)) - n^{-\frac{1}{2}} E \sum w_i(N_i^*(t, \beta) - N_i^*(t, \beta_0))| = o_p(1) \]

and analogically for \( Y^* \). Hence, it suffices to compute the expectation of the sum of indicators. For summand \( i \) we have

\[ E[w_i \Delta_i [I(\cdot) - I(\cdot)] = E[w_i \Delta_i E[I(\cdot) - I(\cdot)]|\Delta_i]]. \]

The inner expectation equals to

\[ E[I(e^{-Z_i^T \beta_0} \leq T_i^* \leq e^{-Z_i^T \beta}) - I(e^{-Z_i^T \beta} \leq T_i^* \leq e^{-Z_i^T \beta_0})] \]
\[ = P(t < T_i^* e^{Z_i^T(\beta_0 - \beta)}) - P(t e^{Z_i^T(\beta_0 - \beta)} < T_i^* e^{Z_i^T \beta_0} \leq t). \]

38
Either the first or the second probability is zero, because the cases are mutually exclusive. Assume first, that $t e^{Z_i^T(\beta_0-\beta)} > t$, which is equivalent with $Z_i^T \beta_0 > Z_i^T \beta$. Because $T_i e^{Z_i^T \beta_0}$ are i.i.d. with the distribution function $F_0$, we have

$$P(t < T_i e^{Z_i^T \beta_0} \leq t e^{Z_i^T(\beta_0-\beta)}) = F_0(t e^{Z_i^T(\beta_0-\beta)}) - F_0(t) = f_0(t) t (e^{Z_i^T(\beta_0-\beta)} - 1) + o_P(1) = f_0(t) t Z_i^T (\beta_0 - \beta) + o_P(1).$$

We used the Taylor expansion for $\beta \to \beta_0$ twice. For $Z_i^T \beta_0 < Z_i^T \beta$ we get the same result.

Therefore we get the desired result with an conditional expectation with $\Delta_i$, we have

$$n^{-\frac{1}{2}} \sum w_i(N_i(t, \beta) - N_i^*(t, \beta_0)) = E\left[\left(1 - \frac{1}{n} \sum w_i \Delta_i f_0(t) t Z_i^T (\beta_0 - \beta) n^{\frac{1}{2}}\right)\right] + o_P(1).$$

Because the censoring is independent, due to SLNN we can replace the expectation with the observed quantity:

$$= \left(\frac{1}{n} \sum w_i \Delta_i f_0(t) t Z_i^T (\beta_0 - \beta) n^{\frac{1}{2}}\right) + o_P(1) = -n^{\frac{1}{2}} f_0^T(t)(\beta - \beta_0) + o_P(1).$$

For the sums of $Y_i^*$, we have

$$n^{-\frac{1}{2}} \sum w_i(Y_i^*(t, \beta) - Y_i^*(t, \beta_0))$$

$$= n^{-\frac{1}{2}} \sum w_i[I(T_i \geq t e^{-Z_i^T(\beta_0-\beta)}) - I(t e^{-Z_i^T(\beta_0-\beta)})]$$

$$= n^{-\frac{1}{2}} \sum w_i[I(t > \min(T_i e^{Z_i^T \beta_0}, C_i e^{Z_i^T \beta_0}) \geq t e^{Z_i^T(\beta_0-\beta)})]$$

$$- I(t e^{Z_i^T(\beta_0-\beta)}) \geq \min(T_i e^{Z_i^T \beta_0}, C_i e^{Z_i^T \beta_0}) \geq t)]].$$

We assumed that $C_i^* = C_i e^{Z_i^T \beta_0}$ have a bounded density and therefore the variables $T_i e^{Z_i^T \beta_0} = \min(T_i^*, C_i) e^{Z_i^T \beta_0}$ can be assumed to have a density $g_0$. Computing again the expectation and using the Taylor expansion, we get

$$n^{-\frac{1}{2}} \sum w_i(Y_i^*(t, \beta) - Y_i^*(t, \beta_0)) = (n^{-1} \sum w_i g_0(t) t Z_i^T (\beta - \beta_0) n^{\frac{1}{2}} + o_P(1))$$

$$= n^{\frac{1}{2}} f_0^T(t)(\beta - \beta_0) + o_P(1).$$

□
The convergence of the statistic $W(t)$ and $\hat{W}(t)$

**Proof of Theorem 9.** We show the asymptotic equivalence by proving the convergence of finite-dimensional distributions and tightness, with the help of multivariate functional central limit theorem given by Pollard (1990).

$$W(t) = n^{-\frac{1}{2}} \sum_i w_i \hat{M}^*_i(t, \hat{\beta})$$

$$= n^{-\frac{1}{2}} \sum_i w_i M^*_i(t, \beta_0) + n^{-\frac{1}{2}} \sum_i w_i (\hat{M}^*_i(t, \hat{\beta}) - M^*_i(t, \beta_0))$$

$$= n^{-\frac{1}{2}} \sum_i w_i M^*_i(t, \beta_0) + n^{-\frac{1}{2}} \sum_i w_i (N^*_i(t, \hat{\beta}) - N^*_i(t, \beta_0))$$

$$- n^{-\frac{1}{2}} \sum_i w_i \int_0^t \left( Y^*_i(s, \hat{\beta}) d\hat{A}_0(s, \hat{\beta}) - Y^*_i(s, \beta_0) dA_0(s) \right)$$

Applying (12) of Lemma 10 and adding and subtracting $Y^*_i(s, \hat{\beta}) dA_0(s)$ and $Y^*_i(s, \beta_0) dA_0(s, \hat{\beta})$ we get

$$W(t) = n^{-\frac{1}{2}} \sum_i w_i M^*_i(t, \beta_0) - n^{-\frac{1}{2}} \sum_i w_i \int_0^t Y^*_i(s, \beta_0) dA_0(s)$$

$$- n^{-\frac{1}{2}} \sum_i w_i \int_0^t \left( Y^*_i(s, \hat{\beta}) - Y^*_i(s, \beta_0) \right) dA_0(s) + o_P(1).$$

With the help of (10) and (11) we have

$$n^{\frac{1}{2}} (\hat{A}_0(s, \hat{\beta}) - A_0(s)) = n^{\frac{1}{2}} (\hat{A}_0(s, \beta_0) - A_0(s)) + b^T(t) n^{\frac{1}{2}} (\hat{\beta} - \beta_0) + o_P(1)$$

$$= n^{\frac{1}{2}} \sum_i \int_0^t dM^*_i(s, \beta_0) S_0^*(s, \beta_0) + b^T(t) n^{-\frac{1}{2}} A^{-1} U(\beta_0) + o_P(1).$$

40
We apply (13) on the last term of \(W(t)\) and then (10) for \(n^{\frac{1}{2}}(\hat{\beta} - \beta_0) = n^{-\frac{1}{2}}A^{-1}U(\beta_0) + o_P(1)\):

\[
W(t) = n^{-\frac{1}{2}} \sum_i w_i M_i^*(t, \beta_0) - n^{\frac{1}{2}} \left( f_N(t) + \int_0^t f_Y(s) dA_0(s) \right) (\hat{\beta} - \beta_0)
\]

\[\]
\[
= n^{-\frac{1}{2}} \sum_i \int_0^t S_w(s, \beta_0) M_i^*(s, \beta_0) dW_M(s) - n^{\frac{1}{2}} \int_0^t S_w(s, \beta_0) dM^T(s) A^{-1}U(\beta_0) + o_P(1)
\]

\[\]
\[
= n^{-\frac{1}{2}} \sum_i \int_0^t (w_i - E_w(s, \beta_0)) dM_i^*(s, \beta_0)
\]

\[\]
\[
= n^{-\frac{1}{2}} \left( f_N(t) + \int_0^t f_Y(s) dA_0(s) + \int_0^t \frac{1}{n} S_w(s, \beta_0) db(s) \right) A^{-1}U(\beta_0) + o_P(1).
\]

The limiting process can be found similarly as in Lin et al. (1998). Write

\[
U_M(t) = n^{-\frac{1}{2}} \sum_i M_i^*(t, \beta_0), \quad U_{MZ}(t) = n^{-\frac{1}{2}} \sum_i Z_i M_i^*(t, \beta_0),
\]

\[
U_{MW}(t) = n^{-\frac{1}{2}} \sum_i w_i M_i^*(t, \beta_0).
\]

For fixed \(t\), each of the processes is a sum of i.i.d. zero-mean terms and therefore the finite-dimensional convergence of \((U_M, U_{MZ}, U_{MW})\) follows from multivariate central limit theorem. For each \(t\), \(M_i^*(t, \beta_0)\), \(Z_i M_i^*(t, \beta_0)\) and \(w_i M_i^*(t, \beta_0)\) can be written as sums and products of monotone functions, and therefore are manageable in sense of Pollard (1990), p. 38. It then follows from the functional central limit theorem (Pollard, 1990, p. 53) that \((U_M, U_{MZ}, U_{MW})\) is tight and converges weakly to a zero-mean Gaussian process, say \((W_M, W_{MZ}, W_{MW})\). By the Skorokhod-Dudley-Wichura theorem (Shorack and Wellner, 1986, p. 47), an equivalent process \((\tilde{U}_M, \tilde{U}_{MZ}, \tilde{U}_{MW})\) in an alternative probability space can be found, in which the convergence becomes almost sure. Because \(Q(t, \beta_0), E^*(t, \beta_0), E_w^*(t, \beta_0), \frac{1}{n} S_w(t, \beta_0), f_N(t)\) and \(f_Y(t)\) have bounded variation and converge almost surely to \(q, e, e_w, s_w, f_0^N(t)\) and \(f_0^Y(t)\), respectively, then \(W(t)\) converges in \(D[0, \tau]\) to

\[
\int_0^\tau dW_{MW}(s) - \int_0^\tau e_w(s, \beta_0) dW_M(s) - c^T(t) \left( \int_0^\tau q(s) dW_{MZ} - \int_0^\tau q(s) e(s, \beta_0) dW_M \right),
\]

where \(c(t) = f_0^N(t) + \int_0^t f_0^Y(s) dA_0(s) + \int_0^t s_w(s, \beta_0) db(s),\) which has zero mean and covariance function \(\sigma(t_1, t_2) = E(\xi(t_1)\xi(t_2)),\) where

\[
\xi(t) = \int_0^t (w_1 - e_w(s, \beta_0)) dM_i^*(s, \beta_0) - c^T(t) A^{-1} \int_0^\tau q(s) [Z_1 - e(s, \beta_0)] dM_i^*(s, \beta_0).
\]
For \( \hat{W}(t) \), we have

\[
\hat{W}(t) = n^{-\frac{1}{2}}U_w(t, \hat{\beta}) - n^{\frac{1}{2}} \left( \bar{f}_N(t) + \int_0^t \bar{f}_Y(s) d\hat{A}_0(s, \hat{\beta}) \right)^T (\hat{\beta} - \beta^*)
\]

\[-n^{-\frac{1}{2}} \int_0^t S_w(s, \hat{\beta}) d(\hat{A}_0(s, \hat{\beta}) - \hat{A}_0(s, \hat{\beta}^*))
\]

\[-n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t (w_i - E_w(s, \hat{\beta})) d\hat{M}_i(s, \hat{\beta}) G_i
\]

\[-n^{-\frac{1}{2}} \left( \bar{f}_N(t) + \int_0^t \bar{f}_Y(s) d\hat{A}_0(s, \hat{\beta}) \right)^T (\hat{\beta} - \beta^*)
\]

\[-n^{-\frac{1}{2}} \int_0^t \frac{1}{n} S_w(s, \hat{\beta}) \hat{b}(s)(\hat{\beta} - \beta^*) + o_P(1)
\]

\[-n^{-\frac{1}{2}} \sum_{i=1}^n \int_0^t (w_i - E_w(s, \hat{\beta})) d\hat{M}_i(s, \hat{\beta}) G_i
\]

\[-n^{-\frac{1}{2}} \left( \bar{f}_N(t) + \int_0^t \bar{f}_Y(s) d\hat{A}_0(s, \hat{\beta}) + \int_0^t \frac{1}{n} S_w(s, \hat{\beta}) \hat{b}(s) \right)^T A^{-1} U(\hat{\beta}^*) + o_P(1).
\]

We used (10) for

\[n^{\frac{1}{2}}(\hat{\beta} - \beta^*) = n^{-\frac{1}{2}} A^{-1} U(\hat{\beta}^*) + o_P(1)
\]

and (11) for

\[n^{\frac{1}{2}}(\hat{A}_0(t, \hat{\beta}) - \hat{A}_0(t, \hat{\beta}^*)) = \hat{b}^T(t) n^{\frac{1}{2}}(\hat{\beta} - \beta^*) + o_P(1).
\]

The score process satisfies \( U(\hat{\beta}^*) = U^G(\hat{\beta}) \) and therefore we see that \( \hat{W}(t) \) consists of the same parts as \( W(t) \), with \( \beta_0, M^*_i(t, \beta_0), f_N(t) \) and \( f_Y(t) \) replaced with \( \hat{\beta}, G_i M^*_i(t, \hat{\beta}), \bar{f}_N(t) \) and \( \bar{f}_Y(t) \). The resampled martingale residuals \( G_i M^*_i(t, \hat{\beta}) \) have the same distribution as their theoretical counterparts, and the kernel estimates of \( f_0 \) and \( g_0 \) converge uniformly to the real densities. Therefore \( \hat{W}(t) \) has the same limiting finite-dimensional distributions as \( W(t) \). Tightness follows also by the same arguments as for \( W(t) \).

\[\square\]

The test statistic - time-varying covariates

We can also work with time-dependent covariates \( Z_i(t) \). Constructing the test is not entirely similar, because the weights \( w_i = f(Z_i(t)) I(Z_i(t) \leq z) \) would be time-dependent. For practical reasons, we work here with time-invariant weights evaluated at a chosen \( t_0 \). It could however also be shown that it is possible to use time-varying weights as long as they are predictable.
With appropriate weights and transformed counting processes we compute $S_{w}^{*}$, $E_{w}^{G}$, and $U^{G}$ in the same way as above. Because $e^{Z_{i}(t)\beta}$ is positive, $h_{i}(t, \beta)$ is increasing in $t$. Therefore for each fixed $\beta$ an inverse function $h_{i}^{-1}(t, \beta)$ can be found, for which $h_{i}^{-1}(h_{i}(t, \beta), \beta) = h_{i}(h_{i}^{-1}(t, \beta), \beta) = t$. Let again $f_{0}$ and $g_{0}$ be the density functions of $h_{i}(T^{*}_{i}, \beta_{0})$ and $h_{i}(T_{i}, \beta_{0})$, respectively. Denote

$$f^{\pm}_{N}(t) = \frac{1}{n} \sum_{i} \Delta_{i} w_{i} f_{0}(t) \frac{\partial}{\partial \beta} \left( - h_{i}(h_{i}^{-1}(t, \beta_{0}), \beta_{0}) \right)_{\beta=\beta_{0}}, \quad (14)$$

$$f^{\pm}_{V}(t) = \frac{1}{n} \sum_{i} w_{i} g_{0}(t) \frac{\partial}{\partial \beta} \left( - h_{i}(h_{i}^{-1}(t, \beta_{0}), \beta_{0}) \right)_{\beta=\beta_{0}} \quad (15)$$

and their empirical counterparts $\hat{f}^{\pm}_{N}(t)$ and $\hat{f}^{\pm}_{V}(t)$ obtained by inserting $\hat{\beta}$ and kernel estimates $\hat{f}_{0}$ and $\hat{g}_{0}$. Again, $h_{i}(T_{i}, \beta_{0})$ can be simply estimated by inserting $\hat{\beta}$.

For $\hat{f}_{0}$ we use the kernel estimator for censored data by Diehl and Stute (1988) utilizing the Kaplan-Meier estimate of the baseline distribution based on $h_{i}(T_{i}, \beta)$.

**Theorem 11.**
Suppose that (i)-(vi) rewritten for the modified variables and processes and also the assumptions (C1)-(C3) for Theorem 8 for $Z_{i}(t)$ hold. Suppose that for fixed $\beta$ the image of $h_{i}(t, \beta), t \in [0, \infty]$ does not depend on $\beta$. Let

$$w_{i} = f(Z_{i}(t_{0})) I(Z_{i}(t_{0}) \leq z)$$

for a fixed time-point $t_{0}$.

Then given the data $(N_{i}(t), Y_{i}(t), Z_{i}(t)), i = 1, ..., n$, the test process

$$W(t) = \frac{1}{\sqrt{n}} \sum_{i} w_{i} \hat{M}_{i}^{*}(t)$$

converges with $n \to \infty$ to the same zero-mean Gaussian process as the resampled process $\hat{W}(t)$ constructed in the same way as in Theorem 8 with modified components $\hat{f}^{\pm}_{N}(t)$ and $\hat{f}^{\pm}_{V}(t)$ from above.

The results (10) and (11) (due to Lin et al., 1998) hold for the case of time-dependent covariates, too. We can also extend (12) and (13):

**Lemma 12.**
Suppose that for fixed $\beta$ the image of $h_{i}(t, \beta), t \in [0, \infty]$ does not depend on $\beta$. Conditional on $Z_{i}(t)$, under the assumptions (i)-(vi) rewritten for the modified variables and processes and (C1)-(C3), for $d_{n} \to 0$:

$$\sup_{t \in [0, \tau], \|\beta-\beta_{0}\|<d_{n}} \left| n^{-\frac{1}{4}} \sum_{i} w_{i}(N_{i}^{*}(t, \beta)-N_{i}^{*}(t, \beta_{0}))+ f_{N}^{\pm}(t) \right| = o_{P}(1),$$

43
Because conditional expectation is sum of indicators (Lin and Ying, 1995, Lemma 1). For each part, the inner □

\[ P \sum_{i} (Y_i^*(t, \beta) - Y_i^*(t, \beta_0) - (f_{i}^+(t))^T n^{\frac{1}{2}}(\beta - \beta_0)) = o_p(1). \]

Proof. We proceed similarly as in the proof of Lemma \[10\]

\[ n^{-\frac{1}{2}} \sum_{i} w_i (N_i^*(t, \beta) - N_i^*(t, \beta_0)) \]

\[ = n^{-\frac{1}{2}} \sum_{i} w_i \Delta_i [I(h_i(T_i^*, \beta) \leq t) - I(h_i(T_i^*, \beta_0) \leq t)] \]

\[ = n^{-\frac{1}{2}} \sum_{i} w_i \Delta_i [I(T_i^* \leq h_i^{-1}(t, \beta)) - I(T_i^* \leq h_i^{-1}(t, \beta_0))] \]

\[ = n^{-\frac{1}{2}} \sum_{i} w_i \Delta_i [I(h_i^{-1}(t, \beta_0) < T_i^* \leq h_i^{-1}(t, \beta)) - I(h_i^{-1}(t, \beta) < T_i^* \leq h_i^{-1}(t, \beta_0))]. \]

Again, it can be shown that it suffices to compute the expectation of the sum of indicators (Lin and Ying, 1995, Lemma 1). For each part, the inner conditional expectation is

\[ E[I(h_i^{-1}(t, \beta_0) < T_i^* \leq h_i^{-1}(t, \beta)) - I(h_i^{-1}(t, \beta) < T_i^* \leq h_i^{-1}(t, \beta_0)) \]

\[ = P(t < h_i(T_i^*, \beta_0) \leq h_i(h_i^{-1}(t, \beta), \beta_0)) - P(h_i(h_i^{-1}(t, \beta), \beta_0) < h_i(T_i^*, \beta_0) \leq t). \]

Both cases are mutually exclusive, suppose first that \( h_i(h_i^{-1}(t, \beta), \beta_0) > t \). Because \( h_i(T_i^*, \beta_0) = e^t \) are i.i.d., we have

\[ P(t < h_i(T_i^*, \beta_0) \leq h_i(h_i^{-1}(t, \beta), \beta_0)) = F_0(h_i(h_i^{-1}(t, \beta), \beta_0)) - F_0(t) \]

\[ = f_0(t) \left( \frac{\partial}{\partial \beta} \left( h_i(h_i^{-1}(t, \beta), \beta_0) \right) \right)_{\beta = \beta_0} (\beta - \beta_0) + o_p(1). \]

using Taylor expansion for \( \beta \to \beta_0 \). For \( h_i(h_i^{-1}(t, \beta), \beta_0) < t \) we get again the same result. Inserting into the sum and replacing the expectation with respect to \( \Delta_i \) with the observed quantity we get

\[ n^{-\frac{1}{2}} \sum_{i} w_i (N_i^{*+}(t, \beta) - N_i^{*+}(t, \beta_0)) \]

\[ = \left( \frac{1}{n} \sum_{i} w_i \Delta_i f_0(t) \frac{\partial}{\partial \beta} \left( h_i(h_i^{-1}(t, \beta), \beta_0) \right) \right)_{\beta = \beta_0} (\beta - \beta_0) n^{\frac{1}{2}} + o_p(1) \]

\[ = -n^{\frac{1}{2}} (f_{i}^+(t))^T (\beta - \beta_0) + o_p(1). \]

In similar way we obtain also the result for sums of \( Y_i^{*+} \).

Proof of Theorem \[11\]. The proof is analogous to the proof of Theorem \[9\] using Lemma \[12\]
Remark 13.
One of the simplest cases would be, if a covariate represents an additional influence which is added in given time $s_i$ for each observed individual,

$$Z_i(t) = \begin{cases} 
1 & t > s_i \\
0 & t \leq s_i.
\end{cases}$$

This means that at the time $s_i$ the observed individual starts to age faster or slower. The covariate processes clearly have bounded variation. We get

$$h_i(t, \beta) = \min(t, s_i) + e^{\beta(t-s_i)^+}, \quad h_i^{-1}(t, \beta) = \min(t, s_i) + e^{-\beta(t-s_i)^+}.$$

The weights for $W(t)$ can be chosen as $w_i = I(s_i \leq z)$ for some $z$, i.e. $z = \text{median}(s_i)$ etc. Or we can simply sum all the residuals ($w_i \equiv 1$). In this case, we have

$$\frac{\partial}{\partial \beta} \left( -h_i(h_i^{-1}(t, \beta), \beta_0) \right)_{\beta=\beta_0} = -\frac{\partial}{\partial \beta} \left( \min(t, s_i) + e^{\beta_0 - \beta(t-s_i)^+} \right)_{\beta=\beta_0} = (t-s_i)^+,$$

therefore $f_N^+(t)$ and $f_Y^+(t)$ are easy to compute.

Also the model with constant covariates can be viewed as a special case, with

$$h_i(t, \beta) = te^{Z_i^T \beta}, \quad h_i^{-1}(t, \beta) = te^{-Z_i^T \beta},$$

therefore

$$\frac{\partial}{\partial \beta} \left( -h_i(h_i^{-1}(t, \beta), \beta_0) \right)_{\beta=\beta_0} = -\frac{\partial}{\partial \beta} \left( te^{Z_i^T(\beta_0 - \beta)} \right)_{\beta=\beta_0} = tZ_i^T,$$

and inserting into (14) and (15) we get $f_N^+(t)$ and $f_Y^+(t)$ as in (9).

Testing procedure

We can now compute $W(t)$ for the studied data set and replicate $\hat{W}(t)$ many times. The desired p-value $p$ of the test is the proportion of cases, in which the statistics computed from the replicated $\hat{W}(t)$ exceed the statistic computed from $W(t)$. If $p < \alpha$, we reject the hypothesis that the data follow the AFT model. The variance for the standardised version can be computed directly from the resampled processes.

It is also possible to divide the interval $[0, \tau]$ into $k$ subintervals, i.e. quartiles, and compute the statistics in each of the parts separately and obtain $k$ p-values $p_1, ..., p_k$. One possibility is then to reject the hypothesis whenever we
would reject in one of the subintervals (if \( \min(p_1, \ldots, p_k) < \alpha \)). This can lead to violating the general level of significance of the test. A possibility to avoid this is to use a Bonferroni approximation for multiple-testing and reject only if \( \min(p_1, \ldots, p_k) < \alpha/k \).

2.5 Further models

The three discussed models present basic tools for handling survival data, as they are easy to use and interpret. However, for some data the dependence between the lifetime distribution and available covariates may not be adequately explained by these models. Here we present some of the various explored generalizations.

Cox-Aalen model

The Cox-Aalen model introduced by Scheike and Zhang (2002) assumes that the hazard has the form

\[
\alpha_i(t) = e^{Z^T_i(t)\beta} X^T_i(t) \alpha_0(t)
\]

with a vector of hazard functions \( \alpha_0(t) = (\alpha_1(t), \ldots, \alpha_q(t))^T \). This models contains both an additive and a multiplicative part, allowing for a broad application. A goodness-of-fit procedure based on martingale residuals has been developed for both parts. This allows to determine the nature of the influence of each covariate, whether it fits better to the additive or to the multiplicative part. When dealing with a large number of covariates with possible interactions, however, the interpretation of the model can become less straightforward.

Cox-AFT model

An interesting combination of the Cox proportional hazards model and the AFT model for time-invariant covariates is the approach suggested by Chen and Jewell (2001). The hazard is assumed to have the form

\[
\alpha_i(t) = \alpha_0(t) e^{Z^T_i \beta_1} e^{Z^T_i \beta_2}.
\]
It is possible to see whether a covariate in question has a multiplicative influence by testing $\beta_{1j} = 0$, or whether the influence is time-accelerating with $\beta_{1j} = \beta_{2j}$. When the baseline distribution is close to Weibull, the models coincide. This leads to both conditions being true, resulting in the covariate having no influence at all, as $\beta_{1j} = \beta_{2j} = 0$.

2.6 Conclusion

In this chapter we presented the most widely used regression models for survival data: the Cox, Aalen and AFT model. For each model we have shown the inference procedures for estimating the respective parts and testing their significance. We also presented possible goodness-of-fit methods for checking whether the model explains the studied data well. New approaches for testing the AFT model were introduced, based on both classic linear regression methods and martingale theory. In the next chapter we study the properties of the testing methods in various situations.
3. Empirical properties of the goodness-of-fit tests

In this chapter we will study the empirical properties of the testing procedures introduced in the previous part. As the methods are asymptotic, it is interesting to see how well they can perform in various situations. We therefore simulated artificial data sets from all three presented basic models in different settings and compared the results of different testing procedures. Statistical software R 3.1.2 was used. The inference and testing procedures for the Cox and Aalen models are implemented in the package \texttt{timereg} (see Martinussen and Scheike, 2006). The procedures for the AFT model were newly programmed and are documented in the Appendix. We then conclude this chapter with a real life application from smelting industry.

3.1 Properties of the AFT test based on martingale residuals

First we focus on the newly proposed test of the AFT model based on martingale residuals from section 2.4.3. We want to study whether the test holds its level of significance and observe the empirical power of the test against certain alternatives for various sample sizes. Each time we consider both censored data and data with about one quarter of the observations randomly and independently censored. As the test statistic, $\sup \left| W(t) \right|$ and $\sup \left| \frac{W(t)}{\sqrt{\text{var} W(t)}} \right|$ were taken, with the variance estimated from the resampled processes. Both statistics were computed over the whole time interval and over four separated subintervals divided by the quartiles of $T_i e^{Z_i \beta}$ or $h_i(T_i, \beta)$ respectively. The lowest and highest 5% of the data (on the transformed scale) were omitted to avoid possible problems at endpoints.

The p-value is taken as the proportion of samples in which the replicated statistics exceed the observed one. For the supremum over the whole interval, we reject the hypothesis if the p-value is lower than the chosen level of significance $\alpha = 5\%$. If we compute the p-values over the quartiles separately, we reject the hypothesis when $\min(p_1, p_2, p_3, p_4) < \alpha/4$, using the Bonferroni
correction. Each time, 500 samples were generated and for each sample, \( \hat{W}(t) \) was generated 200×. To examine the empirical power, we generate data from different models and observe the proportion of rightfully rejected samples. To see if the tests hold the significance level, we generate from the AFT model itself and observe the proportion of wrongfully rejected samples.

3.1.1 Constant covariates

| Test | Statistic | \(|\hat{\delta}, \tau - \delta|\) \(\sup |W(t)|\) \(\sup \frac{|W(t)|}{\sqrt{\text{var}(W(t))}}\) Quartiles - Bonferroni \(\sup |W(t)|\) \(\sup \frac{|W(t)|}{\sqrt{\text{var}(W(t))}}\) |
|------|-----------|----------------|----------------|----------------|----------------|----------------|
| Censoring | NC | C | NC | C | NC | C | NC | C |
| \(Z_i\) median | 0.046 | 0.012 | 0.022 | 0.002 | 0.024 | 0.012 | 0.026 | 0.004 |
| \(Z_i\) \(\infty\) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 median | 0.016 | 0.014 | 0.014 | 0 | 0.012 | 0.012 | 0.014 | 0.010 |
| 1 \(\infty\) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The proportion of wrongfully rejected samples from the AFT model:

| \(Z_i\) median | 0.522 | 0.346 | 0.690 | 0.518 | 0.590 | 0.396 | 0.622 | 0.426 |
| \(Z_i\) \(\infty\) | 0.004 | 0.004 | 0.622 | 0.448 | 0.174 | 0.088 | 0.378 | 0.276 |
| 1 median | 0.678 | 0.470 | 0.840 | 0.688 | 0.746 | 0.544 | 0.792 | 0.608 |
| 1 \(\infty\) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The proportion of rightfully rejected samples from the Cox model:

| \(Z_i\) median | 0.664 | 0.462 | 0.674 | 0.466 | 0.608 | 0.408 | 0.618 | 0.422 |
| \(Z_i\) \(\infty\) | 0.004 | 0.004 | 0.114 | 0.050 | 0.004 | 0.002 | 0.044 | 0.012 |
| 1 median | 0.502 | 0.340 | 0.586 | 0.432 | 0.472 | 0.300 | 0.494 | 0.330 |
| 1 \(\infty\) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Table 1: The empirical level of significance of the test and empirical power against the alternatives of the Cox model and the Aalen model for various weights \(w_i\). C - censoring, NC - without censoring

First, we generated data from the AFT model itself, with gamma baseline hazard \(\Gamma(p = 20, \lambda = 4/100)\), \(\beta = 1\) and one covariate \(Z_i\) generated as i.i.d. from \(N(3,1)\). Censoring was generated independently in the same way with baseline \(\Gamma(p = 20, \lambda = 4/150)\). We use weights \(f(Z_i)I(Z_i \leq z)\) with \(f(Z_i)\) as either \(Z_i\) or equal to 1 and \(z\) as either the sample median of \(Z_i\) or infinity (no observations left out) and compare the results. Each time, 500 samples of 1000 observations were tested.
Next, we generated data from the Cox model \( \alpha_i(t) = e^{Z_i \beta_0(t)} \) with the same baseline hazard, parameter and covariates as above and censoring in the same way. To see which weights yield the highest power against this alternative, we compare the results for the four types of weights used above. Then, we generated data from the Aalen model \( \alpha_i(t) = \beta_0(t) + Z_i \beta_1(t) \) with \( Z_i \sim N(3, 1) \), baseline hazard \( \beta_0(t) \) corresponding to \( \Gamma(p = 20, \lambda = 4/100) \) and the added hazard \( \beta_1(t) \) corresponding to \( \Gamma(p = 20, \lambda = 80/150) \) for failure times and \( \Gamma(p = 20, \lambda = 80/150) \) for censoring times.

On Tab.1 we can see that the empirical level of significance tends to be very low, in some cases even with no rejected samples whatsoever. This indicates that the test is overly conservative, leading possibly to a loss of power. In further examples, we will try to overcome this problem by removing the Bonferroni correction, thus rejecting the hypothesis whenever we would reject in one of the quartiles. One has to be careful, because the removal of the correction may lead to exceeding the level of significance, as will be seen further. We can see that the weights \( w_i = I(Z_i \leq \text{median}(Z_j)) \) yield the best empirical power against the alternative of the Cox model. Against the Aalen model, the weights \( w_i = Z_i I(Z_i \leq \text{median}(Z_j)) \) provide largest power.

We now use these weights \( (I(Z_i \leq \text{median}(Z_j)) \) for AFT and Cox, \( Z_i I(Z_i \leq \text{median}(Z_j)) \) for Aalen) for testing samples of size ranging from 50 to 2000 (Tab.2). The results indicate that with increasing sample size the empirical power gets higher, however, for a reasonable power a large number of observations is still needed. Standardising with the deviation process and dividing into quartiles adds some power. With censoring, the power diminishes somewhat. If we do not use the Bonferroni correction for the division in quartiles, the empirical level of significance is sometimes above 0.05, and therefore is not advisable to disregard, especially for larger samples.

On Fig.6 we see the test process \( W(t) \) for one case generated from the Cox model with \( n = 200 \) and its 50 replications \( ˆW(t) \) under the hypothesis of the AFT model. Around the time index 450, the observed process tends to exceed the resampled values, which suggests that the model does not fit the data well. The variance of the resampled processes increases with time, however, and in the displayed case the observed process is well between the resampled ones at the end of the time interval. Therefore the non-standardised statistic may not detect the deviation from the AFT model because the supremum of the resampled processes is near the end where the variance is larger. Variance standardising and division into quartiles helps to overcome this problem.
Figure 6: The statistic $W(t)$ (bold) with its 50 replications under the AFT model, non-standardised and variance-standardised version.
| Test Statistic | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}(W(t))}}$ | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}(W(t))}}$ |
|---------------|-----------------|---------------------------------|-----------------|---------------------------------|
| Censoring     | NC C C NC C NC C NC | C C C C C C C C | NC C C NC C NC C C | NC C C NC C NC C C |

The proportion of wrongfully rejected samples from the AFT model:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>0.002</td>
<td>0.002</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
<td>0.006</td>
</tr>
</tbody>
</table>

The proportion of rightfully rejected samples from the Cox model:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.058</td>
<td>0.054</td>
<td>0.156</td>
</tr>
</tbody>
</table>

The proportion of rightfully rejected samples from the Aalen model:

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>50</th>
<th>100</th>
<th>200</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>0</td>
<td>0.012</td>
<td>0.056</td>
<td>0.180</td>
<td>0.180</td>
<td>0.156</td>
</tr>
</tbody>
</table>

Table 2: The empirical level of significance and the empirical power against the Cox and Aalen model for various sample sizes

3.1.2 Time-varying covariates

Consider data with a single jump in one covariate, $Z_i(t) = I(t > s_i)$. First, we generated data from the AFT model itself, with gamma baseline distribution $\Gamma(p = 20, \lambda = 4/100)$ and $\beta = 1$. The jump times $s_i$ were generated as i.i.d. $\Gamma(p = 20, \lambda = 4/80)$. Censoring times were generated independently with the same distribution of jumps and baseline distribution $\Gamma(p = 20, \lambda = 4/150)$. We applied the test of the AFT model with weights $w_i = I(s_i \leq \text{median}(s_j))$, with the plain supremum statistic, variance-adjusted version and the supremum computed over the quartiles using the Bonferroni correction. For results, see Table 3. As we see, the test holds its level of significance, which is in all cases below 5%. Sometimes, especially for the smaller samples, is
again overly conservative.

| Test Statistic | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}W(t)}}$ | Quartiles - Bonferroni | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}W(t)}}$ |
|----------------|------------------|------------------|-------------------|------------------|-------------------|
| Censoring      | NC C NC C NC C NC C | NC C NC C NC C NC C |
| 50             | 0 0 0.010 0.002 | 0.008 0 0.006 0 |
| 100            | 0 0 0.012 0.006 | 0.004 0.006 0.012 0.006 |
| 200            | 0 0 0.020 0.020 | 0.012 0.010 0.018 0.012 |
| 500            | 0 0.002 0.026 0.024 | 0.008 0.018 0.020 0.026 |
| 1000           | 0 0 0.026 0.038 | 0.014 0.016 0.016 0.034 |
| 2000           | 0 0 0.024 0.032 | 0.022 0.018 0.024 0.026 |

Table 3: The empirical level of significance when generating from the AFT model with a time-varying covariate

Next, we generated from the Cox model $\alpha_i(t) = \exp(Z_i(s)\beta)\alpha_0(t)$ with the same setting and censoring generated the same way. For the obtained empirical power see Table 4. Without standardising or dividing into quartiles, the empirical power is low - this fact is discussed later. Observing the nonstandardised statistic in the quartiles separately yields better results, the power increases with the sample size. With the standardising, the power is even higher, and for each sample size stays approximately the same regardless of dividing into quartiles or censoring.

| Test Statistic | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}W(t)}}$ | Quartiles - Bonferroni | $\sup |W(t)|$ | $\sup \frac{W(t)}{\sqrt{\text{var}W(t)}}$ |
|----------------|------------------|------------------|-------------------|------------------|-------------------|
| Censoring      | NC C NC C NC C NC C | NC C NC C NC C NC C |
| 50             | 0 0 0 0 | 0 0 0 0 |
| 100            | 0 0 0.014 0.014 | 0 0.010 0.008 0.016 |
| 200            | 0 0.010 0.046 0.036 | 0.024 0.024 0.020 0.032 |
| 500            | 0.072 0.088 0.432 0.398 | 0.296 0.274 0.392 0.354 |
| 1000           | 0.434 0.424 0.884 0.752 | 0.754 0.646 0.890 0.706 |
| 2000           | 0.888 0.926 0.988 0.984 | 0.962 0.942 0.986 0.988 |

Table 4: The empirical power against the Cox model with a time-varying covariate

Finally, we generated data from the AFT model with one jump as above and tried fitting the model with misspecified jump times given as covariates. We
randomly chose a half of the jump times $s_i$ and instead of the actual value, a halved time $r_i = s_i/2$ was used for fitting. This leads to incorrect specification of the covariates $Z_i(s) = I(s > r_i)$. As a result, the parameters estimated by the model may give misleading information. It is therefore interesting to see the proportion of cases in which the fit would be rejected. For results, see Tab[5]. The empirical power is reasonably high, with the highest values for a given sample size obtained using the standardised statistics. Censoring does not reduce the power much. Using the plain statistic without standardising or dividing into quartiles, the power is considerably lower.

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>$[\delta, \tau - \delta]$</th>
<th>Sup</th>
<th>Quartiles - Bonferroni</th>
</tr>
</thead>
<tbody>
<tr>
<td>Censoring</td>
<td>sup $</td>
<td>W(t)</td>
<td>$</td>
</tr>
<tr>
<td>50</td>
<td>NC 0 0 0.012 0.010 0.004 0.006 C 0 0.006</td>
<td>NC 0 0.012 0.006 0.224 0.156 0.124 0.024 0.094 0.024</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.012 0.006 0.224 0.156 0.124 0.024 0.094 0.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.340 0.360 0.962 0.926 0.778 0.754 0.792 0.726</td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>0.986 1 1 1 1 1 1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>1 1 1 1 1 1 1 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>1 1 1 1 1 1 1 1 1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: The empirical power against the AFT model with a misspecified jump

For checking why the test does not reject the alternative with the plain statistic, we want to visualize the observed and resampled processes. For one generated data set of $n = 200$ with the misspecified jump times we can see the test process $W(t)$ and its 50 replications under the hypothesis of the AFT model on Fig[7]. The observed process lies between the replicated processes only for smaller and larger time values but exceeds them in between. We may therefore suppose that the model does not fit the data well.

Again, the variance of the processes increases with time. This is the reason for lower rejection rate using the plain supremum statistic, because the supremum of each replicated process lies near the end of the observed interval. The observed process, however, deviates from the model notably around the time index 500, where the supremum is smaller. This can be overcome by using the variance-standardised processes or disregarding the last quartile of the data.
Figure 7: The statistic $W(t)$ (bold) with its 50 replications under the AFT model, non-standardised and variance-standardised version
3.2 Comparison of the basic models

3.2.1 Time-invariant covariates

We generated data from Cox, Aalen and AFT models with a single time-invariant covariate and fitted all three models to the data. Then we performed goodness-of-fit tests for all models to see whether the model is correctly identified. For the Cox and Aalen models we used the tests as presented in Chapter 2. For the AFT model we used the regression tests using the two sample Wilcoxon test and t-test and also the test based on martingale residuals.

For generating we used Gamma baseline distribution with $\Gamma(p = 20, \lambda = 4/100)$ and lognormal distribution $LN(\mu = 5, \sigma^2 = 1)$. These were used as the baseline $\alpha_0(t)$ of the Cox and AFT models as well as for $\beta_0(t)$ in the considered Aalen model $\alpha_i(t) = \beta_0(t) + Z_i \beta_1(t)$. $\beta_1(t)$ was chosen as either $\Gamma(20, 80/100)$ or $LN(4, 1)$ respectively. The covariates $Z_i$ were taken as continuous generated from the normal distribution $N(3, 1)$. We used $\beta = 1$ and $\beta = 2$ for the Cox and Aft model and $\beta_1(t)$ and $2\beta_1(t)$ for the Aalen model to compare how the empirical properties change with increasing influence of the covariates. Each time two variants were explored, with no censoring and with approximately a $1/4$ of the data randomly independently censored. We generated the datasets of size $n = 500$ and we considered the standard level of significance $\alpha = 0.05$. Each case was generated 500 times to see in how many cases would be rejected by the test of the hypothesis of given model. This way we obtain the empirical level of significance and power against given alternatives.

For the tests based on martingale residuals we generated 200 resampled processes and used the variance-standardised version of the supremum test statistics $\sup \left| \frac{U(t)}{\sqrt{\text{var}U(t)}} \right|$, $\sup \left| \frac{M_{K_j}(t)}{\sqrt{\text{var}M_{K_j}(t)}} \right|$ and $\sup \left| \frac{W(t)}{\sqrt{\text{var}W(t)}} \right|$. For testing the Aalen model, we divided the data along the quartiles of the values of $Z_i$ and we reject if the respective minimal p-value among the groups was lower than the Bonferroni-corrected level of significance. For the test of the AFT model we used the weights which yielded best results, see Tab.I (for Cox and AFT and $w_i = I(Z_i \leq \text{median } Z_j)$ for Cox and AFT and $w_i = Z_i I(Z_i \leq \text{median } Z_j)$ for Aalen).
<table>
<thead>
<tr>
<th>Generated model</th>
<th>Tested model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cox</td>
</tr>
<tr>
<td>Model</td>
<td>$U(t)$</td>
</tr>
<tr>
<td>Cox gamma 1 NC</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox gamma 2 NC</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox lognorm 1 NC</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox lognorm 2 NC</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox Aalen gamma 1 NC</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox Aalen lognorm 1 NC</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox Aalen lognorm 2 NC</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox AFT gamma 1 NC</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>Cox AFT lognorm 1 NC</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
</tr>
<tr>
<td></td>
<td>C</td>
</tr>
</tbody>
</table>

Table 6: Empirical level of significance and power for data generated from the basic models
In Tab.6 we see the results for data generated from the basic models. The tests mostly all hold their level of significance, with the tests for the AFT model being the most conservative. The test for the Cox model performs best against the alternative of the AFT model with lognormal distribution. The Aalen model is rarely rejected by the tests, however, the test of the Aalen model rejects the other two models almost in all cases. The power of the test for the AFT model against other models with no censoring is larger for $\beta = 2$ than for $\beta = 1$ and smaller for censored samples. The martingale-based method has a slightly higher empirical power than the regression tests.

3.2.2 Time-varying covariates

Next we generated from the three models with the same baselines as before, but now with a covariate $Z_i(t) = I(t > s_i)$ for randomly generated jump times $s_i$. This means that at the time $s_i$ the hazard is either increased by the factor of $e^\beta$ according to the Cox model, or there is an added non-constant part of the hazard from the time $s_i$ onward, or the device starts to age faster by the factor of $e^\beta$. We perform an analysis as for the case with constant covariates, with the same baseline distributions, both with and without censoring, with $n = 500$. The distribution of the jumps $s_i$ is taken as $\Gamma(20, 4/80)$ or $LN(4, 1)$ for gamma and lognormal baselines, respectively.

In Tab.7 we see that the models mostly hold their respective level of significance, with the exception of the test of the Cox model with the lognormal distribution. The empirical power of the tests is higher when checking data from the gamma distribution than with lognormal. The test for the Aalen model does not reject the data from the Cox model often, possibly due to its nonparametric nature. With larger $\beta$, the tests for the Cox and AFT models tend to yield larger power.
<table>
<thead>
<tr>
<th>Generated model</th>
<th>Tested model</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cox</td>
<td>Aalen</td>
</tr>
<tr>
<td>Model</td>
<td>Baseline</td>
<td>$\beta$</td>
<td>Cens.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$U(t)$</td>
<td>$M_k(t)$</td>
</tr>
<tr>
<td>Cox gamma</td>
<td>1 NC</td>
<td>0.042</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.044</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>0.043</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.044</td>
<td>0.022</td>
</tr>
<tr>
<td>lognorm</td>
<td>1 NC</td>
<td>0.028</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>1 C</td>
<td>0.062</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>0.034</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.070</td>
<td>0.036</td>
</tr>
<tr>
<td>Aalen gamma</td>
<td>1 NC</td>
<td>0.488</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.476</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>0.696</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.596</td>
<td>0.010</td>
</tr>
<tr>
<td>lognorm</td>
<td>1 NC</td>
<td>0.460</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.418</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>0.632</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.572</td>
<td>0.018</td>
</tr>
<tr>
<td>Aft gamma</td>
<td>1 NC</td>
<td>0.476</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.468</td>
<td>0.576</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>1</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>1</td>
<td>0.282</td>
</tr>
<tr>
<td>lognorm</td>
<td>1 NC</td>
<td>0.278</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.268</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>2 NC</td>
<td>0.828</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.710</td>
<td>0.152</td>
</tr>
</tbody>
</table>

Table 7: Empirical level of significance and power for data generated from the models with a time-varying jump covariate
3.2.3 Gamma baseline distribution

The Cox and AFT models are equivalent if their baseline distribution is Weibull for constant covariates and exponential for time-varying covariates. The Gamma distribution $\Gamma(p, \lambda)$ with the density $f_{p,\lambda}(t) = \frac{\lambda^p}{\Gamma(p)} t^{p-1} e^{-\lambda t}$ can be seen as a generalization of the exponential distribution, and becomes exponential when $p = 1$. Thus with $p$ being further from 1, the Cox and AFT models should be easier to distinguish from each other. We use the martingale residuals test for the AFT model on data generated from the Cox model and vice versa, using the setup from above for both constant and time-varying covariates with $n = 500$ and Gamma baseline distribution $\Gamma(p, p/500)$ with increasing values of $p$ and proportionally decreasing values of $\lambda$ to maintain the same expectation. The results in Tab.8 indicate that as expected, the models are easier to differentiate from one another as $p$ increases. In the case of constant covariates, the test of the Cox model performs better on the data from the AFT model than vice versa. With the one-jump covariate, the Cox model is being rejected by the $W(t)$ test in a larger proportion of cases for larger $p$ and less often for smaller $p$. Censoring seem to lower the empirical power by a small amount.

<table>
<thead>
<tr>
<th>Model Generated</th>
<th>const.</th>
<th>time-varying</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFT wrath Cox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested AFT</td>
<td>Cox</td>
<td></td>
</tr>
<tr>
<td>$p$</td>
<td>NC</td>
<td>C</td>
</tr>
<tr>
<td>0.5</td>
<td>0.326</td>
<td>0.192</td>
</tr>
<tr>
<td>1</td>
<td>0.042</td>
<td>0.062</td>
</tr>
<tr>
<td>2</td>
<td>0.262</td>
<td>0.212</td>
</tr>
<tr>
<td>5</td>
<td>0.642</td>
<td>0.422</td>
</tr>
<tr>
<td>10</td>
<td>0.608</td>
<td>0.386</td>
</tr>
<tr>
<td>15</td>
<td>0.562</td>
<td>0.338</td>
</tr>
<tr>
<td>20</td>
<td>0.542</td>
<td>0.358</td>
</tr>
<tr>
<td>25</td>
<td>0.504</td>
<td>0.254</td>
</tr>
<tr>
<td>30</td>
<td>0.520</td>
<td>0.294</td>
</tr>
<tr>
<td>35</td>
<td>0.590</td>
<td>0.326</td>
</tr>
<tr>
<td>40</td>
<td>0.622</td>
<td>0.372</td>
</tr>
<tr>
<td>50</td>
<td>0.735</td>
<td>0.556</td>
</tr>
</tbody>
</table>

Table 8: Empirical level of significance and power for data generated from the basic models
3.3 Modeling of the impact of a strike

When dealing with real-life data, it is desirable to find a model which explains the distribution of the time to failure reasonably well. Here we demonstrate the model fitting and selection methods for a real problem from smelting industry, utilizing additional information as a time-varying covariate.

During the worker’s strike in Quebec aluminium smelter in 1967, the operation of electrolytic smelting cells was shut down on a short notice. After the production was resumed, the smelting cells were more prone to irreparable failures. This was assumed to be caused by sudden cooling or clogging by the aluminium consolidated during the strike. The aim of a following inquiry was to determine, whether the sudden shutdown had a negative impact on the lifetime distribution of the cells or not (Kalbfleisch and Struthers, 1982).

We have the life times of 395 cells at disposal, along with time points \( s_i \) indicating when in their life time the strike occurred. 103 of the cells failed before and 292 after the strike. We fit the basic regression models with a single time varying covariate \( Z_i(t) = I(t > s_i) \) and test their goodness-of-fit using the variance standartised versions of the statistics \( U(t) \), \( M_K(t) \) and \( W(t) \) for the Cox, Aalen and AFT models respectively. For the AFT model we use the weights \( I(s_i \leq \text{median } s) \) and for the Aalen model we divide the data into two parts along the median of \( s_i \) and perform the test on both parts.

In Table 9 we see the parameter estimates of the Cox and AFT models along with their standard deviation and significance, as well as the p-values of the goodness-of-fit tests. The Aalen model is rejected on the first part with \( s_i \leq \text{median } s \), but both the Cox model and the AFT model seem to fit the data well, as the p-values of the goodness-of-fit tests, 0.385 and 0.310, are well above 0.05. The coefficient estimates indicate that after the strike, the hazard is increased \( \exp(0.929) = 2.53 \) times using the Cox model or the internal time of the cells is accelerated by the factor of \( \exp(0.502) = 1.65 \) according to the AFT model. Both effects are highly significant. Further analysis could concern estimating the cost of potential ”lost” days of the cell operation (Volf, 2004).

One could conclude that as both the Cox and AFT model were not rejected by the tests, the baseline distribution should be close to exponential. Therefore the parameter estimates should also be similar. However, as seen in the previous section, the tests do not always distinguish between the models perfectly,
because of their asymptotic nature.

<table>
<thead>
<tr>
<th>Model</th>
<th>(\hat{\beta})</th>
<th>sd (\hat{\beta})</th>
<th>p-val (\beta = 0)</th>
<th>p-val g.o.f.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cox</td>
<td>0.929</td>
<td>0.125</td>
<td>&lt; 0.001</td>
<td>0.385</td>
</tr>
<tr>
<td>AFT</td>
<td>0.502</td>
<td>0.072</td>
<td>&lt; 0.001</td>
<td>0.310</td>
</tr>
<tr>
<td>Aalen</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>(&lt; 0.001, 0.215)</td>
</tr>
</tbody>
</table>

Table 9: Estimates and goodness-of-fit tests of the models fitted to the smelting cells data

### 3.4 Conclusion

In this chapter we studied and compared the properties of the goodness-of-fit testing procedures. As each of the models and possible selection of covariates present different ways to interpret the data, it is desirable to identify which model fits the data best. We saw that in some cases the tests correctly rejected the wrong model in a large proportion of the simulations, whereas in some cases the tests were not able to distinguish between the models as often.

We studied in detail the properties of the newly proposed test for the AFT model based on the martingale residuals. We have seen that the choice of weights for the test statistics can affect the empirical power against the various alternatives. The selection of weights is therefore important for correct identification whether the dependence is sufficiently described by the model.

Finally, we presented an application from industry and demonstrated the use of a time-varying covariate and possible model selection approach.
4. Regression models for repairable systems

4.1 Introduction

In this part we generalize the methods of survival analysis from the last chapters for use in reliability setting. We study data describing a service record of one or more devices which degrade over time. When a device breaks down, it can be repaired and put to use again. Furthermore, preventive maintenance is performed to avoid breakdowns. To optimize the maintenance and repair costs, it is desirable to estimate the distribution of the time to failure with the help of available information.

The main difference as opposed to before is that we do not consider a failure to be the final state. The subject returns into observation after a failure occurs, and can be again subject to consequent failures. The foundations of dealing with repairable systems have been described by Ascher and Feingold (1984) or Brown and Proschan (1983). The modeling of preventive maintenance and its scheduling has been discussed by Kay (1976), Handlarski (1980) or Percy (1996). The important principle we apply here is that the failures, undertaken repairs and maintenance actions may have an influence on the distribution of the consequent failure times. Our aim is to study the means of modeling this influence.

State of the art

As suggested by Brown and Proschan (1983), the repairs after a failure can be of three kinds:

- a perfect repair, returning the device to a working state as new,
- a minimal repair, where the device is returned to a working state in the condition as it was just before the failure, and
- a partial repair, bringing the device to a state somewhere in between.
Consider first, that the repairs do not affect the time-to-failure distribution. In the case with every repair being perfect, the data record of one device forms a renewal process (see Cox, 1970 or Blackwell, 1953) with the times between consecutive failures being i.i.d.. On the other hand, when the repairs are minimal, the data form a non-homogenous Poisson process, discussed in reliability setting by i.e. Crowder et al. (1991). If the failure time distribution is exponential, both cases form a homogenous Poisson process. Inference for this processes has been developed and the failure time distribution can be estimated using standard methods. In the survival setting, Cox (1972b) and Lin et al. (1998) introduced generalizations of the Cox and AFT models respectively for recurrent events with constant covariates.

For modeling the influence of repairs on the time to failure, Kijima (1989) introduced an age-reduction model, where a repair possibly reduces the virtual age of the device. The reduction can range from resetting the device back to the starting age (perfect repair) to no reduction at all (minimal repair). Dorado et al. (1997) developed inference procedures for these models, including the nonparametric estimation of the failure distribution function.

In this chapter we focus on methods of modeling the lifetime of the devices with available regression models of survival analysis with suitable covariates. The models discussed in chapter 2 need to be adjusted to accommodate recurring repairs and maintenance actions. Based on the proportional intensities model by Cox (1972b), Percy and Kobbacy (1998) and Percy and Alkali (2005) introduced a generalised proportional intensities model as an extension the Cox proportional hazards model (Cox, 1972) including the number of repairs and maintenances as time-varying covariates multiplicatively influencing a parametric baseline hazard. In a similar way, we show the use the Accelerated failure time model with time-varying covariates (Lin and Ying, 1995), which states that the covariates influence multiplicatively the flow of the internal time of the device (Novák, 2014). Further, we show methods for estimating the cumulative baseline hazard nonparametrically if we have data on more devices, which allows us to estimate the regression parameters without posing assumptions on the shape of the baseline. For each of the introduced approaches we present large sample properties of the estimates.
4.2 Modeling the lifetime of one device

Let $T_1, ..., T_m$ be random variables representing the ordered times of actions performed on one observed device, denoting both repairs and preventive maintenances. Denote $\Delta_1, \Delta_m$ the indicators whether in $j$-th time a repair ($\Delta_j = 1$) or a preventive maintenance ($\Delta_j = 0$) was performed and let $Z(t) = (Z_1(t), ..., Z_p(t))^T$ be a vector of additional explanatory variables, possibly time-varying.

The data is available either in form of the ordered times of actions $(T_j, \Delta_j)_{j=1}^m$ or in the form of times elapsed between the actions $(T_j - T_{j-1}, \Delta_j)_{j=1}^m$ with $T_0 = 0$, along with the covariate values $Z(t)$. We assume that the time elapsed during an action does not contribute to the total time elapsed, as the device is considered not to be under workload when a repair or maintenance action is being done. The duration of an action may even not be available if the data is presented in the second form.

We work with counting processes denoting the number of repairs $N_\bullet(t)$ and maintenance actions $M_\bullet(t)$ up to time $t$:

$$N_\bullet(t) = \sum_{j=1}^m I(T_j \leq t, \Delta_j = 1), \quad M_\bullet(t) = \sum_{j=1}^m I(T_j \leq t, \Delta_j = 0).$$

Our aim is to model and possibly predict the distribution of the time to failure, depending on the history of the device and available covariates. We work with hazard function denoting the limit probability of immediate breakdown of the device (an increase of $N_\bullet(t)$) given its history:

$$\alpha(t) = \lim_{h \to 0} \frac{P(N_\bullet(t+h) - N_\bullet(t) \geq 1 | \mathcal{H}(t))}{h}$$

where $\mathcal{H}(t)$ is the filtration denoting history of events up to time $t$. Further denote the cumulative hazard function $A(t) = \int_0^t \alpha(s)ds$ and $S(t) = \exp(-A(t))$ and $f(t) = -S'(t)$ corresponding survival function and density of the time to failure distribution. We assume that each repair returns the device to a working state as it was directly before the failure, and that each subsequent repair or maintenance action somehow affects the hazard function. The aim is to determine whether the repairs and maintenance actions increase or decrease the hazard and by how much. Having data on only one device does not yield enough information to use semiparametric approaches, therefore we parametrize the hazard function and estimate the parameters using the maximum likelihood method.
The times of actions are mutually dependent, as the j-th action can only happen after the (j-1)-th action and the hazard function and thus also the distribution is determined by the previous actions. However, if we take the distribution of each $(T_j, \Delta_j)$ conditional on the service history until $T_{j-1}$, we get independent parts. The joint distribution of the data can be written as

$$g((T_n, \Delta_n), ..., (T_1, \Delta_1)) = g((T_n, \Delta_n)|(T_{n-1}, \Delta_{n-1}), ..., (T_1, \Delta_1)) \cdot g((T_{n-1}, \Delta_{n-1})|(T_{n-2}, \Delta_{n-2}), ..., (T_1, \Delta_1)) \cdot ... \cdot g((T_1, \Delta_1))$$

with $g$ standing for the respective distributions.

The terms denoting the breakdowns of the device ($\Delta_j = 1$) can be seen as realizations of the conditional distribution of the time to failure, given that the device survived up to the last action, resulting in terms $f(T_j^-)S(T_j^-)$. The survival function in the denominator is included because the device is returned to a working state after a repair and the next action at $T_j$ happens after the time $T_{j-1}$. The maintenance actions ($\Delta_j = 0$) can be interpreted as right censoring of that distribution, since we do not know for how long the device operate until a breakdown without the preventive maintenance. This contributes to the likelihood with terms $S(T_j)S(T_j^-)$.

Therefore the likelihood can be written as

$$L = \prod_{j=1}^{n} \left( \frac{f(T_j^-)}{S(T_{j-1})} \right)^{\Delta_j} \left( \frac{S(T_j)}{S(T_{j-1})} \right)^{1-\Delta_j} = \prod_{j=1}^{n} \alpha(T_j^-)^{\Delta_j} \cdot S(T_n) \quad (17)$$

and the log-likelihood has the form

$$l = \sum_{j=1}^{n} \Delta_j \log \alpha(T_j^-) - \int_{0}^{T_n} \alpha(t)dt, \quad (18)$$

or for future ease of use

$$l = \sum_{j=1}^{n} \left( \Delta_j \log \alpha(T_j^-) - (A(T_j^-) - A(T_{j-1})) \right). \quad (19)$$

At this point, we want to incorporate the available history and covariates $N(t), M(t)$ and $Z(t)$ into the likelihood through the hazard function, possibly in an easily interpretable way.
4.2.1 Cox model

In the Cox model the covariates affect the hazard function multiplicatively. We assume that a baseline hazard function \( \alpha_0(t) \) is multiplicatively increased or decreased by each repair, maintenance action and other explanatory variables. We work with the hazard function in the form (Percy and Alkali, 2005)

\[
\alpha(t) = \alpha_0(t)e^{M\cdot(t)\rho + N\cdot(t)\sigma + Z^T(t)\beta} = \alpha_0(t)(e^{\rho})^{M\cdot(t)}(e^{\sigma})^{N\cdot(t)}(e^{\beta^T})^{Z(t)}.
\] (20)

As the explanatory variables \( Z(t) \) we can use either fixed covariates as the type or producer of the device, or time-varying information such as the cost of the last repair before time \( t \). If the covariate values change only in the times of observed events and the baseline hazard \( \alpha_0(t) \) is parametric, it is possible to insert the hazard function into the log-likelihood and maximize. This approach was suggested by (Percy and Alkali, 2005) as the generalization of the proportional intensities model of Cox (1972b).

Given the conditional structure of the data explained above, under certain regularity assumptions and conditionally independent distribution of the maintenance times, the parameters estimated by maximizing the likelihood are consistent and their asymptotic distribution is normal. Because we have data on only one device, we cannot use the framework for random processes for proving the normality as for the inference in chapter 2. Instead we use the central limit theorem for martingale differences:

**Theorem 14 (Brown, 1971).**

Suppose that for all \( m \in \mathbb{N} \) there exist \( X_{1m}, X_{2m}, \ldots, X_{mm} \in L_1 \) real random vectors and \( \sigma \)-arrays \( F_{0m} \subset F_{1m} \subset \ldots \subset F_{mm} \). Let

- For all \( m \in \mathbb{N} \) and \( j = 1, \ldots, n \) be \( X_{jm} \) \( F_{jm} \)-measurable and \( E[X_{jm}|F_{j-1,m}] = 0 \).

\[ \sum_{j=1}^{m} E[X_{jm}^2|F_{j-1,m}] \overset{P}{\rightarrow} I \text{ for } n \rightarrow \infty. \]

\[ \sum_{j=1}^{m} E[X_{jm}^{<2}I_{\|X_{jm}\|\geq\epsilon}|F_{j-1,m}] \overset{P}{\rightarrow} 0 \text{ for } n \rightarrow \infty. \]

Then for \( n \rightarrow \infty \)

\[ \sum_{j=1}^{m} X_{jm} \overset{D}{\rightarrow} N(0, I). \]

**Proof.** See Brown (1971) \qed

67
Denote $\gamma$ the vector of the baseline hazard function parameters, using $\alpha_0(t) = \alpha_0(t, \gamma)$. Further let $\beta = (\rho, \sigma, \beta)^T$, $X(t) = (N(t), M(t), Z^T(t))$ and $\theta = (\beta^T, \gamma^T)^T$. Denote the score and its components

$$U(\theta) = \frac{\partial}{\partial \theta} l(\theta) = \sum_{j=1}^m U_j(\theta) = \sum_{j=1}^m \frac{\partial}{\partial \theta} (\Delta_j \log \alpha(T_j^-) - A(T_j^-) + A(T_j^-_1)).$$

Let $H_k = H(T_k)$ indicate the history of events up to the time $T_k$. Take the parameter estimates $\hat{\theta}$ as the solution of $U(\theta) = 0$. Then we can establish their asymptotic properties as follows:

**Theorem 15.**

Suppose that

- $(T_j, \Delta_j, Z(t))$, $j = 1, 2, \ldots$ follow the Cox model for repairable systems \cite{20},
- $Z(t)$ is bounded on $\mathbb{R}_+$ and $E|\frac{d\alpha_0(T_j)}{\alpha_0(T_j)}| < \infty$,
- the maintenance distribution and the time-to-failure distribution of $T_j$ are conditionally independent given $H_{j-1}$,
- the baseline distribution fulfills the standard regularity conditions (Lehmann, 1994, p.464).

Let $\theta_0$ be the real value of the parameters. Suppose there exists a set $B$ surrounding $\theta_0$ such that for all $\theta \in B$:

(a) there exists a finite positive semidefinite matrix $\Sigma$, that

$$-\frac{1}{m} E[U_j(\theta)U_j(\theta)^T|H_{j-1}] \overset{P}{\to} \Sigma.$$

(b) \[
\frac{1}{m} \sum_{j=1}^m E[U_j(\theta)U_j(\theta)^T I_{\frac{\partial}{\partial \theta} |U_j(\theta)| \geq \epsilon |H_{j-1}|} \overset{P}{\to} 0 \quad \forall \epsilon > 0.
\]

(c) Take $\alpha_j(t, \theta) = \epsilon x^T(T_j-1)^\beta \alpha_0(t, \gamma)$. Suppose that there exist bounded functions $G_j(t)$ and $H_j(t)$, such that

$$\frac{1}{n} \sum_{j=1}^m \int_0^\infty G_j(t)Y_j(t)dt \quad \text{and} \quad \frac{1}{n} \sum_{j=1}^m \int_0^\infty H_j(t)\alpha_j(t, \theta_0)Y_j(t)dt$$

converge in probability to finite quantities and $\forall \epsilon > 0$

$$\frac{1}{n} \sum_{j=1}^m \int_0^\infty H_j(t)I_{\frac{\partial}{\partial \theta} H_j^{-1/2}(t) \geq \epsilon} \alpha_j(t, \theta_0)Y_j(t)dt \overset{P}{\to} 0,$$
for which

$$\sup_{\theta \in \mathcal{B}} \left| \frac{\partial^3}{d\theta_1^3 \theta_2^3 \theta_3^3} \alpha_j(t, \theta) \right| \leq G_j(t) \quad \text{and} \quad \sup_{\theta \in \mathcal{B}} \left| \frac{\partial^3}{d\theta_1^3 \theta_2^3 \theta_3^3} \log \alpha_j(t, \theta) \right| \leq H_j(t)$$

\forall t, j and \(k_1, k_2, k_3\). Then for \(m \to \infty\):

$$\frac{1}{\sqrt{m}} U(\theta_0) \overset{D}{\to} N(0, \Sigma),$$

$$\sqrt{m}(\hat{\theta} - \theta_0) \overset{P}{\to} N(0, \Sigma^{-1}),$$

$$\frac{1}{m} I(\hat{\theta}) = - \frac{1}{m} \frac{\partial^2}{\partial \theta \partial \theta^T} \log l(\theta) \overset{P}{\to} \Sigma.$$

**Proof.** We want to use the Brown’s CLT for the parts of the score process using \(X_{jm} = \frac{1}{\sqrt{m}} U_j(\theta)\). For the Cox model, we have

$$U_j(\theta) = \frac{\partial}{\partial \theta} \left( \Delta_j \log \alpha(T_j^-) - A(T_j^-) + A(T_{j-1}) \right)$$

$$= \frac{\partial}{\partial \theta} \left( \Delta_j (N(T_{j-1}) \rho + M(T_{j-1}) \sigma + Z(T_{j-1}) \beta + \log \alpha_0(T_j)) \right)$$

$$- e^{N(T_{j-1}) \rho + M(T_{j-1}) \sigma + Z(T_{j-1}) \beta}(A_0(T_j^-) + A_0(T_{j-1}))).$$

$$= \begin{pmatrix} 
\Delta_j M_0(T_j^-) - (A_0(T_j) - A_0(T_{j-1}))e^{X^T(T_j^-) \beta} M_0(T_j^-) \\
\Delta_j N_0(T_j^-) - (A_0(T_j) - A_0(T_{j-1}))e^{X^T(T_j^-) \beta} N_0(T_j^-) \\
\Delta_j Z(T_j^-) - (A_0(T_j) - A_0(T_{j-1}))e^{X^T(T_j^-) \beta} Z(T_j^-) \\
\Delta_j \frac{d}{dt} \alpha_0(T_j^-) - \frac{d}{dt} (A_0(T_j) - A_0(T_{j-1}))e^{X^T(T_j^-) \beta}
\end{pmatrix}.$$

Using the indicator processes of the \(j\)-th failure and the corresponding at-risk processes

$$N_j(t) = \Delta_j I(T_j \leq t) \quad \text{and} \quad Y_j(t) = I(T_{j-1} < t \leq T_j)$$

and denoting

$$X^s_j(t) = \left( M_0(T_{j-1}), N_0(T_{j-1}), Z^T(T_{j-1}), \frac{d}{dt} \alpha_0(t) \right)^T$$

and

$$\mathbb{M}_j(t, \theta) = N_j(t) - \int_0^t Y_j(s)e^{X^T(s) \beta} \alpha_0(s) ds$$
we can rewrite the score as

\[ U_j(\theta) = \int_0^\infty X_j^\alpha(t) dM_j(t, \theta). \]

For a fixed \( j \) and \( m \), \( X(t) \) and \( E[|\frac{\partial \alpha_0(T_j)}{\partial \alpha_0}|] \) are bounded. Given the regularity of the baseline distribution we get that \( X_{jm} = \frac{1}{\sqrt{m}} U_j(\theta) \in L_1. \)

Previously we denoted \( \mathcal{H}(t) \) the sigma-field generated by the history of the events up to time \( t \). By taking \( \mathcal{F}_{jm} = \mathcal{H}_j \), we see that the fields form an increasing sequence and that \( X_{jm} \) are \( \mathcal{H}_j \)-measurable.

Given the history up to time \( T_{j-1} \), the processes \( M_j(t, \theta_0) \) are zero mean \( \mathcal{H}_t \)-martingales, as \( Y_j(s)e^{X^T(s^-)}\beta_0\alpha_0(s) \) are the intensities of the processes \( N_j(t) \) given \( \mathcal{H}_{j-1} \) under the Cox model (see (16) and (20)) and the processes \( \int_0^t Y_j(s)e^{X^T(s^-)}\beta_0\alpha_0(s)ds \) form the compensators for \( N_j(t) \). Therefore, because \( X_j^\alpha(t) \) is a non-random quantity given \( \mathcal{H}_{j-1} \), we get

\[ E[U_j(\theta_0)|\mathcal{H}_{j-1}] = E[\int_0^\infty X_j^\alpha(t) dM_j(t, \theta_0)|\mathcal{H}_{j-1}] = 0 \quad \forall j \quad \forall m. \]

The conditions (a) and (b) provide the finiteness of the variance matrix and the Feller-Lindeberg condition necessary for the conditional central limit theorem, giving us

\[ \frac{1}{\sqrt{n}} U(\theta_0) \overset{D}{\to} N(0, \Sigma) \]

using the Slutsky’s theorem for the variance standardisation.

For establishing the asymptotic normality of the parameter estimates \( \hat{\theta} \), we need first to prove their consistency. Because of the martingale representation via \( M_j(t, \theta_0) \), asymptotic theory for parametric survival models can be employed at this point. Taking the conditional likelihood contributions \( \alpha_j(t)S_j(t)/S(T_{j-1}) \) along with conditions (a)-(c), consistency of the parameter estimates is obtained using Theorem 1 of Borgan (1984) or the results of Andersen et al. (1993), Ch.VI.

Using Taylor expansion along with condition (c) gives us

\[ \sqrt{m}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{m}} \Sigma^{-1} \sum_{j=1}^m U_j(\theta_0) + o_P(1), \]
giving us the asymptotic result for the parameter estimates. The convergence of the observed information matrix follows from the strong law of large numbers for martingale differences.

The conditions (a)-(c) impose certain restrictions on the baseline hazard and the regression parameters $\rho$ and $\sigma$, which counterweights the increasing number of repairs and maintenance actions included as covariates as $m \to \infty$. Using the martingale representation, condition (a) can be rewritten as

$$\frac{1}{m} \sum_{j=1}^{m} E[U_j(\theta)U_j(\theta)^T|\mathcal{H}_{j-1}] = \frac{1}{m} \sum_{j=1}^{m} E[(\int_0^\infty X_j^\alpha(t)dM_j(\theta))^{\otimes 2}|\mathcal{H}_{j-1}]$$

$$= \frac{1}{m} \sum_{j=1}^{m} E[\int_0^\infty X_j^\alpha(t)(X_j^\alpha(t))^TY_j(t)e^{X(t)^T\beta} \alpha_0(t)dt|\mathcal{H}_{j-1}] < \infty.$$

Because $X_j^\alpha$ contains as elements $M_*(T_j^-)$ and $N_*(T_j^-)$ which increase with $m$, either one of the parameters $\rho$ or $\sigma$ has to be negative or the baseline hazard needs to be sufficiently decreasing for the sum to be convergent. The Lindeberg condition (b) and the third-derivative bound (c) strengthen this necessity but are not straightforward to interpret.

### 4.2.2 Accelerated failure time model

We can also assume that each repair or maintenance causes that the internal time of the device flows faster or slower (Accelerated failure time model, AFT). We use the time transformation (Lin and Ying, 1995)

$$t \to \int_0^t e^{M_*(s)\rho + N_*(s)\sigma + Z^T(s)\beta} ds =: h(t, \beta),$$

where $\beta = (\rho, \sigma, \beta)^T$. In this framework $t$ represents the observed age and $h(t, \beta)$ the internal age of the device. The hazard function has the form

$$\alpha(t) = \alpha_0(h(t, \beta))e^{M_*(t)\rho + N_*(t)\sigma + Z^T(t)\beta}.$$  \hspace{1cm} (21)

If the baseline hazard function is constant (corresponding with the exponential distribution), both models coincide. It is again possible to insert into the log-likelihood and obtain the parameter estimates by its maximization. The consistency and asymptotic normality of the parameter estimates follows in a similar way as for the Cox model.
The score has a different form. Inserting the hazard function into the likelihood (19) yields

\[ U(\theta) = \sum_{j=1}^{m} \frac{d}{d\theta} \left( \log \alpha_0(h(T_j^-, \beta)) + M(T_j^-)\rho + N(T_j^-)\sigma + Z^T(T_j^-)\beta \right) - A_0(h(T_j^-, \beta)) + A_0(h(T_{j-1}^-, \beta)). \]

**Theorem 16.**
Suppose that

- \((T_j, \Delta_j, Z(t)), j = 1, 2, \ldots\) follow the AFT model for repairable systems (21),
- \(Z(t)\) is bounded on \(\mathbb{R}_+\) and \(E|\frac{d}{d\gamma}\alpha_0(h(T_j, \beta_0))| < \infty\),
- the maintenance distribution and the time-to-failure distribution of \(T_j\) are conditionally independent given \(H_{j-1}\),
- the baseline distribution fulfills the standard regularity conditions (Lehmann, 1994, p.464).

Let \(\theta_0\) be the real value of the parameters. Suppose there exists a set \(B\) surrounding \(\theta_0\) such that for all \(\theta \in B\):

1. there exists a finite positive semidefinite matrix \(\Sigma\), that

\[ -\frac{1}{m} E[U_j(\theta)U_j(\theta)^T|H_{j-1}] \xrightarrow{P} \Sigma. \]

2. \(\frac{1}{m} \sum_{j=1}^{m} E[U_j(\theta)U_j(\theta)^TI_{1\sqrt{m}|U_j(\theta)|\geq\epsilon|H_{j-1}}] \xrightarrow{P} 0 \quad \forall \epsilon > 0, \]

3. Take \(\alpha_j(t, \theta) = e^{X^T(T_j^-)\beta_0}h(t, \beta, \gamma)\). Suppose that there exist bounded functions \(G_j(t)\) and \(H_j(t)\) fulfilling the conditions of Theorem 15 with \(\alpha_j(t, \theta)\) of the AFT model.

Then for \(m \to \infty\):

\[ \frac{1}{\sqrt{m}} U(\theta_0) \xrightarrow{D} N(0, \Sigma), \]

\[ \sqrt{m} (\hat{\theta} - \theta_0) \xrightarrow{D} N(0, \Sigma^{-1}), \]

\[ \frac{1}{m} I(\hat{\theta}) = -\frac{1}{m} \frac{\partial^2}{\partial \theta \partial \theta^T} \log l(\theta) \xrightarrow{P} \Sigma. \]
Proof. Again we want to use the CLT for martingale differences (Theorem 14) by Brown (1971) for $X_{jm} = \frac{1}{\sqrt{m}}U_j(\theta)$. Denote $\alpha'_0(t) = \frac{d}{dt}\alpha_0(t)$ The score parts have the form

$$U_j(\theta) = \left( \begin{array}{c} \Delta_j \left( \frac{\alpha'(h(T_j^-, \beta))}{\alpha(h(T_j^-, \beta))} \int_0^{T_j} X(t)e^{X(t)^\beta}dt + X(T_j^-) \right) \\ -\alpha_0(h(T_j, \beta)) \int_0^{T_j} X(t)e^{X(t)^\beta}dt \\ +\alpha_0(h(T_{j-1}, \beta)) \int_0^{T_{j-1}} X(t)e^{X(t)^\beta}dt \\ \Delta_j \frac{d}{d\gamma} \left( \frac{\alpha'_0(h(T_j^-, \beta))}{\alpha_0(h(T_j^-, \beta))} \right) - \frac{d}{d\gamma} \left( A_0(h(T_j^-, \beta)) - A_0(h(T_{j-1}^-, \beta)) \right) \end{array} \right)$$

The first part consists of $2 + p$ components corresponding to $\rho$, $\sigma$ and $\beta$, the last part corresponds to the baseline parameters $\gamma$. We can rewrite the score using the processes $N_j(t) = \Delta_j I(T_j \leq t)$ and $Y_j = I(T_{j-1} < t \leq T_j)$. If we denote

$$X_j^\alpha(t) = \left( \frac{\alpha'(t^-, \beta)}{\alpha(h(t^-, \beta))} \right) \int_0^t X^T(s)e^{X^T(s)\beta}ds + X^T(t^-), \frac{d}{d\gamma} \left( \frac{\alpha'_0(h(t^-, \beta))}{\alpha_0(h(t^-, \beta))} \right)$$

and

$$M_j(t, \theta) = N_j(t) - \int_0^t Y_j(s)e^{X^T(s)\beta}\alpha_0(h(s, \beta))ds$$

we obtain the score parts in the form

$$U_j(\theta) = \int_0^\infty X_j^\alpha(t) dM_j(t, \theta).$$

For each $m \in \mathbb{N}$ and $j = 1, ..., m$ is then $E[\frac{1}{\sqrt{m}}U_j(\theta)] < \infty$ due to the regularity, boundedness of covariates and the fact that $E\left| \frac{d}{d\gamma} \left( \frac{\alpha'_0(h(T_j, \beta))}{\alpha_0(T_j)} \right) \right| < \infty$. The martingale property also follows, as with given $\mathcal{H}_{j-1}$, the elements of $X_j^\alpha$ are non-random and $M_j(t)$ are $\mathcal{H}(t)$-martingales due to $Y_j(t)e^{X^T(t)\beta}\alpha_0(h(t, \beta))$ being the intensities of $N_j(t)$ (see eq. (16) and (21)), thus $E[U_j(\theta)|\mathcal{H}_{j-1}] = 0$. Using the conditions (a) and (b) to establish the variance matrix and the Lindeberg condition gets us

$$\frac{1}{\sqrt{m}} U(\theta_0) \overset{D}{\rightarrow} N(0, \Sigma).$$

The asymptotics for the parameter estimates and observed information matrix follow in the same way as in the previous proof, using the approach of Borgan (1984), Taylor expansion and SLNN for martingale differences. □
4.3 Inference when observing more devices

Suppose we have data on \( n \) independent devices and for each device \( i = 1, \ldots, n \), we have observed \( m_i \) events. We limit the observed interval to \([0, \tau]\) with \( 0 < \tau < \infty \) and establish the inference using increasing number of observed devices instead of performed actions. Let us have \( \alpha_i(t), T_{ij}, \Delta_{ij}, j = 1, \ldots m_i \) and \( Z_i(t) \) the hazard function, times of events, repair indicators and covariate values for the \( i \)-th device respectively. Thus the \( j \)-th event on the \( i \)-th device occurred at the time \( T_{ij} \) and it was a repair if \( \Delta_{ij} = 1 \) and a preventive maintenance action if \( \Delta_{ij} = 0 \). Denote

\[
N_{ij}(t) = \Delta_{ij} I(T_{ij} \leq t),
M_{ij}(t) = (1 - \Delta_{ij}) I(T_{ij} \leq t),
Y_{ij}(t) = I(T_{i,j-1} < t \leq T_{ij})
\]

the indicators, whether on the \( i \)-th device at time \( t \) the \( j \)-th action already occurred at the time \( t \) and whether it was a repair (\( N_{ij}(t) = 1 \)) or a maintenance (\( M_{ij}(t) = 1 \)) or that the device is still at risk before the \( j \)-th action (\( Y_{ij}(t) = 1 \)). The hazard function for the \( i \)-th device is defined as

\[
\alpha_i(t) = \lim_{h \to 0} P(N_i(t + h) - N_i(t) \geq 1 | H(t))/h
\]

where

\[
H(t) = \sigma\{N_{ij}(t), M_{ij}(t), Y_{ij}(t), Z_i(t), i = 1, \ldots, n, j = 1, \ldots, m_i, 0 \leq s \leq t\}
\]

is the history of events up to time \( t \). We work with \( \bullet \) meaning the sum over corresponding index:

\[
N_\bullet(t) = \sum_{j=1}^{m_i} N_{ij}(t), \quad M_\bullet(t) = \sum_{j=1}^{m_i} M_{ij}(t) \quad \text{and} \quad Y_\bullet(t) = \sum_{j=1}^{m_i} Y_{ij}(t).
\]

Because the devices are independent, the joint likelihood of the data can be taken as the product of likelihoods for each device given by (17). We get the log-likelihood in form

\[
l = \sum_{i=1}^{n} \left( \sum_{j=1}^{m_i} \Delta_{ij} \log \alpha_i(T_{ij}^{-}) - \int_{0}^{T_{i,m_i}} \alpha_i(t) dt \right)
\]
\[ = \sum_{ij} \int_0^\tau \left( \log \alpha_i(t^-) dN_{ij}(t) - Y_{ij}(t) \alpha_i(t^-) dt \right). \] (23)

The hazard function \( \alpha_i \) will contain the counts of repairs and maintenance actions \( N_i(t) \) and \( M_i(t) \).

When using the Cox or AFT models, two options are available at this point. We can either parametrize the baseline hazard and proceed as in the last section, or it is possible to estimate the baseline hazard nonparametrically. This may be desirable, since we then do not need to pose any assumptions on the form of the baseline and can focus on the regression parameters.

### 4.3.1 Semiparametric Cox model

Denote \( X_i^T(t) = (N_i(t), M_i(t), Z_i^T(t)) \). Suppose that

\[ \alpha_i(t) = e^{X_i^T(t) \beta_0(t)}, \] (24)

where \( \alpha_0(t) \) is an unspecified continuous baseline hazard function. Inserting into (23) we get the likelihood under the Cox model in the form

\[ l = \sum_{ij} \int_0^\tau \left( \log \alpha_0(t^-) + X_i^T(t^-) \beta \right) dN_{ij}(t) - Y_{ij}(t) e^{X_i^T(t^-) \beta} \alpha_0(t^-) dt \]

and we obtain the score function by taking derivative with respect to the regression parameters:

\[ U(\beta) = \sum_{ij} \int_0^\tau \left( X_i^T(t^-) dN_{ij}(t) - Y_{ij}(t) X_i^T(t^-) e^{X_i^T(t^-) \beta} dA_0(t) \right). \]

The score depends on an unknown cumulative baseline hazard \( A_0(t) \). This can be replaced by the Nelson-Aalen type estimate

\[ \hat{A}_0(t, \beta) = \sum_{ij} \int_0^t \frac{dN_{ij}(s)}{\sum_{kl} e^{X_k^T(s^-) \beta} Y_{kl}(s)}. \]

Inserting the estimate we get the score function in form

\[ U(\beta) = \sum_{ij} \int_0^\tau \left( X_i(t^-) - \frac{\sum_{kl} X_k(t^-) e^{X_k^T(t^-) \beta} Y_{kl}(t)}{\sum_{kl} e^{X_k^T(t^-) \beta} Y_{kl}(t)} \right) dN_{ij}(t) \]

\[ = \sum_{i} \int_0^\tau \left( X_i(t^-) - \frac{\sum_{kl} X_k(t^-) e^{X_k^T(t^-) \beta} Y_{kl}(t)}{\sum_{kl} e^{X_k^T(t^-) \beta} Y_{kl}(t)} \right) dN_{i}(t). \]
Now we can find the parameter estimates $\hat{\beta}$ by solving the equations $U(\beta) = 0$. The asymptotic properties of the score and parameter estimates are obtained in a different way than when observing just one device, using the central limit theorem for continuous martingales. First, we consider processes $M_i(t, \beta)$ defined as

$$M_i(t, \beta) = N_i(t) - \int_0^t Y_i(s)e^{X_i(s)T\beta}dA_0(s).$$

It follows, that the expectation of $dM_i(t, \beta_0)$ is zero for all $t$, motivating the Nelson-Aalen estimate. With some algebra, the score can be rewritten as

$$U(\beta) = \sum_i \int_0^\tau \left( X_i(t^-) - \frac{\sum_{kl}X_{kl}(t^-)e^{X_{kl}(t^-)T\beta}Y_{kl}(t)}{\sum_{kl}e^{X_{kl}(t^-)T\beta}Y_{kl}(t)} \right) dM_i(t).$$

We want to use the martingale central limit theorem of Rebolledo (1980). Define

$$S_0(t, \beta) = \sum_{ij} Y_{ij}(t)e^{X_{ij}(t^-)T\beta}, \quad S_1(t, \beta) = \sum_{ij} X_{ij}(t^-)Y_{ij}(t)e^{X_{ij}(t^-)T\beta},$$

$$S_2(t, \beta) = \sum_{ij} X_{ij}(t)X_{ij}(t^-)^TY_{ij}(s)e^{X_{ij}(s)T\beta}, \quad E(t, \beta) = \frac{S_1(t, \beta)}{S_0(t, \beta)}.$$

Under the assumptions of the Cox model for repairable systems, $M_i(t, \beta_0)$ are zero-mean $\mathcal{H}(t)$-martingales. Thus we can use the martingale CLT using the counting processes $N_i(t)$, intensities $Y_i(t)e^{X_i(t^-)T\alpha_0(t)}$ and

$$H_i(t) = X_i(t^-) - E(t, \beta_0).$$

As the processes are defined only on a finite interval $[0, \tau]$, we cannot base the asymptotics on letting $m_i \to \infty$. Moreover, $N_i(\tau)$ and $M_i(\tau)$ have to be bounded to some extent.

**Theorem 17.**

Suppose that:

(a) The Cox model for repairable systems (24) holds, with $(N_i(t), M_i, Y_i(t), Z_i(t)), i = 1, \ldots, n$ (i.i.d.).

(b) The preventive maintenance distribution is independent on the time-to-failure distribution of $T_{ij}$ given $\mathcal{H}(T_{i,j-1})$.

(c1) The baseline hazard $\alpha_0(t)$ is finite and continuous on $[0, \tau]$.

(c2) There exists such a neighborhood $\mathcal{B}$ of $\beta_0$, such that

$$E\left[ \sup_{[0,\tau], \beta \in \mathcal{B}} |X_i(t^-)X_i(t^-)^T|e^{X_i(t^-)T\beta} \right] < \infty.$$
(d) There exists a positive definite $p \times p$ matrix $\Sigma(t)$ of functions continuous on $[0, \tau]$, such that

$$\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( \frac{S_2(s, \beta_0)}{S_0(s, \beta_0)} - E(s, \beta_0)^{\otimes 2} \right) S_0(s, \beta_0) dA_0(s) \xrightarrow{P} \Sigma(t)$$

for each $t \in [0, \tau]$ as $n \to \infty$.

(e) $P(Y_{i\bullet}(t) = 1 \forall t) > 0 \forall i$

Then

$$\frac{1}{\sqrt{n}} U(\beta_0) \xrightarrow{D} N(0, \Sigma),$$

$$\sqrt{n}(\beta - \beta_0) \xrightarrow{D} N(0, \Sigma^{-1}),$$

$$\frac{1}{n} I(\theta) = -\frac{1}{n} \frac{\partial^2}{\partial \theta \partial \theta^T} \log l(\theta) \xrightarrow{P} \Sigma.$$

for $n \to \infty$, with $\Sigma = \Sigma(\tau)$.

Proof. The processes $H_i(t)$ consist of terms

$$X_i(t^-) = (N_{i\bullet}(t^-), M_{i\bullet}(t^-), Z_{i\bullet}^T(t^-))^T$$

and

$$\sum_{k} \frac{X_{k}(t^-) e^{X_{k}(t^-) \beta} Y_{k}(t)}{\sum_{l} e^{X_{l}(t^-) \beta} Y_{l}(t)}.$$

The counting processes $N_{i\bullet}$ and $M_{i\bullet}$ as well as the covariates $Z_i$ change only at the times of events. Because we use their values at $t^-$ in the hazard, they are left-continuous and $H_i(t)$ adapted. The same holds for the at-risk processes $Y_{ij}$, therefore $H_i(t)$ are also left-continuous, $H(t)$ adapted and thus locally bounded and predictable.

Verifying of conditions (m-c) and (m-d) of Theorem 3 is now a straightforward extension of the proof of Andersen and Gill (1982) for processes with one event and of that of Lin et al. (2000) for recurrent events. We will present the main steps. Condition (d) yields that

$$\sum_{i=1}^{n} \int_{0}^{t} H_i(s) H_i(s)^T Y_i(s) \alpha_i(s) ds$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( X_i(s^-) - E(s, \beta) \right)^{\otimes 2} Y_{i\bullet}(s) e^{X_i(t^-) \beta} dA_0$$

$$= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( \frac{S_2(s, \beta_0)}{S_0(s, \beta_0)} - E(s, \beta)^{\otimes 2} \right) S_0(t, \beta) dA_0 \xrightarrow{P} \Sigma(t).$$
The Lindeberg condition \((m-d)\) is proven by using the inequality

\[ |a - b|^2 I_{|a - b| \geq \varepsilon} \leq 4|a|^2 I_{|a| \geq \varepsilon} + 4|b|^2 I_{|b| \geq \varepsilon} \]

and proving that each part is asymptotically negligible by using \((c2)\). This gives us the first results.

To prove the consistency of the parameter estimates \(\hat{\beta}\), we observe the function

\[
\chi(\beta) = \frac{1}{n} \sum_{ij} \int_0^\tau (\beta - \beta_0)^T X_i(t) dN_{ij}(t) - \int_0^\tau \log \frac{S_0(t, \beta)}{S_0(t, \beta_0)} dN_{ij}(t).
\]

This function is concave, as its second derivative is a negative semidefinite matrix. Using the strong law of large numbers and convex/concave function theory (Rockafellar, 1970), \(\sup_{\beta : \|\beta - \beta_0\| \leq r} \|\chi(\beta) - E\chi(\beta)\| \to 0\) for any \(r\). Since \(\frac{d}{d\beta} E\chi(\beta_0) = 0\) and \(\frac{d^2}{d\beta^2} E\chi(\beta_0) = -\Sigma(\tau)\) and \(\Sigma(\tau)\) is positive definite using condition \((e)\), \(E\chi(\beta)\) is maximized by \(\beta_0\). This gives \(\sup_{\beta : \|\beta - \beta_0\| = r} E\chi(\beta) < E\chi(\beta_0)\) and thus also \(\sup_{\beta : \|\beta - \beta_0\| = r} \chi(\beta) < \chi(\beta_0)\) for any small \(r\) and large \(n\). Hence there must be a maximizer of \(\chi(\beta)\) on \(\|\beta - \beta_0\| < r\) and it can be uniquely taken as the solution \(\hat{\beta}\) of \(\frac{d}{d\beta} \chi_0(\beta_0) = 0\). This is in fact the score function, resulting in \(\hat{\beta} \to \beta_0\) almost surely. The convergence of \(\sqrt{n}(\hat{\beta} - \beta_0)\) follows immediately using Taylor expansion.

As proven by Andersen and Gill (1982), the condition \((c2)\) yields that the processes \(\frac{1}{n} S_0(t, \beta), \frac{1}{n} S_1(t, \beta), \frac{1}{n} S_2(t, \beta)\) converge uniformly almost surely to some finite functions \(s_0, s_1\) and \(s_2\) for each \(\beta \in \mathcal{B}\). Condition \((e)\) then gives that \(E(t, \beta)\) and \(V(t, \beta) = \frac{S_2(t, \beta)}{S_0(t, \beta_0)} - E(t, \beta)\) converge uniformly to finite functions \(e\) and \(v\). The convergence of \(\frac{1}{n} I(\theta)\) is then verified by observing

\[
\|\frac{1}{n} I(\beta) - \Sigma(\beta_0)\| = \|\frac{1}{n} \int_0^\tau V(\beta) dN_{\bullet\bullet}(t) - \int_0^\tau v(t, \beta_0) s_0(t, \beta_0) dA_0(t)\|
\leq \|\int_0^\tau (V(t, \beta) - v(t, \beta)) \frac{dN_{\bullet\bullet}(t)}{n}\| + \|\int_0^\tau (v(t, \beta) - v(t, \beta_0)) \frac{dN_{\bullet\bullet}(t)}{n}\|
+ \|\int_0^\tau v(t, \beta_0) \frac{dN_{\bullet\bullet}(t)}{n} - S_0(t, \beta_0) dA_0(t)\| + \|\int_0^\tau v(t, \beta_0) \frac{S_0(t, \beta_0)}{n} - s_0(t, \beta_0) dA_0(t)\|.
\]

All terms on the right hand side tend to zero for \(\beta \to \beta_0\) and \(n \to \infty\), giving the final result. \(\square\)

**Remark 18.** In fact, the at-risk process \(Y_{i\bullet}(t)\) is equal to one for all \(t \in [0, T_{im_i}]\). We could introduce a censoring mechanism into the model, specifying that each device is observed only until a certain time \(C_i\) independent on \(T_{ij}\)
Then it would be possible to work with modified processes $Y_{ij} = I(T_{i,j-1} < t \leq \min(T_{ij}, C_i))$, $N_{ij}(t) = \Delta_{ij}\delta_{ij}I(\min(T_{ij}, C_i) \leq t)$ with $\delta_{ij} = (T_{ij} \leq C_i)$. The condition (e) then bounds $S_0(t)$ away from zero.

Remark 19. Condition (c2) poses a restriction on the covariates, including the number of repairs and maintenances $N_{i\cdot}(t)$ and $M_{i\cdot}(t)$ and their respective distributions. However, because the data are observed on a bounded interval, the condition is not as restrictive as with the fully parametric models for one device.

### 4.3.2 Semiparametric AFT model

Using the Accelerated failure time model we assume that the internal time flows differently for each device and that its rate changes at different moments. Therefore for each device we need its time transformation

$$t \rightarrow h_i(t, \beta) = \int_0^t e^{X_i(s)^T \beta} ds$$

and we assume that the hazard function has the form

$$\alpha_i(t) = e^{X_i(t)^T \beta_0} \alpha_0(h_i(t, \beta_0)) \tag{25}$$

with an unspecified baseline $\alpha_0(t)$ and $X_i(t)^T = (N_i(t), M_i(t), Z_i(t)^T)$. Our aim is to establish the inference with respect to the transformed, virtual time. We work with time-transformed processes

$$N_{ij}^*(t, \beta) = \Delta_{ij}I(h_i(T_{ij}, \beta) \leq t),$$

$$M_{ij}^*(t, \beta) = (1 - \Delta_{ij})I(h_i(T_{ij}, \beta) \leq t),$$

$$Y_{ij}^*(t, \beta) = I(h_i(T_{ij-1}, \beta) < t \leq h_i(T_{ij}, \beta)),$$

$$X_i^*(t, \beta) = X_i(h_i^{-1}(t, \beta)).$$

The score obtained by inserting the hazard functions (25) into the log-likelihood (23) and taking the derivative with respect to $\beta$ has the form

$$U(\beta) = \sum_{ij} \int_0^T Q_i(t, \beta) \left( dN_{ij}^*(t, \beta) - Y_{ij}^*(t, \beta) dA_0(t) \right),$$

79
where \( Q_i(t, \beta) = \frac{\alpha'_0(t)}{\alpha_0(t)} \int_0^{h_i^{-1}(t, \beta)} X_i^T(s) e^{X_i^T(s) \beta} ds + X_i^*(t, \beta) \). This form is relatively complicated, with terms \( \alpha'_0 \) and \( \alpha_0 \) not easy to estimate. The exact score can be replaced by the approximate score (Lin and Ying, 1995):

\[
U(\beta) = \sum_{ij} \int_0^T X_i^*(t^-, \beta) \left( dN_{ij}^*(t, \beta) - Y_{ij}^*(t, \beta) dA_0(t) \right).
\]

The unknown cumulative baseline hazard function \( A_0 \) is replaced by its Nelson-Aalen type estimate:

\[
\hat{A}_0(t, \beta) = \sum_{ij} \int_0^t \frac{dN_{ij}^*(s, \beta)}{\sum_{kl} Y_{kl}^*(t, \beta)}.
\]

Note the analogy with the Nelson-Aalen estimate of the cumulative hazard for i.i.d. survival data (3), only now on the transformed scale. We get

\[
U(\beta) = \sum_{ij} \int_0^T \left( X_i^*(t^-, \beta) - \frac{\sum_{kl} X_k^*(t^-, \beta) Y_{kl}^*(t, \beta)}{\sum_{kl} Y_{kl}^*(t, \beta)} \right) dN_{ij}^*(t, \beta).
\]

Because the score is not continuous in \( \beta \), we obtain the parameter estimates by minimizing \( \|U(\beta)\| \). The asymptotic properties are obtained in a similar manner as with the Cox model, using the time transformed processes

\[
M_i^*(t, \beta) = N_{ij}^*(t, \beta) - \int_0^t Y_{ij}^*(s, \beta) dA_0(s)
\]

having zero mean for all \( t \) and the score taking the form

\[
U(\beta) = \sum_i \int_0^T \left( X_i^*(t^-, \beta) - \frac{\sum_{kl} X_k^*(t^-, \beta) Y_{kl}^*(t, \beta)}{\sum_{kl} Y_{kl}^*(t, \beta)} \right) dM_i^*(t, \beta).
\]

Again we obtain the asymptotic normality of \( \sqrt{n}(\hat{\beta} - \beta_0) \) with the help of the martingale central limit theorem. The covariance function of the latter depends on unknown functions \( \alpha_0 \) and \( \alpha'_0 \) and therefore cannot be estimated easily. Define

\[
S_0^*(t, \beta) = \sum_{ij} Y_{ij}^*(t), \quad S_1^*(t, \beta) = \sum_{ij} X_i^*(t^-) Y_{ij}^*(t), \quad E^*(t, \beta) = \frac{S_1^*(t, \beta)}{S_0^*(t, \beta)},
\]

\[
S_2^*(t, \beta) = \sum_{ij} X_i^*(t^-)^2 Y_{ij}^*(t), \quad S_3^*(t, \beta) = \sum_{ij} (X_i^*(t^-) X_i^*(t^-)^T)^2 Y_{ij}^*(t)
\]
Theorem 20.
Suppose that:
(a) The AFT model for repairable systems (25) holds, with
\((N_{i\bullet}(t), M_{i\bullet}, Y_{i\bullet}(t), Z_i(t)), i = 1, \ldots, n, (\text{i.i.d.})\).
(b1) The preventive maintenance distribution is independent on the time-to-
failure distribution of \(T_{ij}\) given \(H(T_{j-1})\).
(b2) \(E[M_{i\bullet}(\tau)] < \infty \forall i\)
(b3) \(Z_i(t)\) have uniformly bounded variation.
(c1) The baseline \(\alpha_0(t)\) is finite and continuous on \([0, \tau]\).
(c2) There exists such a neighborhood \(B\) of \(\beta_0\) that the processes
\(S_0^*, S_1^*, S_2^*, S_3^*, S_4\) and \(E^*\) converge almost surely to finite continuous functions \(s_0, s_1, s_2, s_4\) and \(e\) uniformly on \([0, \tau] \times B\).
(d1) There exists a positive definite \(p \times p\) matrix \(\Sigma(t)\) of functions continuous
on \([0, \tau]\), such that
\[
\frac{1}{n} \sum_{i=1}^{n} \int_0^t \left( \frac{S_2^*(s, \beta_0) - E^*(s, \beta_0)^{\otimes 2}}{S_0^*(s, \beta_0)} \right) S_0^*(s, \beta_0) dA_0(s) \stackrel{P}{\to} \Sigma(t)
\]
for each \(t \in [0, \tau]\) as \(n \to \infty\).
(d2)
\[
\frac{1}{n} \sum_{i=1}^{n} \int_0^\tau (X_i^*(s, \beta_0) - E^*(s, \beta_0)) Q_i(s, \beta) dF_0(s) \stackrel{P}{\to} A
\]
as \(n \to \infty\).
Then for \(\Sigma = \Sigma(\tau)\):
\[
\frac{1}{\sqrt{n}} U(\beta_0) \stackrel{D}{\to} N(0, \Sigma),
\]
\[
\sqrt{n}(\hat{\beta} - \beta_0) \stackrel{D}{\to} N(0, A\Sigma^{-1}A).
\]
as \(n \to \infty\).

Proof. We use the Rebolledo’s martingale central limit theorem (Theorem 3) with
\(H_i(t) = \frac{1}{\sqrt{n}}(X_i^*(t, \beta_0) - E^*(t, \beta_0))\) and \(M_i^*(t)\). Under the assumptions of the AFT model (25), the processes \(M_i^*(t)\) are square integrable \(H(t)\)-martingales. The time transformation \(h_i(t, \beta_0)\) is strictly monotone and continuous and \(H_i(t)\) contain the covariates evaluated at \(t^-\) and left-continuous processes \(Y_{i\bullet}(t)\). Therefore \(H_i(t)\) themselves are left-continuous, \(H(t)\)-adapted and predictable. Further, using (d1)
\[
\sum_{i=1}^{n} \int_0^t H_i(s) H_i(s)^T d\Lambda_i(s) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left( X_i^*(s^-) - E^*(s, \beta) \right)^{\otimes 2} Y_{i\bullet}(s) dA_0(s)
\]
81
Due to SLNN, the convergence occurs if \( \text{EI} \) converges to zero because

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} H_i(s)H_i(s)^T I_{\|H_i(s)\|\geq\epsilon} d\Lambda_i(s)
\]

Using the inequality Hölder’s inequality

\[
\left( \sum_{i=1}^{n} I_{\|x_i^*(s^-)\|\geq\epsilon} Y_i^*(s) dA_0(s) \right)^{\frac{2}{3}} \leq \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} X_i^*(s^-)^{\otimes 2} I_{\|x_i^*(s^-)\|\geq\epsilon} Y_i^*(s) dA_0(s)
\]

The second part converges to zero because \( \frac{1}{n} S_0 \) and \( E^* \) converge uniformly to finite functions and \( \int_{0}^{t} dA_0(t) < \infty \). For the first part, we get using the Hölder’s inequality

\[
\frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( X_i^*(s^-) - E^*(s, \beta) \right)^{\otimes 2} I_{\|x_i^*(s^-) - E^*(s, \beta)\|\geq\epsilon} Y_i^*(s) dA_0(s)
\]

For the Lindeberg condition we have

\[
\sum_{i=1}^{n} \int_{0}^{t} H_i(s)H_i(s)^T I_{\|H_i(s)\|\geq\epsilon} d\Lambda_i(s)
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( X_i^*(s^-) - E^*(s, \beta) \right)^{\otimes 2} I_{\|x_i^*(s^-) - E^*(s, \beta)\|\geq\epsilon} Y_i^*(s) dA_0(s)
\]

The second part converges to zero because \( \frac{1}{n} S_0 \) and \( E^* \) converge uniformly to finite functions and \( \int_{0}^{t} dA_0(t) < \infty \). For the first part, we get using the Hölder’s inequality

The first part converges to a finite function because of the convergence of \( S_4 \).

The second part converges to zero if \( \frac{1}{n} \sum_{i=1}^{n} I_{\|x_i^*(s^-)\|\geq\epsilon} \) converges to zero.

Due to SLNN, the convergence occurs if \( EI_{\|x_i^*(s^-)\|\geq\epsilon} \rightarrow 0 \) uniformly on \([0, \tau]\). Because \( X_i^*(t^-) = (N_i^*(t^-), M_i^*(t^-), Z_i^*(t^-)^T) \) and \( N_i^*(t) \) and \( M_i^*(t) \) are nondecreasing in \( t \), we have

\[
\sup_{t} E \left[ \frac{1}{\sqrt{n}} \|X_i^*(s^-)\| \geq \epsilon \right] = \sup_{t} P(\frac{1}{\sqrt{n}} \|X_i^*(s^-)\| \geq \epsilon) \leq \sup_{t} \frac{E \|X_i^*(s^-)\|}{\sqrt{n} \epsilon}
\]

\[
\leq \frac{EN_i^*(\tau)}{\sqrt{n} \epsilon} + \frac{EM_i^*(\tau)}{\sqrt{n} \epsilon} + \sup_{t} \frac{E \|Z_i^*(s^-)\|}{\sqrt{n} \epsilon}.
\]
As \( \text{EN}_i^*(\tau) = E \int_0^\tau Y_i^*(s)dA_0(s) < \infty \) due to (c1), \( \text{EM}_i^*(\tau) < \infty \) due to (b2) and \( Z_i(t) \) have bounded variation, the numerator terms are finite and the theorem holds, giving the asymptotics of the score.

Consistency of \( \hat{\beta} \) follows using the analogy of Theorem 1 by Lin and Ying (1995), stating that under given assumptions, there exists a function \( u(\beta) \), such that for any \( B > 0 \) \( d_n \to \infty \)

\[
\sup_{\|\beta\| \leq B} \|U(\beta) - u(\beta)\| = o_P(n^{1/2} + \epsilon),
\]

meaning that the score can be approximated by a non-random function. By Ying (1993) it follows that \( u(\beta) \) has a unique root \( \beta_0 \). Given the uniform convergence, the root of the score thus necessarily also converges to \( \beta_0 \), giving the consistency of \( \hat{\beta} \). The convergence of \( \sqrt{n}(\hat{\beta} - \beta_0) \) is obtained using the Taylor expansion and necessitating the use of the original score with the functions \( Q_i \).

\[\Box\]

### 4.3.3 Properties of the cumulative baseline estimates

For each of both semiparametric models we consider the process

\[ W(t) = \sqrt{n}(\hat{A}_0(t, \hat{\beta}) - A_0(t)). \]

It is possible to prove that \( W(t) \) converges weakly to a zero mean Gaussian process with a finite covariance function for \( n \to \infty \), ensuring the consistency of the Nelson-Aalen estimate. The covariance function differs between the models. In the AFT model, it depends on unknown \( \alpha_0 \) and \( \alpha'_0 \) and it is not possible to estimate it easily. In this section we present a resampling approach. Consider first the Cox model. We generate \( G_1, ..., G_n \) i.i.d. from \( N(0,1) \). Let

\[
U_G(\beta) = \sum_{ij} \int_0^T (X_i(t^-) - E(t, \beta)) G_idN_{ij}(t).
\]

Find \( \hat{\beta}_G \) as the solution of the equation \( U(\hat{\beta}_G) = U_G(\hat{\beta}) \) and let

\[
\hat{W}(t) = n^{1/2} \left( \hat{A}_0(t, \hat{\beta}) - \hat{A}_0(t, \hat{\beta}_G) + \hat{A}_{\text{OC}}(t, \hat{\beta}) \right),
\]

with

\[
\hat{A}_{\text{OC}}(t, \beta) = \sum_i \int_0^t \frac{G_idM_i(s)}{S_0(s, \beta)}
\]
\[ \hat{M}_i = N_i(t) - \int_0^t Y_{is}(s) d\hat{A}_0(s). \]

Using the martingale CLT it is possible to prove that under the Cox model \( \hat{W}(t) \) converges to the same Gaussian process as \( W(t) \).

**Theorem 21.**

Under the assumptions of Theorem 17, the processes \( W(t) \) and \( \hat{W}(t) \) under the Cox model converge weakly on \([0, \tau]\) to the same zero mean Gaussian process with a finite variance.

**Proof.** The proof is analogous to that of Andersen and Gill (1982) and Lin et al. (2000). First consider

\[
W(t) = \sqrt{n}(\hat{A}_0(t, \beta) - \hat{A}_0(t, \beta_0)) + \sqrt{n}(\hat{A}_0(t, \beta_0) - A_0(t))
\]

\[
= \sqrt{n} \sum_{ij} \int_0^t \left( \frac{1}{S_0(s, \beta)} - \frac{1}{S_0(s, \beta_0)} \right) dN_{ij}(s)
\]

\[
+ \sqrt{n} \sum_{ij} \int_0^t \frac{dN_{ij}(s)}{S_0(s, \beta_0)} - \int_0^t \frac{S_0(s, \beta_0)}{S_0(s, \beta_0)} dA_0(s)
\]

Using the Taylor expansion for the first term we see that

\[
\sqrt{n} \sum_{ij} \int_0^t \left( \frac{1}{S_0(s, \beta)} - \frac{1}{S_0(s, \beta_0)} \right) dN_{ij}(s) = -b(t, \beta_0) \sqrt{n}(\beta - \beta_0) + o_P(1)
\]

with \( b(t, \beta) = \int_0^t e(s, \beta) dA_0(s) \). Therefore

\[
W(t) = -b(t, \beta_0) \sqrt{n}(\beta - \beta_0) + \sqrt{n} \sum_i \int_0^t \frac{dM_i(s, \beta_0)}{S_0(s, \beta_0)} + o_P(1)
\]

\[
= -b(t, \beta_0) \Sigma^{-1} \frac{1}{\sqrt{n}} U(\beta_0) + \sqrt{n} \sum_i \int_0^t \frac{dM_i(s, \beta_0)}{S_0(s, \beta_0)} + o_P(1).
\]

By the arguments of Andersen and Gill (1982) it follows that the two terms are mutually independent. Thus using the Rebolledo’s central limit theorem on both parts we get that \( W(t) \) converges to a zero mean Gaussian process with the variance function \( \sigma(t_1, t_2) = E(\varsigma(t_1)\varsigma(t_2)) \) with

\[
\varsigma(t) = \int_0^t \frac{dM_i(s)}{S_0(s, \beta_0)} - b^T(t, \beta_0) \Sigma^{-1} U(\beta_0).
\]
For the replicated process we have

\[ \hat{W}(t) = \sqrt{n}(\hat{A}_0(t, \hat{\beta}) - \hat{A}_0(t, \beta_0)) - \sqrt{n}(\hat{A}_0(t, \hat{\beta}_G) - \hat{A}_0(t, \beta_G)) + \sum_i \int_0^t G_i \frac{d\hat{M}_i(s)}{S_0(s, \beta)}. \]

As \( \hat{\beta}_G \to \beta_0 \), using twice the same Taylor expansion as above we get that the first two parts equal to

\[ \hat{W}_1(t) = -b(\beta_0)\sqrt{n}(\hat{\beta} - \hat{\beta}_G) + o_P(1) = -b(\beta_0)\frac{1}{\sqrt{n}} U(\hat{\beta}_G) + o_P(1) \]

\[ = -b(\beta_0)\frac{1}{\sqrt{n}} U_G(\hat{\beta}) + o_P(1). \]

Because \( \hat{M}_i(t, \hat{\beta}) \) are not martingales, instead of the Rebolledo’s theorem we use the functional central limit theorem of Pollard (1990). Conditional on the data \((N_i, M_i, Y_i, Z_i(t))\), using multivariate central limit theorem, \( \hat{W}(t) \) converges in finite-dimensional distributions to a zero mean Gaussian process with a covariance function which in turn converges to \( \sigma(t_1, t_2) \). Tightness follows from the fact that \( \hat{M}_i(t, \beta_0), Z_i(t)\hat{M}_i(t, \beta_0), N_i(t^-)\hat{M}_i(t, \beta_0) \) and \( M_i(t^-)\hat{M}_i(t, \beta_0) \) consist of monotone functions and therefore are manageable in sense of Pollard (1990), p.38. The modern empirical process theory can thus be applied similarly as in the proof of Theorem 9.

For the Accelerated failure time model we proceed similarly and work on the time-transformed scale. We take replicates of the approximate score

\[ \tilde{U}_G(\beta) = \sum_{ij} \int_0^\tau (X_i^*(t^-) - E^*(t)) G_i dN_i^*(t) \]

find \( \hat{\beta}_G \) as the solution of \( \tilde{U}(\hat{\beta}_G) = \tilde{U}_G(\beta) \) and let

\[ \hat{A}_{0G}(t, \beta) = \sum_i \int_0^t G_i \frac{d\tilde{M}_i^*(s, \beta)}{S_0^*(s, \beta)} \]

where

\[ \tilde{M}_i^*(t, \beta) = N_i^*(t, \beta) - \int_0^t Y_i^*(s, \beta) d\hat{A}_0(t, \beta). \]

**Theorem 22.**

Under the assumptions of Theorem 20, \( W(t) = \sqrt{n}(\hat{A}_0(t) - A_0(t)) \) estimated under the AFT model converges weakly on \([0, \tau]\) to a zero-mean Gaussian process with a finite variance. The process

\[ \hat{W}(t) = \sqrt{n}(\hat{A}(t, \hat{\beta}) - \hat{A}(t, \hat{\beta}_G) + \hat{A}_{0G}(t, \hat{\beta})) \]
has the same limit.

Proof. We proceed similarly as above, using the time-transformed processes of Lin et al. (1998)

\[ W(t) = \sqrt{n}(\hat{A}_0(t, \hat{\beta}) - \hat{A}_0(t, \beta_0)) + \sqrt{n}(\hat{A}_0(t, \beta_0) - A_0(t)) \]

\[ = -b(t, \beta_0)A^{-1} \frac{1}{\sqrt{n}} \tilde{U}(\beta_0) + \sqrt{n} \sum_i \int_0^t \frac{dM_i^*(s, \beta_0)}{S_0^*(s, \beta_0)} + o_P(1). \]

Using the Taylor expansion

\[ \sqrt{n} \int_0^t (\hat{A}_0(\hat{\beta}) - \hat{A}_0(\beta_0)) = -b(t, \beta_0)\sqrt{n}(\hat{\beta} - \beta_0) + o_P(1) \]

with \( b(t, \beta) = \int_0^t e^*(s, \beta)d(\alpha'_0(s)s). \) The two parts are again independent. Thus using the Rebolledo’s central limit theorem we get that \( W(t) \) converges to a zero mean Gaussian process with the variance function \( \sigma(t_1, t_2) = E(\varsigma(t_1)\varsigma(t_2)) \) with

\[ \varsigma(t) = \int_0^t dM_i^*(s) - b^T(t, \beta_0)A^{-1}\tilde{U}(\beta_0) \]

For the replicated process we have

\[ \hat{W}(t) = n^{1/2} \left( \hat{A}_0(t, \hat{\beta}) - \hat{A}_0(t, \beta_0) + \sum_i \int_0^t G_i d\tilde{M}_i^*(s) \right). \]

\[ = -b(\beta_0)A \frac{1}{\sqrt{n}} U_G(\hat{\beta}) + \sqrt{n} \sum_i \int_0^t G_i d\tilde{M}_i^*(s) + o_P(1). \]

The convergence of finite-dimensional distributions to a zero mean Gaussian process with variance function \( \sigma(t_1, t_2) \) follow from the multivariate CLT. Tightness is established similarly as in Theorem using the empirical process theory and manageability of \( M_i^*(t, \beta_0), Z_i^*(t^-)M_i^*(t, \beta_0), N_i^*(t^-)M_i^*(t, \beta_0) \) and \( M_i^*(t^-)M_i^*(t, \beta_0) \).

\[ \square \]

4.3.4 Testing hypotheses on the baseline hazard shape

When we replicate \( \hat{W}(t) \) many times, we can empirically estimate the variance of \( W(t) \) and compute the pointwise confidence intervals for the cumulative baseline hazard as

\[ \hat{A}_0(t, \hat{\beta}) \pm u_{1-\alpha/2}n^{-1/2}\sqrt{\text{var}W(t)}. \]
or using the log-transformation as

\[ \hat{A}_0(t, \hat{\beta}) \exp \left( \pm u_{1-\alpha/2} n^{-\frac{1}{2}} \frac{\sqrt{\text{var}(W(t))}}{\hat{A}_0(t, \hat{\beta})} \right), \]

where \( u_{1-\alpha/2} \) is the respective quantile of \( N(0,1) \).

If we want to test hypotheses on the shape of the whole hazard, it is necessary to find the respective confidence band for a supremum test. We take \( q_{1-\alpha} \) as the sample \( 1 - \alpha \) quantile of the generated values sup\( [\tau_1, \tau_2] \left| \frac{\hat{W}(t)}{\sqrt{\text{var}(W(t))}} \right| \) where \([\tau_1, \tau_2]\) covers the examined part of the time interval. We can then compute the confidence band using the log-transformation as

\[ \hat{A}_0(t, \hat{\beta}) \exp \left( \pm q_{1-\alpha} n^{-1/2} \sqrt{\text{var}(W(t))/\hat{A}_0(t, \hat{\beta})} \right). \]

We reject the hypothesis on level \( \alpha \), if the tested cumulative baseline hazard does not lie in the confidence band. On Fig. 8 we see an example of the Nelson-Aalen estimate with the respective pointwise confidence intervals, the confidence band and the parametric estimates for exponential, Weibull, gamma, lognormal and truncated Gumbel distributions. In this case, the Weibull baseline fits the data best. The test would reject the exponential and Gumbel distributions, as they are well outside of the confidence band at the beginning of the observed interval.

### 4.4 Conclusion

We explored methods for modeling the influence of maintenance and repairs on the lifetimes of the observed devices. In the Cox model the covariates representing the count and extent of repairs and maintenance actions influence the hazard function multiplicatively, whereas in the AFT model they accelerate or decelerate the flow of the internal time of the device. When we parametrize the baseline hazard function, the service record of one device is enough to obtain the estimates of the regression parameters. If we have data on more devices, it is possible to estimate the cumulative baseline hazard function nonparametrically. For both models we introduced an asymptotic test on the shape of the cumulative baseline hazard based on resampling. Further
research could concern developing goodness-of-fit tests, exploring other transformations in the Accelerated failure time model and taking into account the variability of parametric distributions in the testing procedure.

Figure 8: Comparison of the Nelson-Aalen estimate and the parametric estimates of the cumulative baseline hazard.
5. Inference for repairable systems - examples and properties

In this chapter we present the applications of the methods for dealing with repairable systems. First, we study the properties of the testing procedure for the baseline hazard, using simulated data with various distributions and settings and compare the results. Finally, we deal with real data from oil industry, studying the service records of oil pumps. We estimate the influence of repairs and preventive maintenance using both parametric and semiparametric approaches from sections 4.2 and 4.3.

5.1 Properties of the test on the baseline hazard

We generated data sets from both the Cox and the AFT model of size $n = 20$ and $n = 50$ with various hazard functions and parameters. Each device was followed until the 10th event, the preventive maintenance was undertaken randomly with the same baseline distribution. The parameters $\rho$ and $\sigma$ were taken so that each repair increased the hazard or accelerated the flow of the internal time ($\rho = 0.1$) and each maintenance had an opposite effect ($\rho = -0.1$).

We tested on the level of significance $\alpha = 0.05$, whether the baseline distribution is exponential, Weibull with $\alpha_0(t) = a\lambda t^{a-1}$, Gamma, truncated Gumbel with $\alpha_0(t) = \lambda a'$ or lognormal with the maximum likelihood parameter estimates of the original model considered fixed and we observed the proportion of the rejected hypotheses. Each case was generated 200×, we tested on the interval between the 5% and 95% quantiles of the generated data. $\hat{W}(t)$ was computed using 100 replications.

On table we can see, that the tests are more accurate with larger number of observed devices, meaning that they do not reject the original and reject the other baseline distributions. When observing data with Weibull of Gamma...
<table>
<thead>
<tr>
<th>Generated distribution</th>
<th>Tested distribution - proportion of rejected samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_0$ $\lambda$ $a$ $n$ Exp. Weibull Gamma Gumbel LN</td>
</tr>
<tr>
<td>Weibull</td>
<td>1/10 5 20 0.895 0 0.005 0 0.005</td>
</tr>
<tr>
<td>Cox</td>
<td>50 1 0 0.030 0 0.200</td>
</tr>
<tr>
<td>Weibull</td>
<td>1/10 1/2 20 0.960 0 0.010 0.940 0.010</td>
</tr>
<tr>
<td>Cox</td>
<td>50 1 0 0.030 1 0.365</td>
</tr>
<tr>
<td>Gamma</td>
<td>1/10 5 20 0.210 0 0 0.050 0</td>
</tr>
<tr>
<td>Cox</td>
<td>50 0.930 0 0 0.530 0.014</td>
</tr>
<tr>
<td>Gamma</td>
<td>1/10 1/2 20 0.415 0 0 0.365 0.060</td>
</tr>
<tr>
<td>Cox</td>
<td>50 0.830 0 0 0.765 0.635</td>
</tr>
<tr>
<td>Gumbel</td>
<td>1/10 1.2 20 0.065 0 0.005 0 0.105</td>
</tr>
<tr>
<td>Cox</td>
<td>50 0.685 0 0.200 0 0.705</td>
</tr>
<tr>
<td>LN $\mu=2$ $\sigma=2$</td>
<td>50 1 0.005 0.415 0.995 0</td>
</tr>
<tr>
<td>Cox</td>
<td>50 1 0.030 0.805 1 0</td>
</tr>
<tr>
<td>Weibull</td>
<td>1/10 5 20 0.910 0.010 0.030 0.010 0.065</td>
</tr>
<tr>
<td>AFT</td>
<td>50 1 0 0.210 0 0.490</td>
</tr>
<tr>
<td>Weibull</td>
<td>1/10 1/2 20 0.905 0.005 0.020 0.790 0.090</td>
</tr>
<tr>
<td>AFT</td>
<td>50 1 0 0.120 1 0.645</td>
</tr>
<tr>
<td>Gamma</td>
<td>1/10 5 20 0.915 0.005 0.030 0.010 0.070</td>
</tr>
<tr>
<td>AFT</td>
<td>50 1 0 0.210 0 0.490</td>
</tr>
<tr>
<td>Gamma</td>
<td>1/10 1/2 20 0.900 0.005 0.020 0.795 0.090</td>
</tr>
<tr>
<td>AFT</td>
<td>50 1 0 0.101 1 0.645</td>
</tr>
<tr>
<td>Gumbel</td>
<td>1/10 1.2 20 0.375 0.015 0.095 0.010 0.590</td>
</tr>
<tr>
<td>AFT</td>
<td>50 0.990 0.050 0.695 0 0.995</td>
</tr>
<tr>
<td>LN $\mu=2$ $\sigma=2$</td>
<td>50 0.995 0.480 0.140 0.880 0.090</td>
</tr>
<tr>
<td>AFT</td>
<td>50 1 0.020 0.910 1 0</td>
</tr>
</tbody>
</table>

Table 10: The proportion of rejected hypotheses on the shape of the baseline hazard when generating data from various distributions.

baseline distribution, depending whether $A_0$ is concave or convex, the behavior is sometimes indistinguishable from that of either the Gumbel or Lognormal distribution. The exponential distribution is rejected in large number of cases - providing the analogy with testing whether $a = 1$ in Weibull or Gamma distribution.

5.2 Modeling the lifetime of oil pumps

We explore data on the service of oil pumps during several years (Kobbacy et al., 1997 and Percy and Alkali, 2007). For one device we have detailed data on $m_1 = 65$ times of repairs, maintenance actions and the cost of each action in man-hours. This data has been studied by Percy and Alkali (2005) using the
parametric Cox model. We try to model the lifetime using both the Cox and the AFT model as shown above with various parametrized baseline hazard functions and compare the results. In the parametric case, it is possible to directly maximize the likelihood for all cases and see in which it was largest.

For four other pumps we have only the times of actions at disposal, with \((m_2, \ldots, m_5) = (51, 90, 30, 30)\). We use both the semiparametric methods and parametrized baseline hazards with the two described models to estimate the regression parameters utilizing data of all the five pumps. The likelihood in semiparametric methods depends on the unknown baseline hazard and therefore is not available for comparison of the used methods.

The data is given as the time elapsed between each of the successive actions. We assume that the duration of each action does not contribute to the total time elapsed, because the pumps are inoperable and are not under workload at that time. It can still be argued that the devices do age even during a repair or a maintenance action, but that could possibly lead to failures occurring at that time, necessitating further repairs, which would require more complex models.

### 5.2.1 Parametric modeling of the service of one pump

We have the times of repairs, maintenance actions and cost of each action for one pump. Using methods from section 4.2 we estimate the parameters \(\rho\), \(\sigma\) and \(\beta\) in both the Cox model and the AFT model. We try to maximize the likelihood for exponential, Weibull \(\alpha_0(t) = a \lambda e^{a(t-1)}\), gamma \(f(t) \propto t^{a-1}e^{-\lambda t}\), truncated Gumbel \(\alpha_0(t) = \lambda a^t\) and lognormal baseline distributions.

Comparing the likelihood values in Table 11 we find that it is highest for both the Cox and AFT model with the truncated Gumbel distribution. Further we see that the more each action did cost, the more it increased the hazard function or accelerated the internal time, because \(e^{\hat{\beta}} > 1\). Each man-hour of the action means an increase of hazard or acceleration of time by about 0.5–0.7%. A repair itself has a positive influence \((e^{\hat{\rho}} < 1)\), with the exception of the AFT model with lognormal baseline distribution, but that is the case with the lowest likelihood value. It is interesting that according to all cases except the Gumbel distribution in Cox model, the maintenance actions tend to have a negative influence \((e^{\hat{\rho}} > 1)\). This could be due to repairs often taking much more man-hours than maintenances (on average, a repair took
The log-likelihood and parameter estimates from parametric models of the lifetime of one oil pump.

26.8 whereas a maintenance action took only 9.4 man-hours), resulting in negative influence of both.

5.2.2 Semiparametric modeling of the lifetime of five pumps

For five devices we have only the times of repairs and maintenance available. The data on the cost of the actions was not available for all pumps, therefore we estimate only the regression parameters $\rho$ and $\sigma$. We tried the Cox and the AFT models, both semiparametric and parametric with the same baseline distributions as above. In the parametric cases we maximize the log-likelihood whereas in the semiparametric approach we insert the estimate of the cumulative baseline hazard into the score function and solve the equations $U(\beta) = 0$ for the Cox model and minimize $\|U(\beta)\|$ for the AFT model. We also perform a test whether a parametric baseline hazard fits in the confidence bands of the respective semiparametric model, using 200 resampled processes.

In Table 12 we see that in all cases each repair increases the hazard or accelerates the internal time ($e^\hat{\sigma} > 1$). Among the parametric models, the Gumbel distribution with AFT model has the highest likelihood. In that case and also in the cases with lognormal baseline hazard and the semiparametric models, the maintenance actions have also a negative influence, whereas in the other

<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha_0$</th>
<th>log - lik</th>
<th>$e^\rho$</th>
<th>$e^{\hat{\sigma}}$</th>
<th>$e^{\hat{\beta}}$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>-213.8</td>
<td>1.407</td>
<td>0.920</td>
<td>1.0065</td>
<td>0.0015</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cox Weibull</td>
<td>-213.5</td>
<td>1.266</td>
<td>0.924</td>
<td>1.0062</td>
<td>0.0017</td>
<td>1.671</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>-213.8</td>
<td>1.405</td>
<td>0.918</td>
<td>1.0065</td>
<td>0.0016</td>
<td>1.027</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>-210.2</td>
<td>0.701</td>
<td>0.745</td>
<td>1.0061</td>
<td>0.0006</td>
<td>1.010</td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>-214.8</td>
<td>1.541</td>
<td>0.913</td>
<td>1.0067</td>
<td>$\hat{\mu}$=6.3</td>
<td>$\hat{\sigma}$=1.66</td>
<td></td>
</tr>
<tr>
<td>AFT Weibull</td>
<td>-212.7</td>
<td>1.278</td>
<td>0.918</td>
<td>1.0060</td>
<td>0.0011</td>
<td>1.639</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>-213.8</td>
<td>1.418</td>
<td>0.916</td>
<td>1.0065</td>
<td>0.0014</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>-210.2</td>
<td>1.237</td>
<td>0.892</td>
<td>1.0051</td>
<td>0.0006</td>
<td>1.001</td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>-218.2</td>
<td>1.301</td>
<td>1.050</td>
<td>1.0069</td>
<td>$\hat{\mu}$=5.25</td>
<td>$\hat{\sigma}$=0.89</td>
<td></td>
</tr>
</tbody>
</table>

Table 11: The log-likelihood and parameter estimates from parametric models of the lifetime of one oil pump.
<table>
<thead>
<tr>
<th>Model</th>
<th>$\alpha_0$</th>
<th>log - lik</th>
<th>$e^\hat{\rho}$</th>
<th>$e^\hat{\sigma}$</th>
<th>$\hat{\lambda}$</th>
<th>$\hat{a}$</th>
<th>pval-$A_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp.</td>
<td>-880.3</td>
<td>0.985</td>
<td>1.016</td>
<td>0.016</td>
<td>$-$</td>
<td>$-$</td>
<td>0.655</td>
</tr>
<tr>
<td>Cox Weibull</td>
<td>-880.2</td>
<td>0.976</td>
<td>1.016</td>
<td>0.014</td>
<td>1.063</td>
<td>0.680</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>-880.1</td>
<td>0.988</td>
<td>1.016</td>
<td>0.015</td>
<td>0.811</td>
<td>0.575</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>-880.3</td>
<td>0.994</td>
<td>1.016</td>
<td>0.016</td>
<td>0.999</td>
<td>0.710</td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>-894.4</td>
<td>1.090</td>
<td>1.016</td>
<td>$\hat{\mu}=3.22$</td>
<td>$\hat{\sigma}=0.89$</td>
<td>0.540</td>
<td></td>
</tr>
<tr>
<td>AFT Weibull</td>
<td>-880.2</td>
<td>0.980</td>
<td>1.015</td>
<td>0.014</td>
<td>1.038</td>
<td>0.795</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>-880.1</td>
<td>0.988</td>
<td>1.016</td>
<td>0.015</td>
<td>0.812</td>
<td>0.635</td>
<td></td>
</tr>
<tr>
<td>Gumbel</td>
<td>$-875.1$</td>
<td>1.022</td>
<td>1.036</td>
<td>0.013</td>
<td>0.999</td>
<td>0.450</td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>-879.5</td>
<td>1.284</td>
<td>1.158</td>
<td>$\hat{\mu}=2.67$</td>
<td>$\hat{\sigma}=1.56$</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td>Cox nonparam.</td>
<td>$-$</td>
<td>1.065</td>
<td>1.023</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td></td>
</tr>
<tr>
<td>AFT nonparam.</td>
<td>$-$</td>
<td>1.061</td>
<td>1.164</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td></td>
</tr>
</tbody>
</table>

Table 12: The log-likelihood and parameter estimates from modeling the lifetime of five pumps.

cases it is positive. The only baseline hazard rejected by the test on $\alpha = 0.05$ was the lognormal case in the AFT model. As there are only five observed devices, the test likely does not distinguish between the distributions more precisely because of its asymptotic nature.

5.2.3 Model selection and validation methods

When comparing two models, it is possible to perform a $\chi^2$ test based on scaled deviance $D = 2 \cdot (l_1 - l_2)$. Based on the asymptotics established by Theorems 15 and 16 for the parametric models using the Taylor expansion of the log-likelihood (19) we obtain that $2 \left( l(\hat{\beta}) - l(0, \hat{\beta}_{(2,p)}) \right) \rightarrow \chi^2_1$. This can be used for testing $\beta_1 = 0$ to see whether one particular covariate significantly improves a model or not. For the one pump data using Cox model with Gumbel baseline, no significant improvement was found when adding the influence of the cost of the action ($\beta$) to a model containing only the regression parameters $\rho$ and $\sigma$ (Percy and Alkali, 2005), while it can be argued that the covariate still adds some relevant information. It can be also tested, whether a Weibull or Gamma baseline hazard can be simplified to exponential with $a = 1$. This approach is, however, not usable when comparing models with different baseline hazards or the Cox model with the AFT model.
Aside from the likelihood value, we do not have direct means to determine which model fits the data best, especially when comparing the parametric and semiparametric approaches. For classic survival regression data, we provided an overview on the goodness-of-fit tests in chapter 2. This methods could be adapted to accommodate repairs and maintenance actions. They are however, based on resampling approach and asymptotic convergence of certain martingale processes, and therefore it remains to be seen how well they would perform in such cases as with above data representing only a few independent devices.
Conclusion

Survival analysis and reliability theory constitute an important part of mathematical statistics, with many applications mainly in medical and technical research. In this work we studied the regression models available for dealing with survival and reliability data, allowing for estimating the effect of covariates on the distribution of the time to failure.

We focused on the Cox proportional hazards model, Aalen additive model and mainly on the Accelerated failure time model and its extensions. The AFT model assumes that each studied subject is having the flow of its internal time accelerated or decelerated depending on the influencing factors. Therefore it has many applications and a straightforward interpretation. In Chapter 2, we presented the respective inference for the models and also studied possible methods for testing their goodness-of-fit, allowing to see which model and covariates provide the best interpretation of the data. As one of the main contributions of this work we developed a goodness-of-fit testing procedure based on martingale residuals and resampling techniques for the AFT model, providing an addition to existing model checking procedures for the Cox and Aalen models. In Chapter 3, the properties of the proposed test were examined and its performance was compared with the existing procedures. We explored the empirical level of significance and empirical power of the tests against various alternatives, giving possible information on the behavior of the methods in real applications.

In Chapter 4, as a further contribution, we extended the ways for dealing with repairable systems. We introduced new parametric and semiparametric models taking into account the service history of each device, incorporating the number of repairs and maintenance actions as time-varying covariates. We derived the large sample properties of the estimators based on the central limit theorems for martingale differences and continuous time martingales. Using fully parametric models, under certain conditions it is possible to establish the inference with data on only one device. Having data on more devices allows us to use semiparametric approach. We have also shown a method for testing the shape of the baseline hazard of the models based on martingale resampling. In Chapter 5, we studied the properties of introduced methods in various settings using simulation experiments and have shown the applications on real data from oil industry.
As a final contribution, presented methods have been implemented as procedures for the R language, making them available for use in real applications.

Further work may concern establishing goodness-of-fit testing procedures for the repair models. It would be also interesting to study and compare other possible time transformations for the Accelerated failure time model, both in survival setting as well as for repairable systems.
Appendix

Implementation of the AFT model in R

Using the methods studied in Chapter 2 we implemented estimation and goodness-of-fit testing procedures for the Accelerated failure time model with constant covariates

$$log(T_i^*) = -Z_i^T \beta + \epsilon_i$$

as well as for one important case of time-varying covariates

$$e^{\epsilon_i} = \int_0^{T_i^*} e^{Z_i(s)^T \beta} ds$$

for data with with one jump $Z_i(t) = I(t > s_i)$.

Usage:

```r
aft(times, status, z = NULL, s = NULL, wg = NULL,
    sim.test = T, n.sim = 100)
```

Arguments:

- `tx`: vector of observed times of events $T_i$
- `d`: vector of censoring indicators $\Delta_i$ (1=failure, 0=censoring)
- `z`: matrix or data frame of covariates to be fitted, either continuous or two-level factors
- `s`: jump times
- `wg`: vector of weights to be used in the test based on the martingale residuals
- `sim.test`: logical denoting whether the martingale residuals test should be performed
- `it`: number of resampled processes for the test.
Output: a list containing:

- `coef` a data.frame containing the estimates of the regression coefficients, their standard deviation, significance and p-values of the goodness-of-fit using t-test and Wilcoxon test
- `pval` p-values of the martingale residuals goodness-of-fit test, computed from its plain version, variance standardised version and both of them computed over quartiles of the data.

Example:

```r
## modeling the impact of a sudden shutdown during a strike
## on the lifetime of aluminium smelting cells

> data(strike)
> times=strike$failure;
> status=rep(1,nrow(strike));
> s=strike$strike
> aft(times,status,s,wg=I(s<=median(s)),
+     sim.test=TRUE,n.sim=100)

Fitting AFT model with a single jump.

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-value&lt;&gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td>I(t&gt;si)</td>
<td>0.502</td>
<td>0.069</td>
<td>0</td>
</tr>
</tbody>
</table>

Testing goodness-of-fit using martingale residuals:

<table>
<thead>
<tr>
<th></th>
<th>pval_sup</th>
<th>pval_sd</th>
<th>pval_qsup</th>
<th>pval_qsd</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value:</td>
<td>0.08</td>
<td>0.26</td>
<td>0.07</td>
<td>0.1</td>
</tr>
</tbody>
</table>
```
Implementation of the repair models

We implement the repair models using the methods introduced in Chapter 4 along with the estimation procedures. Both the parametric and semiparametric versions of the Cox and AFT models are encompassed in a single function.

Usage:

```r
regrepair(times, status, z = NULL, model, baseline = "semi",
           sim.test = FALSE, n.sim = 100)
```

Arguments:

- `times`: a matrix of observed times. Rows denote devices and columns the ordered times of actions performed. Fields after the last observed actions are coded as 0 or Inf.
- `status`: a matrix of censoring indicators of the actions performed on the devices, repair=1, maintenance=0.
- `z`: a three-dimensional array of possibly time-varying covariates to be fitted to the model. First dimension denotes the covariate, the other two its values at the observed times.
- `model`: character string of values "cox" or "aft" specifying which model should be fitted
- `baseline`: character string denoting the baseline distribution to be fitted: ex exponential, wb Weibull, ga Gamma, ln Lognormal, gu truncated Gumbel and semi denoting that the baseline should be estimated nonparametrically.
- `sim.test`: logical, denoting whether the test of the shape of the chosen baseline should be performed (only if baseline is not semi).
- `n.sim`: number of replications of the baseline hazard.
Output: a list containing

lik the log likelihood, if a parametric baseline was chosen.

coeff a data.frame containing the estimates of the regression coefficients for $M(t)$, $N(t)$ and further covariates.

pval_baseline p-value of the test whether the chosen parametric baseline fits into the confidence bands for the nonparametric estimate.

Example:

```r
## Modeling the lifetime distribution of five oil pumps
## depending on their service history

> data(pumpdata5)
> times=pumpdata5$times
> status=pumpdata5$status
> regrepair(times,status,model="Aft",baseline="gu",
> +    sim.test=TRUE,n.sim=100)

Fitting parametric AFT model for repairable systems with Gumbel baseline distribution

Maximum likelihood value:
-875.0853

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M(t)$</td>
<td>0.021</td>
</tr>
<tr>
<td>$N(t)$</td>
<td>0.035</td>
</tr>
<tr>
<td>lambda</td>
<td>0.013</td>
</tr>
<tr>
<td>a</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Testing the parametric shape of the baseline hazard among the semiparametric AFT model:

p-value:
0.45
Bibliography


