



Norwegian University of
Science and Technology

Residuals and Functional Form in Accelerated Life Regression Models

Stein Aaserud

Master of Science in Physics and Mathematics

Submission date: June 2011

Supervisor: Bo Henry Lindqvist, MATH

Problem Description

- Give an introduction to log-location-scale regression models and the corresponding Cox–Snell residuals.
- Study methods for retrieving functional form of misspecified covariates by using Cox–Snell residuals.
- Illustrate methods with both simulated and real data.

Assignment given: 24 January 2011,
Supervisor: Bo Henry Lindqvist.

Abstract

This thesis examines misspecified log-location-scale regression models. Particularly how the models' Cox–Snell residuals can be used to infer the functional form of possibly misspecified covariates in the regression. Two different methods are considered. One is using a transformation of the expected value of the residuals. The second is based on estimating the hazard rate function of the residuals using the covariate order method. Simulations and computations in the statistical computing environment **R** are used to obtain relevant and illustrative results. The conclusion is that both methods are able to recover the functional form of a misspecified covariate, but the covariate order method is best when high levels of censoring are introduced. The Kullback–Leibler theory, applied to misspecified regression models, is a part of the basis for the investigations. The thesis shows that a theoretical approach to this theory is consistent with the methods used in **R**.

Acknowledgements

This thesis is a continuation of the problems reviewed in the project “Inferring the functional form of covariates in misspecified survival regression models” [1]. I would like to show my gratitude to my supervisor Bo Henry Lindqvist for all the help he has given me, and all the constructive conversations we had throughout working with the thesis. Then I would like to thank Jan Terje Kvaløy for providing the R code for the covariate order method. Finally, special thanks go to my fellow students in reading hall 393C for motivational digressions and support in times of distress.

Trondheim, June 2011

Stein Aaserud

Contents

List of Figures	vii
List of Tables	ix
1 Introduction	1
2 Survival analysis	3
2.1 The exponential distribution	4
2.2 The Weibull distribution	4
2.3 The Gumbel distribution of the smallest extreme	5
2.4 The normal distribution	6
2.5 The lognormal distribution	6
2.6 Censoring	7
2.7 Cox–Snell residuals	7
2.8 Covariate order method	8
3 Model	9
4 Weibull distributed lifetimes	11
4.1 Coefficients minimizing the Kullback–Leibler distance	12
4.1.1 Using Cox–Snell residuals to infer the form of $f(x_i)$	14
4.2 Simulation and model fitting of Weibull distributed lifetimes	16
4.2.1 Uncensored data set	17
4.2.2 Censored data sets	19
5 Lognormally distributed lifetimes	41
5.1 Coefficients minimizing the Kullback–Leibler distance	42
5.1.1 Using Cox–Snell residuals to infer the form of $f(x_i)$	42
5.1.2 Uncensored data set	43
5.1.3 Censored data sets	46
6 Analytical Kullback–Leibler	61
6.1 Lognormal	62
6.2 Weibull	65
7 Investigation of a real data set	71
7.1 Distribution of the data	71
7.2 Covariate order method	74
8 Discussion and conclusions	81
A Simulated data set of Weibull distributed lifetimes with censoring	85
B Simulated data set of lognormally distributed lifetimes with censoring	89

C	Real data set with censoring	93
D	R-code	95
D.1	Covariate order method	95
D.2	Anderson-Darling trend test	97
D.3	Bootstrapping the Anderson-Darling statistics	97
D.4	Likelihood cross-validation	100

List of Figures

1	Illustration of left, right and interval censoring. The crosses illustrate when the failure occurs, while the short lines perpendicular to the time axis are the actual times observed.	7
2	Constructed artificial point process with $(Y_1, \delta_1 = 1), (Y_2, \delta_2 = 0), (Y_3, \delta_3 = 1), \dots, (Y_n, \delta_n = 1)$	8
3	Illustration of the Kullback–Leibler distance.	10
4	Plot of the logarithm of the Cox–Snell residuals, $\ln(\hat{R}_i^*)$, versus the covariates z_{i1}, z_{i2} and x_i (from left to right).	14
5	Plot of the estimation of the function $f(x_i)$ from (4.9).	15
6	Plot of the estimation of the function $f(x_i)$ from (4.10).	16
7	Plot of the estimation of the function $f(x_i)$ from (4.10) with 100 observations.	19
8	Plots of the covariates (z_1, z_2 , and x respectively from left to right) versus the logarithm of the Cox–Snell residuals from the misspecified models (x_i is modelled linearly, γx_i) for each of the censoring levels. The circles represent residuals from uncensored observations, while the dots represent residuals from censored observations.	22
9	Plot of the estimation of the function $f(x_i)$ with LOWESS (solid line) and quadratic regression (dashed line) estimates for the different censoring levels.	23
10	Plot of the covariates (z_1, z_2 , and x respectively from left to right) versus the logarithm of the Cox–Snell residuals from the respecified models in (4.16) (x_i is modelled quadratically, γx_i^2) for the 20% and 50% censoring levels. The circles represent residuals from uncensored observations, while the dots represent residuals from censored observations.	27
11	Plots checking if the Cox–Snell residuals are unit exponentially distributed. 20% censoring in the left column, 50% censoring in the right. (a) is a plot of the Cox–Snell residuals from the model (4.16) with theoretical mean (red dashed line) and estimated mean (blue solid line). (b) is a histogram of the Cox–Snell residuals with theoretical density function. (c) is the exponential probability plot. 1 is added to the censored residuals.	29
12	Figure illustrating the correspondance between the smoothing parameter on the covariate axis, x , and the event axis, S	33
13	Plots of the estimated function from (4.25) for each covariate (z_1, z_2 and x from left to right) for each censoring level (20%, 50% and 80% from top to bottom).	34
14	Plots of x_i^2 vs. $\ln(\hat{\lambda}(x_i^2))$ for 20%, 50% and 80% censoring (from left to right).	35
15	Density plots of the six bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.	39

16	Plots of the estimation of $f(x_i)$ using the transformation method, covariate order method and the real function. The transformation method is illustrated by the circles, the black solid LOWESS line and the red dashed line (fitted quadratic function). The covariate order method is the blue squares connected by the blue line. The real function is the black dash-dot line.	40
17	Plot of the estimation of the function $f(x_i)$ from (5.8).	43
18	Plots of the estimation of the function $f(x_i)$ from (5.8) for the complete data set of lognormally distributed lifetimes. The solid black line in (b) is the LOWESS, the red dashed line is a fitted quadratic function.	44
19	Plots comparing the estimation of the function $f(x_i)$ for Weibull and lognormally distributed lifetimes.	45
20	Plot of the logarithm of the Cox–Snell residuals versus x_i for censoring levels of 20%, 50% and 80% (left to right). Circles from uncensored observations, solid dots from censored observations.	48
21	Plot of the estimation of the function $f(x_i)$ with LOWESS (black solid line) and quadratic regression (red dashed line) estimates for the different censoring levels for lognormally distributed data.	49
22	Plot of the logarithm of the Cox–Snell residuals versus x_i for censoring levels of 20% and 50% from the respecified models in (5.12). Circles from uncensored observations, solid dots from censored observations.	52
23	Plots checking if the Cox–Snell residuals are unit exponentially distributed. 20% censoring in the left column, 50% censoring in the right. (a) is a plot of the Cox–Snell residuals from the models in (5.12) with theoretical mean (red dashed line) and estimated mean (blue solid line). (b) is a histogram of the Cox–Snell residuals with theoretical density function. (c) is the unit exponential probability plot. 1 is added to the censored residuals.	54
24	Estimation of the $f(x_i)$ using the covariate order method for censoring levels of 20%, 50% and 80%.	56
25	Plots of x_i^2 vs $\ln(\hat{\lambda}(x_i^2))$ for 20%, 50% and 80% censoring (from left to right).	57
26	Density plots of the six bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.	59
27	Plots of the estimation of $f(x_i)$ using the transformation method, covariate order method and the real function. The transformation method is illustrated by the circles, the black solid LOWESS line and the red dashed line (fitted quadratic function). The covariate order method is the blue squares connected by the blue line. The real function is the black dash-dot line.	60
28	(a) Scatter plot of the covariate x_i , $\ln(\text{pseudostress})$, versus the response Y_i , k-Cycles. Circles represent failures, while dots are censored events. (b) Histogram of the lifetimes, k-Cycles, from the real data set.	71

29	Probability plots from the model in (7.1).	73
30	(a) Plot of the covariate x plotted against the estimated log-hazard rate $\ln(\hat{\lambda}(x))$. Circles represent failures, while dots are censored events. (b) Plot of x against $f(x)$ from (7.5). Circles represent failures, dots are censored events.	75
31	The covariate x plotted against the estimated log-hazard rate $\ln(\hat{\lambda}(1/x))$ from the model in (7.6)	76
32	Density plots of the two bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.	77
33	Plots verifying that the Cox–Snell residuals truly are unit exponentially distributed. (a) is a plot of the Cox–Snell residuals from the model (7.6) with theoretical mean (red dashed line) and estimated mean (blue solid line). Circles represent failures, solid squares are censored events. (b) is a histogram of the Cox–Snell residuals with theoretical density function. (c) is the exponential probability plot. 1 is added to the censored residuals.	79

List of Tables

1	Maximum value for $LCV(h)$ with corresponding h for each of the covariates z_{i1} , z_{i2} and x_i for censoring levels of 20%, 50% and 80%.	32
2	Maximum value for $LCV(h)$ with corresponding h for x_i^2 for censoring levels of 20%, 50% and 80%.	35
3	Observed Anderson–Darling statistics and 95%-quantiles from the estimated distributions of the Anderson–Darling statistics.	37
4	Maximum value for $LCV(h)$ with corresponding h for the covariate x_i for censoring levels of 20%, 50% and 80%.	55
5	Maximum value for $LCV(h)$ with corresponding h for the covariate x_i^2 for censoring levels of 20%, 50% and 80%.	57
6	Observed Anderson–Darling statistics and 95%-quantiles of the estimated distributions of the Anderson–Darling statistics.	58
7	Observed Anderson–Darling statistic and 95%-quantile from the estimated distribution for the superalloy data.	78

1 Introduction

In survival analysis a common problem is to determine whether or not some variables (covariates) are correlated with the lifetime (response variable). For covariates which are correlated with the response, it is not always straightforward to determine what kind of relationship they have. This is the main problem considered in this thesis. Say the true model of a variable, T , is of the form

$$\ln T = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + f(x) + \sigma W. \quad (1.1)$$

This is a log-location-scale regression model. Examples of log-location scale models are models with Weibull, log-logistic and lognormal lifetimes. In this thesis we investigate Weibull and lognormally distributed lifetimes. In (1.1) T is the lifetime, z_1 , z_2 and x are covariates, β_0 , β_1 and β_2 are model coefficients, and σW is the error term. The distribution of W varies with the distribution of T , but it is assumed to be known. The function $f(x)$, β_0 , β_1 and β_2 are unknown, but we will first assume that covariate x has a linear relationship with the response variable $\ln T$. By maximum likelihood the possibly misspecified model ($f(x)$ might be linear) is fitted,

$$\ln T = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \gamma x + \sigma W. \quad (1.2)$$

Then by using the misspecified model's Cox–Snell residuals we use two different methods for recovering the correct functional form of the covariate x . The first method uses a transformation of the expected value of the Cox–Snell residuals, while the second is by investigation of the estimated hazard rate of the Cox–Snell residuals, for each covariate, using the covariate order method [2]. The covariate order method is a method of non-parametric censored exponential regression.

We start the report by introducing the reader to the concept of survival analysis and some frequently encountered terms and distributions. Then, for both the Weibull and lognormal distribution, we use a large data set, simulating a population, to find the model coefficients which minimize the Kullback–Leibler distance [3] for a misspecified model. This is done by maximum likelihood fitting of (1.2) with the regression function `survreg` in R [4]. The model minimizing the Kullback–Leibler distance is denoted

$$\ln T = \beta_0^* + \beta_1^* z_1 + \beta_2^* z_2 + \gamma^* x + \sigma^* W. \quad (1.3)$$

The starred coefficients are the coefficients minimizing the Kullback–Leibler distance. The Kullback–Leibler distance can illustratively be thought of as the minimized *distance* between the true model (1.1) and the linearly constrained (misspecified) model (1.2). When trying to recover the functional form of covariate x , we discover that the terms that are modelled correctly, z_1 and z_2 , can be neglected because the Kullback–Leibler coefficients are approximately the same as the real ones for these covariates. Towards the end of the report we try to obtain the coefficients which minimizes the Kullback–Leibler

distance analytically, comparing it to the ones found by maximum likelihood.

Smaller, more practical and complete (i.e. non-censored) data sets are simulated and used to recover the functional form of a covariate with the transformation method for both probability distributions. Censored data sets are investigated with both the transformation method and the covariate order method. This is where the covariate order method is expected to excel. We use levels of 20%, 50% and 80% censoring to analyze the effect censoring has on the results for the different methods.

Finally we investigate the functional form of a covariate in a data set from real life. The data is from [5] and describes the relationship between the lifetimes of nickel-base superalloy specimens, in terms of low-cycle fatigue failures, as a function of different levels of pseudostress being applied to the specimens during their lifetimes.

2 Survival analysis

Survival analysis is a part of statistics which focus is on examining the *lifetimes*, often described by the variable T , of an item. The term *lifetime* is used for simplicity. It does not necessary denote the time a human being is alive. It can also be the time until the failure of a mechanical component etc. Neither is it always measured in traditional time units as hours or minutes. It can be the number of times a switch has been used, or the number of cycles for a tire. However, the lifetime is never negative, $T \geq 0$. In survival analysis we try to model the lifetimes using appropriate probability density functions. The probability density function, abbreviated *PDF*, is a well known term for any statistician. The PDF for the random variable T is often denoted as $f_T(t)$, or simply $f(t)$. Using the terminology of a *survival analyst*, it is the instantaneous probability of failure at time t . There are discrete and continuous distributions, but in this project we only encounter some of the continuous ones. Common for the two kinds are that the PDF must always sum to one. In the continuous case this means that

$$\int_0^{\infty} f_T(t)dt = 1. \quad (2.1)$$

The cumulative distribution function, abbreviated *CDF* and denoted $F_T(t)$, is defined as

$$F_T(t) = \Pr(T \leq t) = \int_0^t f_T(u)du. \quad (2.2)$$

$F_T(t)$ is thus the probability that the item being the focus of attention will fail in the time interval $(0, t]$. $F_T(t)$ takes values in $[0, 1]$. The complement of the CDF is the reliability function

$$R_T(t) = 1 - F_T(t) = \Pr(T > t) = \int_t^{\infty} f_T(u)du. \quad (2.3)$$

$R_T(t)$ is the probability the item does not fail in $(0, t]$. It is the probability that it survives up until time t . This is why it is also known as the survivor function. The last function we will present is the failure rate function, or the hazard rate function. The hazard function gives the instantaneous probability of failure at time t , given that it has survived until t :

$$\begin{aligned} \lambda_T(t) &= \lim_{\Delta T \rightarrow 0} \frac{\Pr(t < T \leq t + \Delta T | T > t)}{\Delta t}, \\ \lambda_T(t) &= \lim_{\Delta T \rightarrow 0} \frac{F_T(t + \Delta t) - F_T(t)}{\Delta t} \cdot \frac{1}{R_T(t)} = \frac{f_T(t)}{R_T(t)}. \end{aligned} \quad (2.4)$$

Having introduced the reader to survival analysis, we now present some of the distributions frequently encountered in the project, the concept of censoring, Cox–Snell residuals and the covariate order method.

2.1 The exponential distribution

The exponential distribution is the most frequently used distribution in survival analysis. This is mainly because it is easy to work with. Lifetimes that are exponentially distributed have PDF, mean and variance

$$\begin{aligned} f_T(t) &= \begin{cases} \lambda e^{-\lambda t} & \text{for } t > 0, \lambda > 0, \\ 0 & \text{otherwise,} \end{cases} \\ E[T] &= \frac{1}{\lambda}, \\ \text{Var}[T] &= \frac{1}{\lambda^2}. \end{aligned} \tag{2.5}$$

If λ in (2.5) equals one, we say that the variables are unit exponentially distributed. The reliability function is

$$R_T(t) = e^{-\lambda t} \text{ for } t > 0. \tag{2.6}$$

One of the main things to notice about the exponential distribution is that the hazard rate function is constant

$$\lambda_T(t) = \frac{f_T(t)}{R_T(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda. \tag{2.7}$$

This again leads to the *memoryless* property of the exponential distribution. The conditional reliability function is defined as the probability that an item will survive for an additional time x , given that it has survived up to t :

$$\begin{aligned} R_T(x|t) &= \Pr(T > t + x | T > t) = \frac{\Pr(T > t + x)}{\Pr(T > t)}, \\ &= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x} = \Pr(T > x) = R_T(x). \end{aligned} \tag{2.8}$$

The consequence of this is that an old item is as good as a new item. The memoryless property is a reason for why the exponential distribution is so easy to work with, but it is also a weakness. It is not realistic that an item will have the same hazard function throughout its lifetime.

2.2 The Weibull distribution

The Weibull distribution, named after the Swedish professor Waloddi Weibull, is also one of the most used distributions in survival analysis. It has PDF, mean and variance

$$\begin{aligned}
f_T(t) &= \begin{cases} \alpha \lambda^\alpha t^{\alpha-1} e^{-(\lambda t)^\alpha} & \text{for } t > 0, \lambda > 0, \\ 0 & \text{otherwise,} \end{cases} \\
E[T] &= \frac{1}{\lambda} \Gamma \left[\frac{1}{\alpha} + 1 \right], \\
\text{Var}[T] &= \frac{1}{\lambda^2} \left(\Gamma \left[\frac{2}{\alpha} + 1 \right] - \Gamma^2 \left[\frac{1}{\alpha} + 1 \right] \right).
\end{aligned} \tag{2.9}$$

Here α is the shape parameter, λ is the scale parameter and $\Gamma[\cdot]$ is the gamma function. The reliability function is

$$R_T(t) = e^{-(\lambda t)^\alpha} \text{ for } t > 0. \tag{2.10}$$

If $\alpha = 1$ we can compare (2.9) with (2.5) and see that the Weibull distribution becomes the exponential distribution. The exponential distribution is therefore a special case of the Weibull distribution.

2.3 The Gumbel distribution of the smallest extreme

The Gumbel distribution of the smallest extreme is an example of an extreme value distribution [6, p. 54]. Say we have a set of independent, identically distributed (*iid*) lifetimes, T_i 's. If the PDF of the lifetimes goes exponentially towards zero when $t \rightarrow \infty$, then the limiting distribution of a normalized version of $U_n = T_{(1)} = \min\{T_1, T_2, \dots, T_n\}$ is known to be

$$F_{T_{(1)}}(t) = 1 - e^{-e^{(t-\nu)/\alpha}} \text{ for } -\infty < t < \infty. \tag{2.11}$$

Here $\alpha > 0$, the mode, and ν , the scale parameter, are constants. If the lifetimes are standardized

$$Y = \frac{T - \nu}{\alpha},$$

we get the CDF of the standardized Gumbel distribution of the smallest extreme

$$F_{Y_{(1)}}(y) = 1 - e^{-e^y} \text{ for } -\infty < y < \infty. \tag{2.12}$$

Which leads to the corresponding PDF, mean and variance

$$\begin{aligned}
f_{Y_{(1)}}(y) &= e^y e^{-e^y} \text{ for } -\infty < t < \infty, \\
E[Y_{(1)}] &= -\phi, \\
\text{Var}[Y_{(1)}] &= \frac{\pi^2}{6},
\end{aligned} \tag{2.13}$$

where $\phi = 0.5772\dots$ is Euler's constant. The reliability function is

$$R_{Y_{(1)}}(y) = e^{-e^y} \text{ for } -\infty < y < \infty. \tag{2.14}$$

2.4 The normal distribution

The normal distribution, also called Gaussian distribution, is a frequently encountered continuous probability distribution in statistics. It has PDF, mean and variance

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2} \text{ for } -\infty < y < \infty, \\ E[Y] &= \mu, \\ \text{Var}[Y] &= \sigma^2. \end{aligned} \quad (2.15)$$

A normally distributed variable Y , with mean and variance equal μ and σ^2 , is denoted: $Y \sim \mathcal{N}(\mu, \sigma^2)$. If Y is $\mathcal{N}(0, 1)$ we say that Y is standard normally distributed. A normally distributed variable can have negative values, but it is sometimes used as a lifetime distribution nonetheless. The CDF of the normal distribution is found by integrating the PDF and is denoted

$$F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-t^2/2} dt = \frac{1}{2} \left[1 + \text{erf}\left(\frac{y-\mu}{\sqrt{2}\sigma}\right) \right]. \quad (2.16)$$

Here $\text{erf}(\cdot)$ is the error function

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (2.17)$$

The reliability function is

$$R_Y(y) = 1 - F_Y(y) = 1 - \Phi\left(\frac{y-\mu}{\sigma}\right) \quad (2.18)$$

2.5 The lognormal distribution

A lognormally distributed variable has PDF, mean and variance

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi}\sigma y} e^{-(\ln y - \mu)^2/2\sigma^2} \text{ for } y > 0, \\ E[Y] &= e^{\mu + \sigma^2/2}, \\ \text{Var}[Y] &= e^{2\sigma} (e^{2\mu^2} - e^{\mu^2}). \end{aligned} \quad (2.19)$$

The lognormal and normal distribution are related by: $Y \sim \text{lognormal}(\mu, \sigma^2)$ if $X = \ln T \sim \mathcal{N}(\mu, \sigma^2)$. The reliability function is

$$\begin{aligned} R_Y(y) &= \Pr(T > t) = \Pr(\ln T > \ln t) \\ &= \Pr\left(\frac{\ln T - \mu}{\sigma} > \frac{\ln t - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - \ln t}{\sigma}\right). \end{aligned} \quad (2.20)$$

Here $\Phi(\cdot)$ is the CDF of the standard normal distribution.

2.6 Censoring

A lifetime, T , is said to be censored if we are not able to observe it entirely. We have *right censoring* when we know when the item was put to test, but not when it failed. *Left censoring* is when we know that the item has failed, but not when. A third version is *interval censored* observations. This is when we know that the item failed within an interval of two observed values. An illustration of censoring can be seen in Figure 1. Censoring is said to be *random* if the censoring time and failure time are independent random variables.

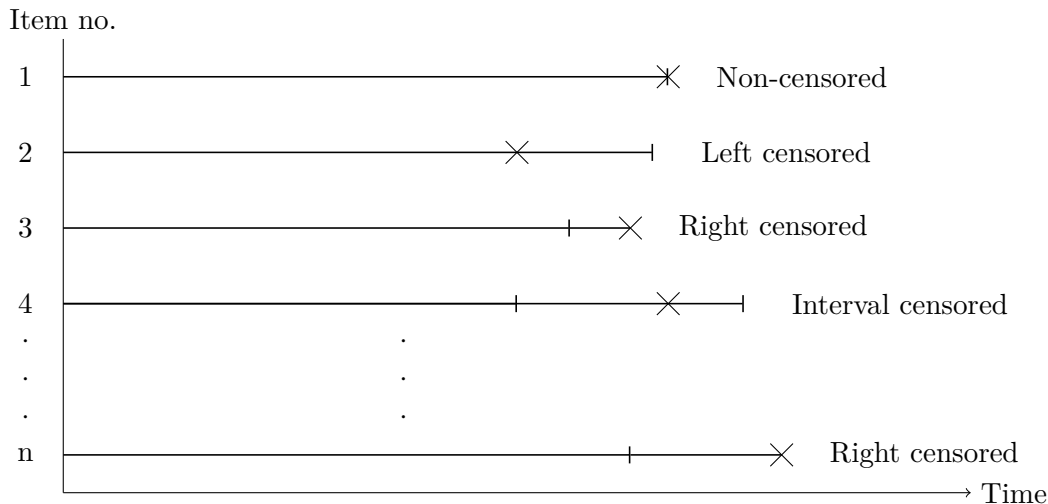


Figure 1: Illustration of left, right and interval censoring. The crosses illustrate when the failure occurs, while the short lines perpendicular to the time axis are the actual times observed.

2.7 Cox–Snell residuals

Cox–Snell residuals [7] are used to check if an assumed reliability function models a set of lifetimes adequately. We write the Cox–Snell residuals with a hat, \hat{R} , to avoid being mistaken for the reliability function. The Cox–Snell residuals are given by

$$\hat{R} = -\ln[R(T)] = -\ln[1 - F(T)], \quad (2.21)$$

where T is an observed lifetime, $R(\cdot)$ is the reliability function of the assumed distribution, and $F(\cdot)$ is the corresponding cumulative distribution function. If the distribution assumed is correct, the Cox–Snell residuals should be unit exponentially distributed.

Using a *log-location-scale model* means that we use lifetimes that comes from a log-location-scale distribution. A log-location-scale distributed variable is parametrized by a location parameter and a non-negative scale parameter.

$$\ln Y = \mu + \sigma X. \quad (2.22)$$

Here μ is the location parameter and σ is the scale parameter. From the models above we see that the Weibull distribution and the lognormal distribution are log-location scale distributions.

For more theory on survival analysis, we recommend the book “System Reliability Theory” [6] by M. Rausand and A. Høyland. Section 2.1-2.3 is a reproduction of parts from this book. Another recommended book on survival analysis is “Statistical Methods for Reliability Data” [5] by William Q. Meeker and Luis A. Escobar, which was used as a source for Section 2.6.

2.8 Covariate order method

The covariate order method is a nonparametric method for exponential regression. We use it to investigate the functional form of misspecified accelerated lifetime models. A short description of the covariate order method follows here, however, for a more detailed description, the reader is advised to look at [2].

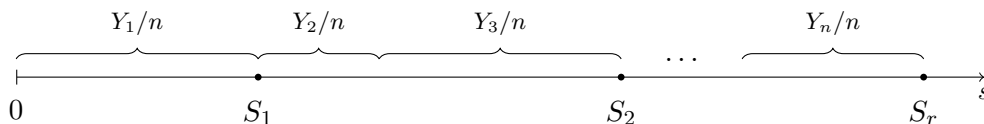


Figure 2: Constructed artificial point process with $(Y_1, \delta_1 = 1), (Y_2, \delta_2 = 0), (Y_3, \delta_3 = 1), \dots, (Y_n, \delta_n = 1)$.

A set of n independent observations $(Y_1, \delta_1, X_1), \dots, (Y_n, \delta_n, X_n)$, where Y_i is the observation time, δ_i the censoring status and X_i the covariate, which is assumed to be one-dimensional, are ordered ascendingly with respect to the covariate. I.e. $X_1 \leq X_2 \leq \dots \leq X_n$. The observations are, for convenience, scaled Y_i/n . $Y_1/n, \dots, Y_n/n$ are set as interarrival times of an artificial point (Poisson) process. If the endpoint of an interarrival time corresponds to a uncensored observation, this point is an event occurring at S_1, S_2, \dots, S_r , while the endpoints corresponding to censored observations are not considered as events. Here $r = \sum_{j=1}^n \delta_j$. An illustration of the process can be seen in Figure 2. The time of the events, S_i , are given by

$$S_i = \sum_{j=1}^{k(i)} \frac{Y_j}{n}, \text{ where } k(i) = \min\left(s \mid \sum_{j=1}^s \delta_j = i\right). \quad (2.23)$$

The covariate order method estimates, by using a kernel density estimator [8], the intensity of the artificial point process and then transforms this to an estimator of $\lambda(x)$, the hazard of a lifetime Y with covariate x . Theorems and accompanying proofs that the estimator is uniformly consistent can be found in [9].

3 Model

We have the lifetime model

$$\ln T = \boldsymbol{\beta}^T \mathbf{z} + f(x) + \sigma W. \quad (3.1)$$

This is a log-location-scale regression model. Here T is the lifetime ($\ln T$ is the response), $\boldsymbol{\beta}$ is a vector of the intercept β_0 and constant weights β_1, \dots, β_n , \mathbf{z} is a vector of covariates z_0, \dots, z_n ($z_0 \equiv 1$ because of the intercept term β_0), $f(\cdot)$ is an unknown function of the covariate x , and σW is the error term. W is from the distribution which give us the desired distribution for T . The functional form of $f(x)$ is unknown. However, we assume as a starting point that it is linear and fit the following model using maximum likelihood,

$$\ln T = \boldsymbol{\beta}^T \mathbf{z} + \gamma x + \sigma W. \quad (3.2)$$

The estimates obtained from the regression via maximum likelihood of this misspecified model, $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$ and $\hat{\sigma}$, are consistent estimators for $\boldsymbol{\beta}^*$, γ^* and σ^* . $\boldsymbol{\beta}^*$, γ^* and σ^* are the values which minimize the Kullback–Leibler distance [3]. The Kullback–Leibler distance is defined as

$$KL(\theta) = \mathbb{E} \left[\ln \frac{g(x)}{h(x, \theta)} \right], \quad (3.3)$$

where $g(x)$ is the density of the true model and $h(x, \theta)$ is the density of the misspecified model. These x 's have nothing to do with (3.1) and (3.2), they are purely illustrational. The expectation in (3.3) is taken with respect to the true model. In our case $g(x)$ and $h(x, \theta)$ are the densities of T given by (3.1) and (3.2), respectively. Figure 3 illustrates the idea where the misspecified model constrains $f(x)$ to a linear function. It follows that $\hat{\boldsymbol{\beta}}$, $\hat{\gamma}$ and $\hat{\sigma}$ are consistent estimators for $\boldsymbol{\beta}^*$, γ^* and σ^* for sufficiently large data sets.

Our idea is to find the form of $f(x)$ using the misspecified model's Cox–Snell residuals, \hat{R}^* . The Cox–Snell residuals are given by the definition in (2.21).

$$\hat{R}^* = -\ln[R^*(T)] = -\ln[1 - F^*(T)], \quad (3.4)$$

where $R^*(\cdot)$ is the reliability function, and $F^*(\cdot)$ is the cumulative distribution function of the misspecified distribution, i.e

$$\begin{aligned} F^*(t) &= \Pr(T \leq t) = \Pr(\ln T \leq \ln t) \\ &= \Pr\left(\frac{\ln T - \boldsymbol{\beta}^{*T} \mathbf{z} - \gamma^* x}{\sigma^*} \leq \frac{\ln t - \boldsymbol{\beta}^{*T} \mathbf{z} - \gamma^* x}{\sigma^*}\right) \\ &= \Phi\left(\frac{\ln t - \boldsymbol{\beta}^{*T} \mathbf{z} - \gamma^* x}{\sigma^*}\right). \end{aligned}$$

From this, the Cox–Snell residual for an observation, T_i , from the misspecified model (3.2) is

$$\hat{R}_i^* = -\ln \left[1 - \Phi \left(\frac{\ln T_i - \boldsymbol{\beta}_i^{*T} \mathbf{z}_i - \gamma^* x_i}{\sigma^*} \right) \right], \quad (3.5)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the distribution in question. The stars in the superscript of the coefficients indicate that these coefficients are the ones obtained from the fitted model. We want $f(x_i)$ to be a part of the expression for the Cox–Snell residuals, so we add and subtract terms from the underlying “real” model.

$$\begin{aligned} \hat{R}_i^* &= -\ln \left[1 - \Phi \left(\frac{\ln T_i - \boldsymbol{\beta}^T \mathbf{z}_i - f(x_i)}{\sigma} \cdot \frac{\sigma}{\sigma^*} + \frac{(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \mathbf{z}_i + f(x_i) - \gamma^* x_i}{\sigma^*} \right) \right] \\ &= -\ln \left[1 - \Phi \left(\frac{\sigma}{\sigma^*} W_i + \frac{(\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \mathbf{z}_i + f(x_i) - \gamma^* x_i}{\sigma^*} \right) \right], \end{aligned} \quad (3.6)$$

where W_i is distributed as Φ . Then solving this expression for $f(x_i)$

$$f(x_i) = -\sigma W_i - (\boldsymbol{\beta} - \boldsymbol{\beta}^*)^T \mathbf{z}_i + \gamma^* x_i + \sigma^* \cdot \Phi^{-1} \left(1 - e^{-\hat{R}_i^*} \right). \quad (3.7)$$

This expression can be used for any Φ when $\boldsymbol{\beta}$, $\boldsymbol{\beta}^*$, γ^* , σ and σ^* are well defined theoretically. For real data sets, we will not know the values of σ and $\boldsymbol{\beta}$, but we are able to estimate $\boldsymbol{\beta}^*$, γ^* and σ^* from the misspecified model. (3.7) is the expression we will use at first to recover the functional form of misspecified covariates. This is the method we refer to as the “transformation method”.

Kullback–Leibler distance

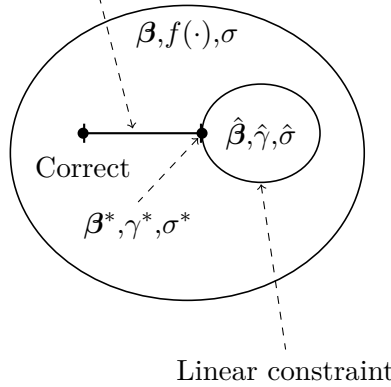


Figure 3: Illustration of the Kullback–Leibler distance.

4 Weibull distributed lifetimes

We start with lifetimes which are Weibull distributed with probability density function as in (2.9)

$$f_T(t) = ab^a t^{a-1} e^{-(bt)^a}, \quad t \geq 0. \quad (4.1)$$

Here a is the shape parameter and b is the scale parameter. When simulating the model given by (3.1), we can ensure the T_i 's to be Weibull distributed by using W_i 's which are from the standard Gumbel distribution of the smallest extreme. The W_i 's have probability density function

$$f_W(w) = e^w e^{-e^w}, \quad -\infty < w < \infty. \quad (4.2)$$

By using the transformation formula for functions of random variables, which can be found in any introductory statistics book, we can show that the lifetime, T , in (3.1) is indeed Weibull distributed. Assuming that the shape and scale parameter in (4.1) are respectively equal to $1/\sigma$ and $e^{-(\beta^T \mathbf{z} + f(x))}$, then isolate T in (3.1) and transform:

$$\begin{aligned} T &= e^{\beta^T \mathbf{z} + f(x) + \sigma W}, \\ \left| \frac{dT}{dW} \right| &= \sigma e^{\beta^T \mathbf{z} + f(x) + \sigma W}, \\ f_W(w) &= f_T \left(e^{\beta^T \mathbf{z} + f(x) + \sigma w} \right) \cdot \sigma e^{\beta^T \mathbf{z} + f(x) + \sigma w} \\ &= \frac{1}{\sigma} \left(e^{-(\beta^T \mathbf{z} + f(x))} \right)^{1/\sigma} \left(e^{\beta^T \mathbf{z} + f(x) + \sigma w} \right)^{1/\sigma - 1} \\ &\quad \cdot e^{\left(-e^{-(\beta^T \mathbf{z} + f(x))} e^{\beta^T \mathbf{z} + f(x) + \sigma w} \right)^{1/\sigma}} \cdot \sigma e^{\beta^T \mathbf{z} + f(x) + \sigma w} \\ &= \frac{1}{\sigma} \left(e^{-(\beta^T \mathbf{z} + f(x))} \right)^{1/\sigma} \frac{\left(e^{\beta^T \mathbf{z} + f(x)} \right)^{1/\sigma}}{\left(e^{\beta^T \mathbf{z} + f(x)} \right)} \frac{(e^{\sigma w})^{1/\sigma}}{e^{\sigma w}} \cdot e^{-e^w} \cdot \sigma e^{\beta^T \mathbf{z} + f(x)} e^{\sigma w} \\ &= e^w e^{-e^w}. \end{aligned}$$

Which equals (4.2). Using the definition of the Cox–Snell residuals in (2.21), we can find the Cox–Snell residuals for the misspecified model. Now we must use the CDF of standard Gumbel distributed variables of the smallest extreme. The residuals become

$$\begin{aligned} \hat{R}_i^* &= -\ln [1 - F_W(W_i)] = -\ln [R_W(W_i)] \\ &= -\ln \left[e^{-e^{W_i}} \right] = e^{W_i} = e^{\frac{\ln T_i - \beta^{*T} \mathbf{z}_i - \gamma^* x_i}{\sigma^*}}. \end{aligned} \quad (4.3)$$

Here we solved (3.2) for W .

4.1 Coefficients minimizing the Kullback–Leibler distance

In order to find the coefficients minimizing the Kullback–Leibler distance, β^* , γ^* and σ^* , we simulate a data set with a very large number of observations. The motivation for this is to be sure that the methods used can provide accurate estimates for the parameter coefficients from the misspecified model. When we know the original coefficients, it is easy to see if the method generates satisfying estimates. The model used in the simulation of the large data set was the “true” model:

$$\ln T_i = \beta^T z_i + f(x_i) + \sigma W_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + f(x_i) + \sigma W_i. \quad (4.4)$$

The coefficients and function were set to

$$\begin{aligned} \beta_0 &= 0, \\ \beta_1 &= 5, \\ \beta_2 &= 0.2, \\ f(x_i) &= x_i^2, \\ \sigma &= 2. \end{aligned} \quad (4.5)$$

When using R [4] to simulate the covariates in the model for T_i , one has to be careful regarding the parametrization of the Gumbel distributed error terms. The `rgumbel()` function in the `VGAM` library in R samples from the Gumbel distribution of the largest extreme. Therefore we implemented our own simulating procedure for variables from the Gumbel distribution of the smallest extreme. The CDF of any variable takes values in $[0,1]$. Letting u be uniformly distributed in $[0,1]$, i.e. $u \sim \text{Unif}[0, 1]$, and solving for w in the expression for the CDF of a standard Gumbel distributed variable of the smallest extreme

$$\begin{aligned} F_W(w) &= 1 - e^{-e^w} = u, \\ \Rightarrow \ln[-\ln(1 - u)] &= w, \\ \Rightarrow \ln[-\ln(u)] &= w, \end{aligned}$$

where the last relationship utilizes the fact that $1 - \text{Unif}[0,1] = \text{Unif}[0,1]$. By simulating u uniformly on $[0,1]$ we can create standard Gumbel distributed variables of the smallest extreme by using this relationship. z_{i1} , z_{i2} and x_i were drawn from the standard normal distribution. By inserting the covariates and error term into

$$T_i = e^{\beta^T z + f(x_i) + \sigma W_i},$$

we have simulated Weibull distributed lifetimes. 1 000 000 vectors of $(T_i, z_{i1}, z_{i2}, x_i, W_i)$ were simulated. All the vectors were saved in a data frame.

The next step in order to obtain the values for β^* , γ^* and σ^* , is to fit the misspecified model. This was done by using the `survreg` function in the `survival` library in R. The function call and summary of the model was:

```
> weibullModLarge <- survreg(Surv(wData$T_i) ~ wData$z_i1 + wData$z_i2
+ wData$x_i, data, dist='weibull')
> summary(weibullModLarge)
```

Call:

```
survreg(formula = Surv(wData$T_i) ~ wData$z_i1 + wData$z_i2 +
  wData$x_i, data = data, dist = "weibull")
              Value Std. Error      z      p
(Intercept)  1.2479   0.003316  376.37 0.00e+00
wData$z_i1   5.0060   0.003129 1599.63 0.00e+00
wData$z_i2   0.2169   0.003142   69.03 0.00e+00
wData$x_i    0.0099   0.001920    5.16 2.45e-07
Log(scale)   1.1433   0.000618 1848.61 0.00e+00
```

Scale= 3.14

Weibull distribution

```
Loglik(model)= -2442543   Loglik(intercept only)= -3105705
Chisq= 1326324 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 1000000
```

From the summary we can find the estimated coefficients. However, again we have to take into consideration the parametrization R uses. The `survreg` function's parametrization of the Weibull distribution is embedded in a general location-scale family. The relationship between the parameters in `survreg`'s scale and the shape in (4.1) is

$$\text{survreg's scale} = 1/a.$$

Knowing this we find the coefficients from the summary of the regression model

$$\begin{aligned}\beta_0^* &= 1.2479, \\ \beta_1^* &= 5.0060, \\ \beta_2^* &= 0.2169, \\ \gamma^* &= 0.0099, \\ \sigma^* &= 3.14.\end{aligned}\tag{4.6}$$

The values for β_1^* and β_2^* are close to the values they were set to in (4.5). This is, as expected due to the misspecified term, not the case for β_0^* and σ^* . γ^* is seen to be small

as well. The reason for all this will be clarified later. With these results the Cox–Snell residuals can be calculated according to (4.3). Plots of the Cox–Snell residuals versus the covariates are illustrated in Figure 4

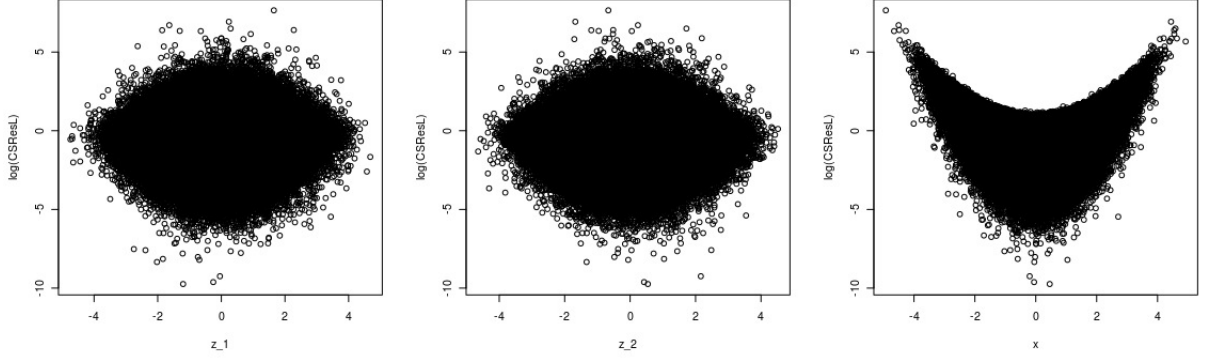


Figure 4: Plot of the logarithm of the Cox–Snell residuals, $\ln(\hat{R}_i^*)$, versus the covariates z_{i1} , z_{i2} and x_i (from left to right).

From Figure 4 it looks as there is no dependency between the residuals and z_{i1} and z_{i2} . The large absolute values of the logarithm of the Cox–Snell residuals in the middle are due to the fact that the normal distribution has the most observations in the middle. However, for x_i there is a clear tendency that the residuals increase rapidly for large absolute values of x_i . Knowing that the fitted model provides satisfying values for β_1^* and β_2^* , we need to clarify what the model’s intentions are for the values of β_0^* and γ^* . If we look at the parameters isolatedly, they essentially describe a linear function with intercept equal β_0^* and slope equal γ^* . From the summary of the model we see that 0 lies within one standard deviation of the estimated value of γ^* . Thus a zero hypothesis stating that $\gamma^* = 0$ would not be rejected. The misspecified models solution for estimating $f(x_i)$ is simply to add the constant β_0^* . The model value for σ , σ^* , is also far from the true value in (4.5), also caused by the misspecified covariate.

4.1.1 Using Cox–Snell residuals to infer the form of $f(x_i)$

As mentioned earlier we want to investigate the form of $f(x_i)$ using the Cox–Snell residuals. This is done using (3.7). If we take another look at the summary of the regression model, we see that the value of β_1^* and β_2^* are within one standard deviation of the values they were set to in (4.5). This means that the terms $(\beta_1 - \beta_1^*)z_{i1}$ and $(\beta_2 - \beta_2^*)z_{i2}$ in (3.7) are so small that they are negligible. The expression can therefore be approximated by

$$f(x_i) \approx -\sigma W_i - (\beta_0 - \beta_0^*) + \gamma^* x_i + \sigma^* \cdot \Phi^{-1} \left(1 - e^{-\hat{R}_i^*} \right). \quad (4.7)$$

Using $\Phi(\cdot)$ from the Gumbel distribution of the smallest extreme, given by equation (2.12), the inverse function becomes

$$\begin{aligned}\Phi^{-1}(u) &= \ln(-\ln(1-u)) \\ \Phi^{-1}\left(1 - e^{-\hat{R}_i^*}\right) &= \ln\left[\hat{R}_i^*\right]\end{aligned}\tag{4.8}$$

Inserting this into (4.7) gives us

$$f(x_i) \approx -\sigma W_i - (\beta_0 - \beta_0^*) + \gamma^* x_i + \sigma^* \ln\left[\hat{R}_i^*\right].\tag{4.9}$$

By inserting the computed Cox–Snell residuals from (4.3) for \hat{R}_i^* in (4.9), the values for $f(x_i)$ were computed using the coefficients in (4.6). Figure 5(a) shows a plot of all the estimated values for $f(x_i)$, while Figure 5(b) shows the estimated function plotted as a solid line.

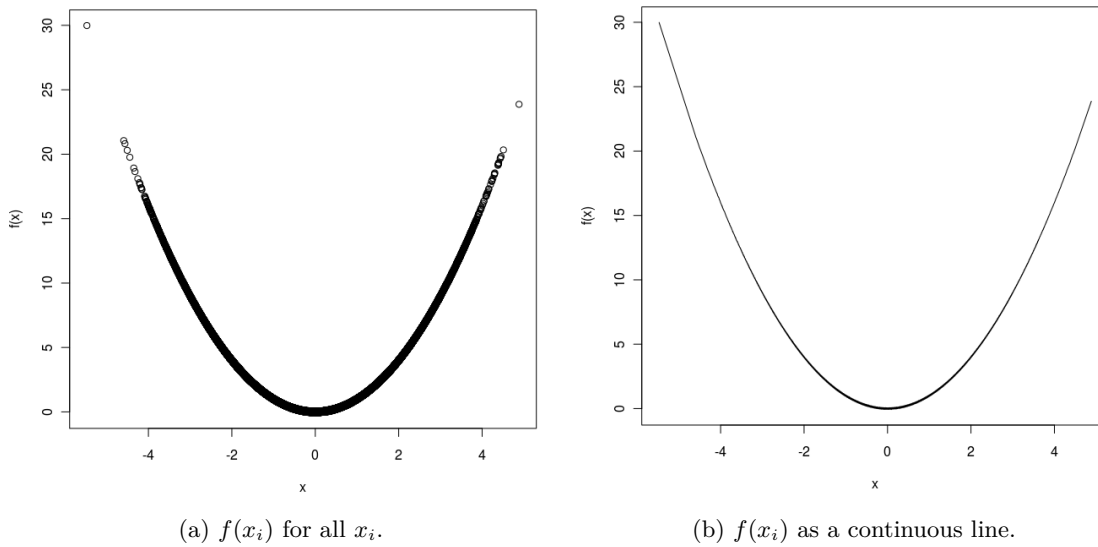


Figure 5: Plot of the estimation of the function $f(x_i)$ from (4.9).

From these plots we see that the form $f(x_i) = x_i^2$ is recovered.

4.2 Simulation and model fitting of Weibull distributed lifetimes

In Section 4.1 the simulated data set had 1 million observations. In practice we may have data sets which are much smaller, and this will be the focus in this section. However, for smaller, more realistically sized data sets, we have to be aware of some moments. Remember (4.9), the approximation of $f(x_i)$. If we have a real data set, we will not know the values of σ and β_0 . Neither will the values for $\beta^*, \gamma^*, \sigma^*$ be known, we would only know the corresponding estimates $\hat{\beta}, \hat{\gamma}, \hat{\sigma}$. The knowledge about $f(x_i)$ will be restricted to the relationship

$$f(x_i) \sim \ln \left[\hat{R}_i \right] \cdot \hat{\sigma} + \hat{\beta}_0 + \hat{\gamma}x_i. \quad (4.10)$$

The points will be displaced by the value $-\sigma W_i + \beta_0$. However, the main thing to notice is that the displacement does not depend on x_i . It is also worth noticing that if the $(\beta_1 - \beta_1^*)z_{i1}$ and $(\beta_2 - \beta_2^*)z_{i2}$ terms in (3.7) are not negligible, they will also lead to a displacement of the curve. This displacement is independent of the x_i 's as long as the z_i 's and x_i 's are independent. The plot obtained from the expression in (4.10) in Figure 6(a) is blurred compared to the plot obtained from the expression in (4.9) in Figure 5(a). The blurring comes from the stochastic term, σW_i , left out. The rest of the displacement is best seen in the smoothed plot. The minima of the smoothed curved in Figure 5(b) is approximately zero, while the minima in Figure 6(b) is less than zero. This smoothed line was obtained by taking the mean value of all approximated values of $f(x_i)$ within intervals of length 0.1 on the x -axis.

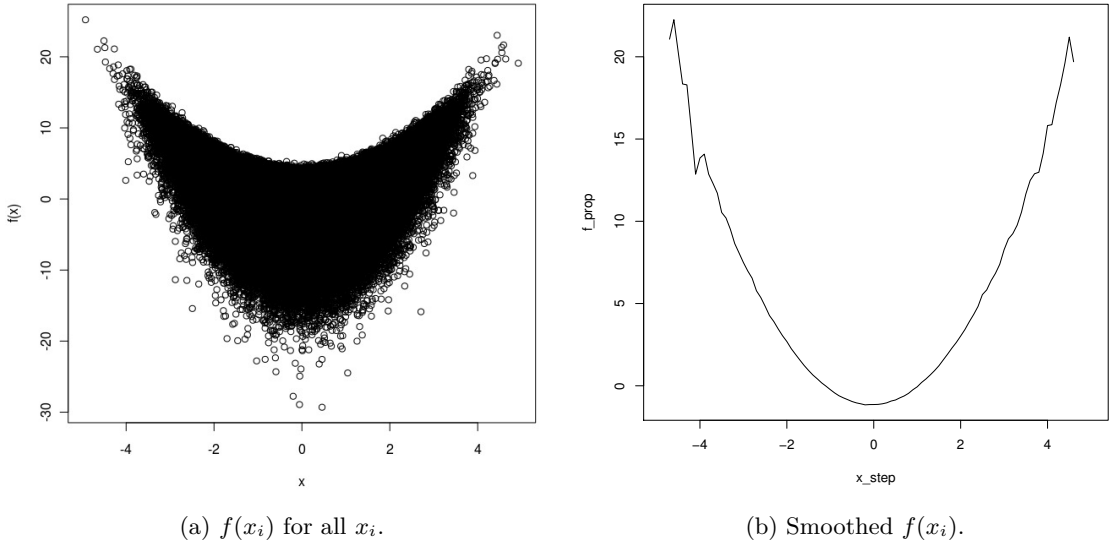


Figure 6: Plot of the estimation of the function $f(x_i)$ from (4.10).

In order for us to be able to neglect the $(\beta_1 - \hat{\beta}_1)z_{i1}$ and $(\beta_2 - \hat{\beta}_2)z_{i2}$ terms, we need to know how large the data set will have to be to provide adequate estimates for $\beta^*, \gamma^*, \sigma^*$. In the specialization project [1], which is the basis for this thesis, we found that 100 observations would be sufficient for this kind of problem.

4.2.1 Uncensored data set

We start with simulating an uncensored set of 100 lifetimes. An uncensored set is said to be complete. The lifetimes come from the same model

$$\ln T_i = \beta^T \mathbf{z}_i + f(x_i) + \sigma W_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + f(x_i) + \sigma W_i, \quad (4.11)$$

with the same coefficients and function

$$\begin{aligned} \beta_0 &= 0, \\ \beta_1 &= 5, \\ \beta_2 &= 0.2, \\ f(x_i) &= x_i^2, \\ \sigma &= 2, \end{aligned} \quad (4.12)$$

as the lifetimes for the large data set in Section 4.1. The covariates are drawn from the standard normal distribution. Using `survreg`, a misspecified model

$$\ln T_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \gamma x_i + \sigma W_i, \quad (4.13)$$

was fitted to the data. We want to check if the linear term, γx_i , should be replaced by a function $f(x_i)$, and what this function may look like. The summary from `survreg` for the misspecified model was:

```
> summary(weibullMod)
```

```
Call:
```

```
survreg(formula = Surv(data$T_i) ~ data$z_i1 + data$z_i2 + data$x_i,
        data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.4597	0.3181	4.589	4.45e-06
data\$z_i1	5.0963	0.2584	19.720	1.44e-86
data\$z_i2	0.5184	0.2973	1.744	8.12e-02
data\$x_i	0.0771	0.2133	0.361	7.18e-01
Log(scale)	1.0925	0.0751	14.547	6.08e-48

```
Scale= 2.98
```

```
Weibull distribution
```

```
Loglik(model)= -254   Loglik(intercept only)= -328.7
Chisq= 149.49 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 100
```

As mentioned previously, with a small data set the estimated coefficients $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\sigma}$ are not necessarily the true values for β^* , γ^* and σ^* . Nevertheless, $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\sigma}$ are consistent estimators for β^* , γ^* and σ^* . This because we have seen that when having a large number of observations, the coefficients for the correctly modelled covariates looked to converge to their true values. From the summary we see that the estimated values $\hat{\beta}_1$ and $\hat{\beta}_2$ are not significantly different from the known true values. It is also seen that the `Scale` estimate of 2.98 is pretty good considering we only have 100 observations. We calculate the Cox–Snell residuals according to (4.3), with σ^* , β_0^* and γ^* replaced by $\hat{\sigma}$, $\hat{\beta}_0$ and $\hat{\gamma}$, respectively, and use (4.10) to find the estimates for $f(x_i)$. In Figure 7(a) the values for the estimated $f(x_i)$ are plotted against the x_i 's. We see that there is a “U”-shape characteristic for quadratic functions. Using the function `lm` in R, a quadratic regression model of the form $f(x) = b_1x + b_2x^2$ was fitted.

```
> quad <- lm(f_prop ~ x + I(x^2)-1)
> summary(quad)
```

Call:

```
lm(formula = f_prop ~ x + I(x^2) - 1)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-8.8250	-2.4583	-0.7959	0.9134	4.2382

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
x	-0.3192	0.2832	-1.127	0.263
I(x^2)	0.8247	0.1526	5.405	4.56e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.922 on 98 degrees of freedom

Multiple R-squared: 0.234, Adjusted R-squared: 0.2183

F-statistic: 14.97 on 2 and 98 DF, p-value: 2.127e-06

From this summary we see that only the coefficient in front of the quadratic term is significantly different from zero. The regression line is plotted as the red dashed line in Figure 7(b) together with the original dots, as well as another estimated line. This estimated line, the solid black one, is obtained from the `lowess` function in R. The `lowess` function estimates the *locally weighted scatterplot smoothed* [10] line from the

points. From the plot we conclude that 100 observations give us a good indication of what $f(x_i)$ looks like.

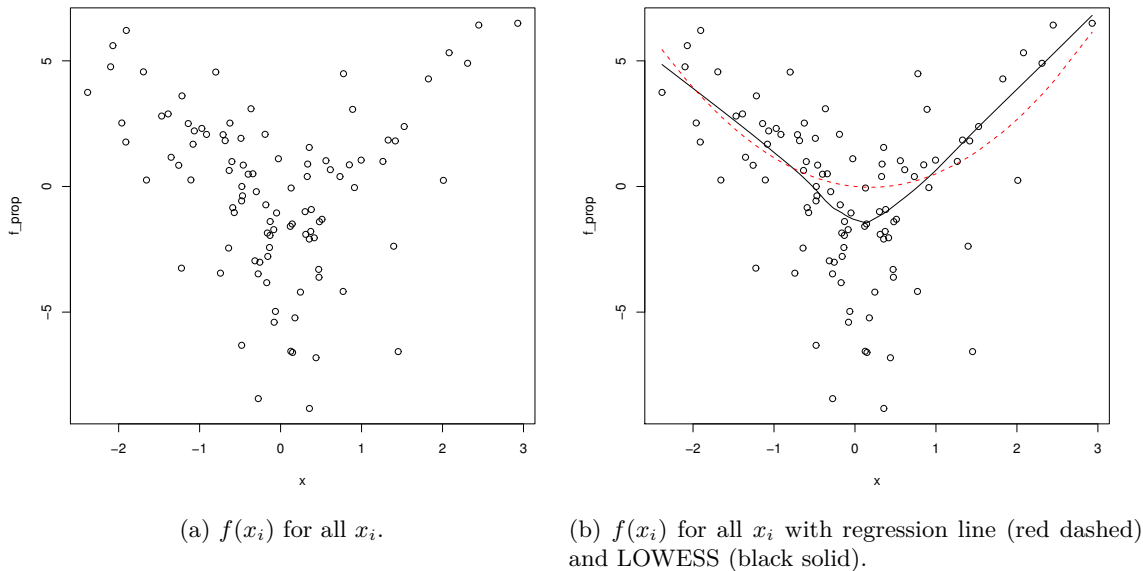


Figure 7: Plot of the estimation of the function $f(x_i)$ from (4.10) with 100 observations.

4.2.2 Censored data sets

Now we simulate censored data sets. A background of censoring is described in Section 2.6. We use random right censoring. Independent censoring times, C , drawn from the exponential distribution with $\lambda_{20} = 0.005$, $\lambda_{50} = 0.5$ and $\lambda_{80} = 100$ gave approximately 20%, 50% and 80% censoring in the data set. The simulated data set with censoring times is listed in its entirety in Appendix A. Two methods for recovering the functional form of the misspecified covariate is tested for censored data; the transformation method, which we have used up until now, and an approach using the covariate order method.

4.2.2.1 Transformation method

Using `survreg` we again fitted a model to the simulated censored data analogously to the uncensored data investigated earlier. The only difference is the `Surv()` element in the call to the `survreg` function. For the uncensored data set we used `Surv(weibullData$T_i)`, but for the censored data set we used `Surv(weibullData$Y_i, weibullData$s_i)`. Here `weibullData$Y_i` is given by $Y_i = \min(T_i, C_i)$ and `weibullData$s_i` indicates the censoring status. The censoring status, s_i , is given by

$$s_i = 1 \text{ if } T_i \leq C_i,$$

$$s_i = 0 \text{ if } T_i > C_i.$$

The misspecified model fitted with `survreg` was

$$\ln Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \gamma x_i + \sigma W_i. \quad (4.14)$$

This model was fitted for the three different censoring levels. The summary from `survreg` for the misspecified models were, for ascending levels of censoring:

```
> summary(weibull20Mod)
```

Call:

```
survreg(formula = Surv(data$Y20_i, data$s20_i) ~ data$z_i1 + data$z_i2 +
  data$x_i, data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.4904	0.3475	4.2886	1.80e-05
data\$z_i1	4.9520	0.3602	13.7487	5.19e-43
data\$z_i2	0.5023	0.3191	1.5739	1.16e-01
data\$x_i	0.0201	0.2444	0.0821	9.35e-01
Log(scale)	1.0724	0.0872	12.3011	8.94e-35

Scale= 2.92

Weibull distribution

Loglik(model)= -52 Loglik(intercept only)= -109.6

Chisq= 115.16 on 3 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 6

n= 100

```
> summary(weibull50Mod)
```

```
Call:
```

```
survreg(formula = Surv(data$Y50_i, data$s50_i) ~ data$z_i1 +
  data$z_i2 + data$x_i, data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.153	0.526	2.191	2.85e-02
data\$z_i1	4.684	0.499	9.381	6.55e-21
data\$z_i2	0.523	0.408	1.283	2.00e-01
data\$x_i	-0.041	0.307	-0.134	8.94e-01
Log(scale)	1.043	0.113	9.189	3.97e-20

```
Scale= 2.84
```

```
Weibull distribution
```

```
Loglik(model)= 93.3 Loglik(intercept only)= 50.2
```

```
Chisq= 86.14 on 3 degrees of freedom, p= 0
```

```
Number of Newton-Raphson Iterations: 5
```

```
n= 100
```

```
> summary(weibull80Mod)
```

```
Call:
```

```
survreg(formula = Surv(data$Y80_i, data$s80_i) ~ data$z_i1 +
  data$z_i2 + data$x_i, data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	1.566	1.519	1.031	3.02e-01
data\$z_i1	5.003	0.991	5.050	4.42e-07
data\$z_i2	0.796	0.586	1.358	1.74e-01
data\$x_i	0.397	0.464	0.855	3.92e-01
Log(scale)	0.978	0.181	5.409	6.33e-08

```
Scale= 2.66
```

```
Weibull distribution
```

```
Loglik(model)= 108.8 Loglik(intercept only)= 81
```

```
Chisq= 55.7 on 3 degrees of freedom, p= 4.9e-12
```

```
Number of Newton-Raphson Iterations: 7
```

```
n= 100
```

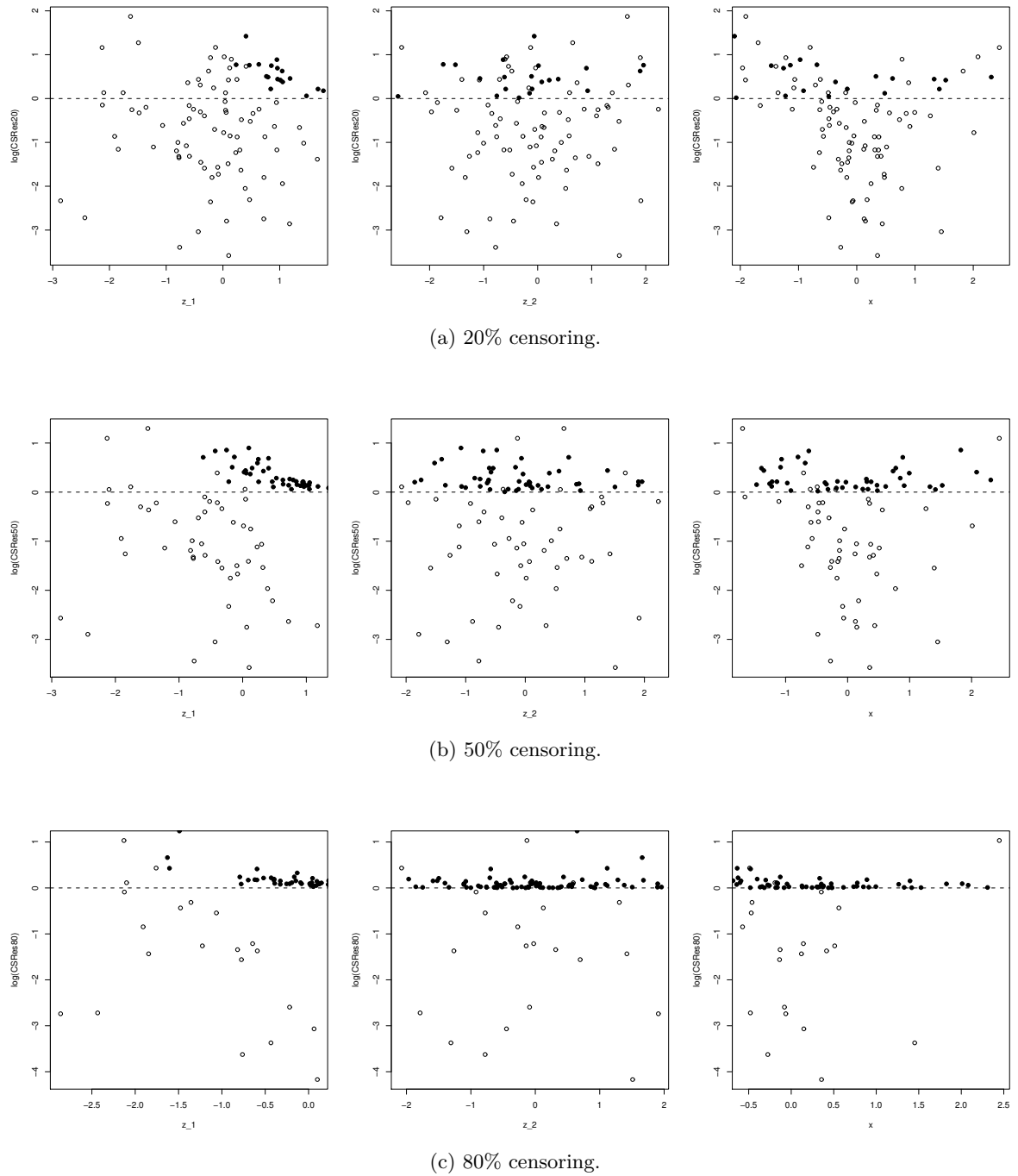
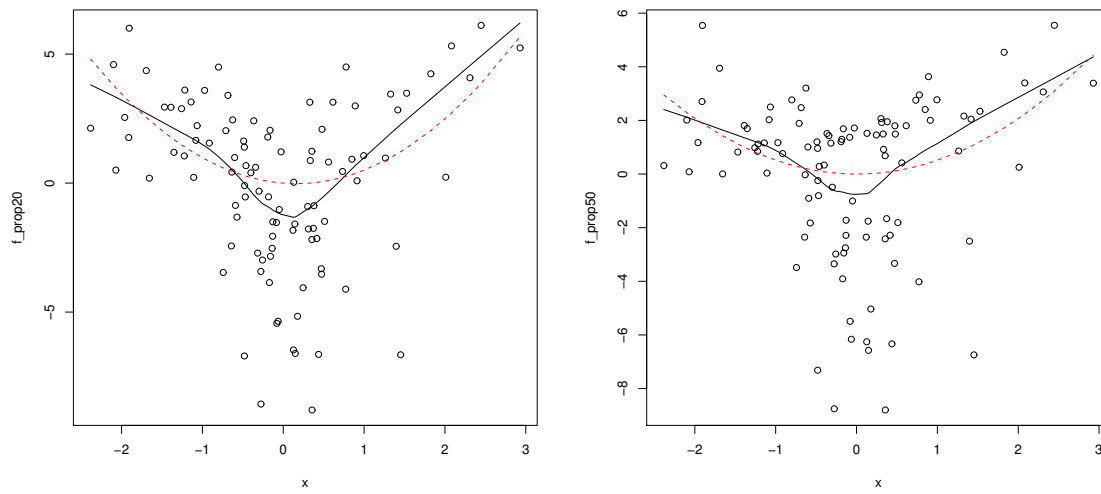
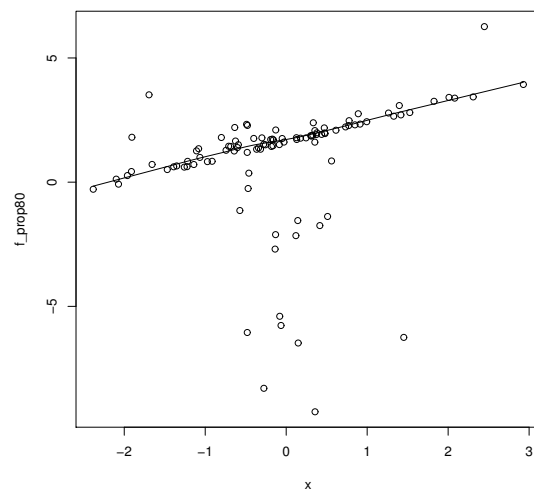


Figure 8: Plots of the covariates (z_1 , z_2 , and x respectively from left to right) versus the logarithm of the Cox–Snell residuals from the misspecified models (x_i is modelled linearly, γx_i) for each of the censoring levels. The circles represent residuals from uncensored observations, while the dots represent residuals from censored observations.



(a) 20% censoring.

(b) 50% censoring.



(c) 80% censoring.

Figure 9: Plot of the estimation of the function $f(x_i)$ with LOWESS (solid line) and quadratic regression (dashed line) estimates for the different censoring levels.

The Cox–Snell residuals for the uncensored observations from the misspecified models were calculated analogously as in (4.3),

$$\hat{R}_i = e^{\frac{\ln Y_i - \hat{\beta}_0 - \hat{\beta}_1 z_{i1} - \hat{\beta}_2 z_{i2} - \hat{\gamma} x_i}{\hat{\sigma}}}. \quad (4.15)$$

For the censored observations it is not quite that simple. We start by calculating temporary Cox–Snell residuals according to (4.15), but since the observations are right censored, it is obvious that the actual lifetimes, T_i , would have been longer than the recorded observation times, $Y_i (= C_i)$ (censoring times). Assuming that the regression model is correct, the Cox–Snell residuals in (4.15) should be unit exponentially distributed. Due to the memoryless property of the exponential distribution (Section 2.1), we simply add one to the temporary Cox–Snell residuals to compensate for the observation times cut short. This corresponds to the fact that the remaining time of a unit exponential variable is always unit exponential. Together with the Cox–Snell residuals for the uncensored observations we now have the Cox–Snell residuals for all of the observations. This was done for all three levels of censoring. We plot the logarithm of the residuals against the covariates to investigate any dependencies between them. This can be seen in Figure 8. For correct models the residuals are independent of the covariates. From Figure 8(a), with plots from the model with 20% censoring, we see that the residuals are random when plotted against z_{i1} and z_{i2} , but for the x_i 's they seem to have large positive values for large absolute values of x_i , and they have large negative values for small absolute values of x_i . This indicates that the x_i 's are not modelled correctly. The same observation can be seen in the plot with 50% censoring (Figure 8(b)), but not as clear. For the plot with 80% censoring (Figure 8(c)) the logarithm of the residuals are close to zero for the censored observations. This is because the censoring times are very small for many of the observations, and the Cox–Snell residuals are then approximately one because we add one to compensate for the observation times being cut short.

The Cox–Snell residuals were again used to estimate the functional form of $f(x_i)$ by using (4.10), for each censoring level. For the 20% censoring level the plot (Figure 9) shows that a “U”-shaped function would fit the data better. The same relationship can, with a stretch, be seen with 50% censoring as well. While for 80% censoring there is not much to learn from the plot. The censored observations ruins the plot by approximating a vast majority of the residuals to one. For the two lowest censoring levels we fit quadratic regression models to the respective plots:


```

> summary(quad20)

Call:
lm(formula = t(f_prop20) ~ x + I(x^2) - 1)

Residuals:
    Min       1Q   Median       3Q      Max
-8.8028 -2.3615 -0.4062  1.2521  4.2397

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
x          -0.2431     0.2891  -0.841   0.402
I(x^2)     0.7436     0.1558   4.774 6.31e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.983 on 98 degrees of freedom
Multiple R-squared:  0.1911, Adjusted R-squared:  0.1745
F-statistic: 11.57 on 2 and 98 DF,  p-value: 3.076e-05

> summary(quad50)

Call:
lm(formula = t(f_prop50) ~ x + I(x^2) - 1)

Residuals:
    Min       1Q   Median       3Q      Max
-9.130 -1.323  1.411  4.109  6.424

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
x          -0.3643     0.3788  -0.962   0.339
I(x^2)     1.3550     0.2041   6.640 1.76e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.908 on 98 degrees of freedom
Multiple R-squared:  0.3122, Adjusted R-squared:  0.2981
F-statistic: 22.24 on 2 and 98 DF,  p-value: 1.089e-08

```

When looking at the summaries from the quadratic regressions, we see that a quadratic term is significant for the models with 20% and 50% censoring. We respecify these models as

$$\ln Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \gamma_2 x_i^2 + \sigma W_i. \quad (4.16)$$

Fitting the models in (4.16) with `survreg` gives the following summaries:

```
> summary(weibull20Mod)
```

Call:

```
survreg(formula = Surv(data$Y20_i, data$s20_i) ~ data$z_i1 +
        data$z_i2 + I((data$x_i)^2), data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.277	0.3293	-0.84	4.01e-01
data\$z_i1	5.059	0.2749	18.40	1.25e-75
data\$z_i2	0.249	0.2146	1.16	2.46e-01
I((data\$x_i)^2)	1.459	0.2465	5.92	3.26e-09
Log(scale)	0.789	0.0864	9.13	6.96e-20

Scale= 2.2

Weibull distribution

Loglik(model)= -27 Loglik(intercept only)= -109.6

Chisq= 165.17 on 3 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 6

n= 100

```
> summary(weibull50Mod)
```

Call:

```
survreg(formula = Surv(data$Y50_i, data$s50_i) ~ data$z_i1 +
        data$z_i2 + I((data$x_i)^2), data = data, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.771	0.418	-1.84	6.51e-02
data\$z_i1	4.765	0.365	13.05	6.50e-39
data\$z_i2	0.273	0.270	1.01	3.13e-01
I((data\$x_i)^2)	1.624	0.336	4.84	1.32e-06
Log(scale)	0.730	0.112	6.50	8.30e-11

Scale= 2.07

Weibull distribution

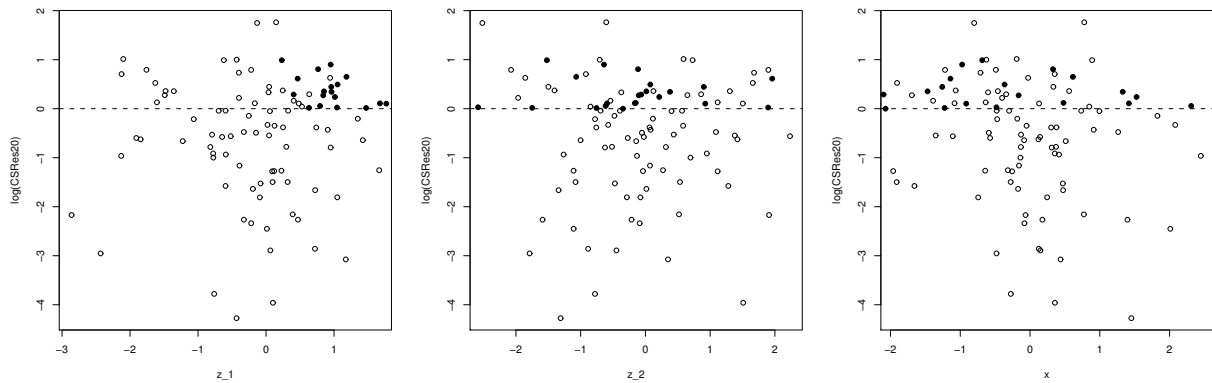
Loglik(model)= 112.6 Loglik(intercept only)= 50.2

Chisq= 124.71 on 3 degrees of freedom, p= 0

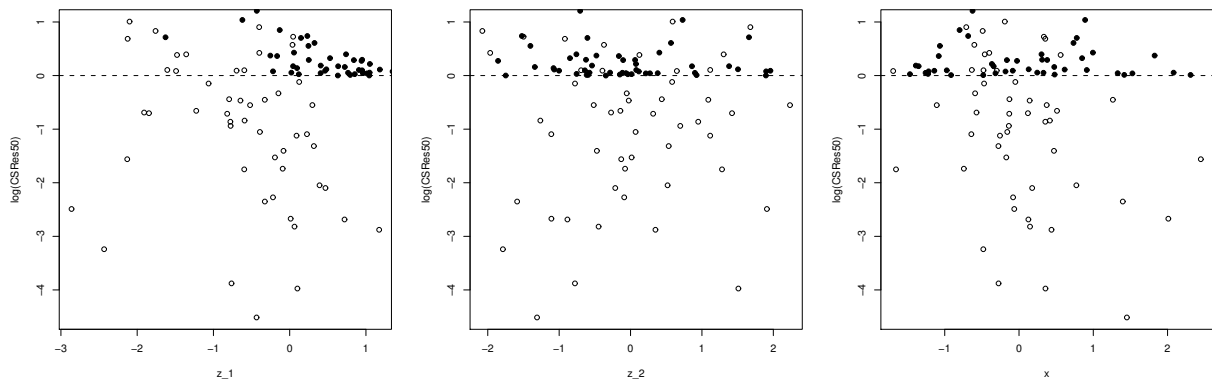
Number of Newton-Raphson Iterations: 7

n= 100

We immediately see that the other coefficients are estimated better now than previously, compared to the correct values. This is best seen in the **Scale** coefficient (i.e. $\hat{\sigma}$). However, this was expected due to the large residuals from the misspecified models. New Cox–Snell residuals were computed and log-plotted versus the covariates in Figure 10. The plots can be hard to read due to the covariates being clustered around zero, because they are normally distributed. However, from the plot we see that the relationship with large residuals for large absolute values of the x_i 's are gone. We believe that the functional form of the x_i 's is recovered.



(a) 20% censoring.



(b) 50% censoring.

Figure 10: Plot of the covariates (z_1 , z_2 , and x respectively from left to right) versus the logarithm of the Cox–Snell residuals from the respecified models in (4.16) (x_i is modelled quadratically, γx_i^2) for the 20% and 50% censoring levels. The circles represent residuals from uncensored observations, while the dots represent residuals from censored observations.

If the fitted model is correct, the Cox–Snell residuals should be unit exponentially distributed. To check if the residuals are distributed this way can be done in several ways. We start with a plot of the Cox–Snell residuals with the theoretical mean as well as the estimated mean (Figure 11(a)). These plots show that the unit exponential distribution might be the correct distribution for both censoring levels. Secondly, a histogram of the computed Cox–Snell residuals together with the theoretical density function for the unit exponential distribution (Figure 11(b)). These plots show that the assumed distribution might fit for a 20% censoring level, but the 50% censoring level does not fit as well. In particular we notice the peak for residuals approximately equal to one. This comes from the censored observations where we add one to the residuals. The same peak can be seen for 20% as well, but not as distinct. Finally, an exponential probability plot (Figure 11(c)). A probability plot [11] is constructed such that we can check if the data comes from an assumed theoretical distribution. This is done, in our case, by plotting the Cox–Snell residuals versus the probability the value of the residual has in the unit exponential distribution. If the assumed distribution fits, the plotted points should lie close to a straight line. From the left plot in Figure 11(a) one could suspect, if this was a real data set, that the residuals larger than five could come from observations characterized as outliers. From a practical viewpoint, an outlier might be the result of error during the recording of the data set, that it comes from a different population, or that it is an unusual observation (an observation with small probability) from the assumed distribution [12]. The probability plots supports the observation in the histograms. The reason why the Cox–Snell residuals are not exact unit exponentially distributed, is that the estimated values, $\hat{\beta}_0, \hat{\gamma}, \hat{\sigma}$, are used. Nonetheless, the plots indicate that the unit exponential is the true distribution for the Cox–Snell residuals for a 20% censoring level. This strengthens our belief that the model in (4.16) is correct for this level of censoring. However, it looks like this method for recovering the functional form for a misspecified covariate fails at high levels of censoring. The plot in 9(c) clearly illustrates this.

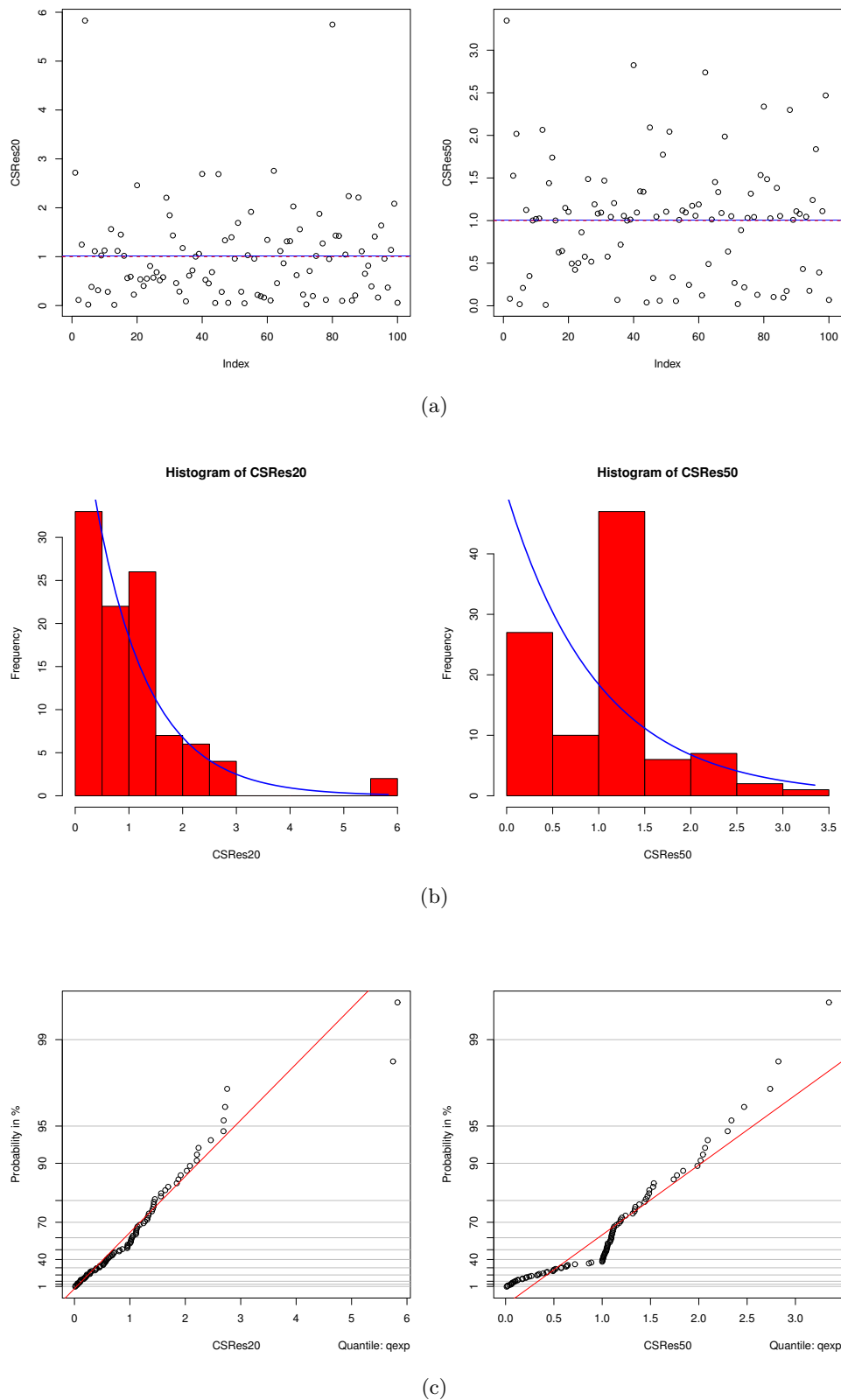


Figure 11: Plots checking if the Cox–Snell residuals are unit exponentially distributed. 20% censoring in the left column, 50% censoring in the right. (a) is a plot of the Cox–Snell residuals from the model (4.16) with theoretical mean (red dashed line) and estimated mean (blue solid line). (b) is a histogram of the Cox–Snell residuals with theoretical density function. (c) is the exponential probability plot. 1 is added to the censored residuals.

4.2.2.2 Covariate order method

Now we turn to the covariate order method, described in Section 2.8, to recover the functional form of the misspecified covariate for the three censoring levels. As mentioned in the description, this method should be better for high levels of censoring.

We start by fitting an empty model. The empty model is a model fitted without any covariates. The results from this can help us to determine how the covariates are modelled. We fit the empty model using `survreg`.

$$\ln Y_i = \beta_0 + \sigma W_i. \quad (4.17)$$

From the summary of the regression in R, estimates for $\hat{\beta}_0$ and $\hat{\sigma}$ are found. The Cox–Snell residuals are, according to (4.3), for the empty model

$$\hat{R}_i = e^{\frac{\ln Y_i - \hat{\beta}_0}{\hat{\sigma}}}. \quad (4.18)$$

These should ideally be unit exponentially distributed under the assumption that the empty model is correct. The residuals are now used as the response variable when using the covariate order method. The covariate order method estimates the hazard function, $\hat{\lambda}(x)$, for a covariate x . Implementation of the covariate order method in R was provided by Jan Terje Kvaløy. This code can be found in Appendix D.1. Assuming that the correct model is given by

$$\ln Y_i = \beta_0 + f(x_i) + \sigma W_i. \quad (4.19)$$

The expected value of the Cox–Snell residuals, \hat{R}_i , in (4.18), are estimated as $1/\hat{\lambda}(x_i)$ by the covariate order method. We need to find the expression for the expected value of the Cox–Snell residuals. Start by solving (4.17) for W_i . Then we find the expected value of the Cox–Snell residuals when Y_i has distribution given by (4.19),

$$\begin{aligned} E[\hat{R}_i^*] &= E \left[e^{\frac{\ln Y_i - \beta_0^*}{\sigma^*}} \right] \\ &= E \left[e^{\frac{\ln Y_i - \beta_0 - f(x_i)}{\sigma} \cdot \frac{\sigma}{\sigma^*} e^{\frac{(\beta_0 - \beta_0^*) + f(x_i)}{\sigma^*}}} \right]. \end{aligned} \quad (4.20)$$

The first exponential term in the expectation is shown to be Weibull distributed. The second term is only a constant depending on observation number i . Setting $U = e^{\frac{\ln Y_i - \beta_0 - f(x_i)}{\sigma}}$ and recognizing this as a unit exponentially distributed variable, i.e. $U \sim f_U(u) = e^{-u}$. Then setting the variable in the first exponential term as $Y = U^{\sigma/\sigma^*}$ and transforming:

$$\begin{aligned} Y = U^{\frac{\sigma}{\sigma^*}} &\Rightarrow U = Y^{\frac{\sigma^*}{\sigma}} \Rightarrow \frac{dU}{dY} = \frac{\sigma^*}{\sigma} Y^{\frac{\sigma^*}{\sigma} - 1}, \\ f_Y(y) &= f_U\left(y^{\frac{\sigma^*}{\sigma}}\right) \cdot \frac{\sigma^*}{\sigma} y^{\frac{\sigma^*}{\sigma} - 1} = \frac{\sigma^*}{\sigma} y^{\frac{\sigma^*}{\sigma} - 1} e^{-y^{\frac{\sigma^*}{\sigma}}}. \end{aligned}$$

Which we recognize, from (2.9), as a Weibull distributed variable with scale parameter 1 and shape parameter σ^*/σ . The expectation of a Weibull distributed variable is known from (2.9). Thus the expectation of the Cox–Snell residual becomes

$$E[\hat{R}_i^*] = \Gamma\left(\frac{\sigma}{\sigma^*} + 1\right) e^{\frac{(\beta_0 - \beta_0^*) + (f(x_i) - \gamma^* x_i)}{\sigma^*}}. \quad (4.21)$$

With this expression and the Cox–Snell residuals we have a tool which can help us to determine what $f(x_i)$ looks like. It is also worth noticing that if we do not take the expectation of (4.18), but still expand the terms as in (4.20) and use the same transformation as above, we see that

$$\hat{R}_i^* = U\sigma/\sigma^* e^{\frac{(\beta_0 - \beta_0^*) + (f(x_i) - \gamma^* x_i)}{\sigma^*}}, \quad (4.22)$$

which means that the \hat{R}_i^* 's are Weibull distributed with scale parameter $e^{\frac{(\beta - \beta^*)^T z_i + (f(x_i) - \gamma^* x_i)}{\sigma^*}}$ and shape parameter σ^*/σ . However, our focus will be on the expectation, $1/\hat{\lambda}(x_i)$, since this is where we can utilize $\hat{\lambda}(x_i)$, which is estimated by the covariate order method.

$$E[\hat{R}_i] = E\left[e^{\frac{\ln Y_i - \hat{\beta}_0}{\hat{\sigma}}}\right] \approx \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \cdot e^{\frac{(\beta_0 - \hat{\beta}_0) + f(x_i)}{\hat{\sigma}}} \approx \frac{1}{\hat{\lambda}(x_i)}. \quad (4.23)$$

Solving this equation for $f(x_i)$ gives us

$$f(x_i) = \hat{\beta}_0 - \beta_0 + \hat{\sigma} \ln \left[\left(\hat{\lambda}(x_i) \cdot \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \right)^{-1} \right]. \quad (4.24)$$

However, we do not know β_0 and σ , so the expression will be horizontally displaced by the value $\beta_0 + \hat{\sigma} \ln [\Gamma[\sigma/\hat{\sigma} + 1]]$. The relationship is then

$$f(x_i) = \hat{\beta}_0 - \hat{\sigma} \ln [\hat{\lambda}(x_i)]. \quad (4.25)$$

Plotting this versus x_i can give us an indication of the functional form of the covariates. We will do this for all covariates, z_{i1} , z_{i2} and x_i , for all three censoring levels. Before we use the covariate order method, we need to find the optimal smoothing parameters in each of the nine cases.

4.2.2.3 Finding the optimal smoothing parameter using a likelihood cross-validation criterion

When using the covariate order method, one has the choice of smoothing over the covariate axis or the event axis (the S_i 's in Figure 2). Since the covariates are standard normal distributed, with more observations around zero than in the tails, we choose to smooth over the covariate axis. A constant smoothing parameter on the covariate axis with a varying number of points within each interval would be unfortunate. The reader is referred to [9] for a more detailed description of the smoothing concept in the method. To find the optimal value of the smoothing parameter we use a likelihood cross-validation

criterion described in [9]. Letting X_i denote the covariate used and \hat{R}_i the Cox–Snell residual, then the likelihood of the Cox–Snell residuals are

$$L(\lambda(\cdot)) = \prod_{i=1}^n [\lambda(X_i) e^{-\lambda(X_i)\hat{R}_i}]^{\delta_i} [e^{-\lambda(X_i)\hat{R}_i}]^{1-\delta_i}, \quad (4.26)$$

which again gives us the log-likelihood function

$$l(\lambda(\cdot)) = \sum_{i=1}^n [\delta_i \ln \lambda(X_i) - \lambda(X_i)\hat{R}_i]. \quad (4.27)$$

Here the δ_i 's are the censoring status. The idea of the cross validation, here *leave-one-out*-cross validation [13, ch. 17] is used, is to estimate the hazard function, $\lambda(x)$, with all of the data except for observation i . Then use this hazard rate to find the hazard rate for the observation left out. Let $\hat{\lambda}^{-i}(x|h)$ denote this hazard rate, where the smoothing parameter, h , is held constant. This leads to the likelihood cross-validation criterion

$$lCV(h) = \sum_{i=1}^n [\delta_i \ln \hat{\lambda}^{-i}(X_i|h) - \hat{\lambda}^{-i}(X_i|h)R_i]. \quad (4.28)$$

The optimal smoothing parameter is found by maximizing this expression for h . We could not find a function that did this for us, so we have done this in R, and the code is included as Appendix D.4. We used model (4.17) when estimating the hazard rates in the likelihood cross-validation criterion. The maximum value for $lCV(h)$ for each covariate was found using the `optimize` function in R. The results are seen in Table 1.

Covariate	20%		50%		80%	
	h	$lCV(h)$	h	$lCV(h)$	h	$lCV(h)$
z_{i1}	1.773136	-51.32165	3.788885	-29.50246	0.2693558	-19.21889
z_{i2}	1.951446	-81.43132	7.329702	-51.16201	8.88204	-22.36975
x_i	1.887488	-76.57993	1.463886	-46.68948	4.017604	-21.95032

Table 1: Maximum value for $lCV(h)$ with corresponding h for each of the covariates z_{i1} , z_{i2} and x_i for censoring levels of 20%, 50% and 80%.

One could suspect that the smoothing parameter, which is the length of each interval on the covariate axis we smooth over, should increase for increasing censoring levels. This is because there are fewer *events* in the artificial point process when the censoring level is high. This should lead to an increase in the smoothing parameter, in order to be able to cover a sufficient amount of events. See Figure 12 (this plot has nothing to do with the data set used, it is merely included for illustrational purposes). It is seen that the smoothing parameter for the covariate z_{i2} increases along with the censoring level, but this is not the case for z_{i1} and x_i . The smoothing parameter for z_{i1} increases from 20% to 50%, and decreases to 80% censoring. For x_i it decreases from 20% to 50%, and increases to 80%. In the specialization project [1] it was seen that a preliminary naïve

choice of 1 for the smoothing parameter also lead to nice results. When computing the $LCV(h)$, it was also noted that this expression diverged for some h 's, so it is uncertain how important this parameter is for the final results. We used the smoothing parameters in Table 1 when the Cox–Snell residuals from the empty model was the basis for the point process in the covariate order method. The functional form in (4.25) was estimated for each covariate for each censoring level.

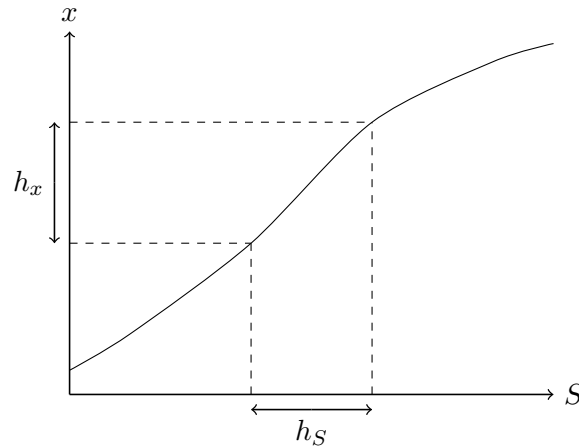


Figure 12: Figure illustrating the correspondance between the smoothing parameter on the covariate axis, x , and the event axis, S .

The plots of the functional form (Figure 13) of each covariate clearly suggests a linear function for z_1 for all censoring levels. z_2 's functional form varies for all censoring levels. The reason for this could be that according to the p-values in the summaries for both the misspecified and assumed correct model in Section 4.2.2.1, the coefficient of z_2 is not significantly different from zero, and hence the covariate's influence is negligible in the models. This may also be the reason for the scale being so different on the $f(\cdot)$ -axis for 80% censoring compared to the two lower censoring levels. The functional form of x is clearly "U"-shaped at 20% censoring, visible at 50%, but not as clear cut at 80%. The next step is to fit a model with functional form of the covariates as suggested by the plots for each censoring level. We fit the model

$$\ln Y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \beta_3 x_i^2 + \sigma W_i. \tag{4.29}$$

New Cox–Snell residuals are calculated,

$$\hat{R}_i = e^{\frac{\ln Y_i - \hat{\beta}_0 - \hat{\beta}_1 z_{i1} - \hat{\beta}_2 z_{i2} - \hat{\beta}_3 x_i^2}{\hat{\sigma}}}. \tag{4.30}$$

This model is believed to be correct. Which means that the hazard rate estimated with the covariate order method, or the inverse of the expected value of the Cox–Snell residuals, should not have any trends. We have chosen to look specifically at the hazard

rates estimated from the covariate x , since we know that this is the one that initially was misspecified. We again find the optimal smoothing parameter for each censoring level. Values can be found in Table 2.

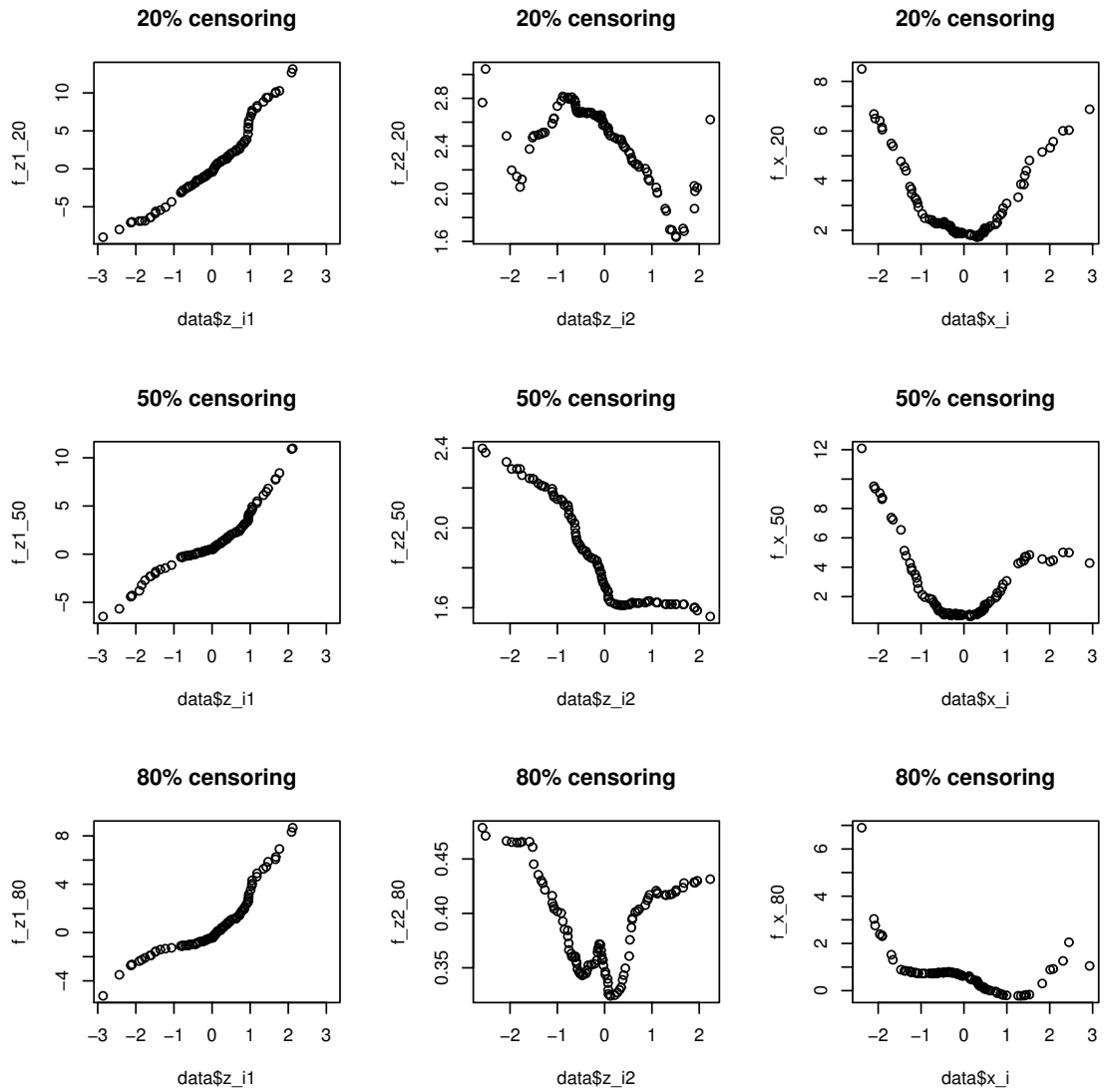


Figure 13: Plots of the estimated function from (4.25) for each covariate (z_1 , z_2 and x from left to right) for each censoring level (20%, 50% and 80% from top to bottom).

Censoring level	h	$lCV(h)$
20%	8.157306	-78.58567
50%	8.35743	-48.93229
80%	7.923083	-19.70174

Table 2: Maximum value for $lCV(h)$ with corresponding h for x_i^2 for censoring levels of 20%, 50% and 80%.

The hazard rates are estimated with the covariate order method. The plot of the logarithm of the hazard rate versus x for all censoring levels can be seen in Figure 14. If the model is correct, the logarithm of the hazard rate should be zero. From the plots we see that all the points lie close to zero, but it looks as though there is an increasing trend for small values of x_i^2 and a decreasing trend for large values. To check for trends more rigorously we perform a more formal trend test.

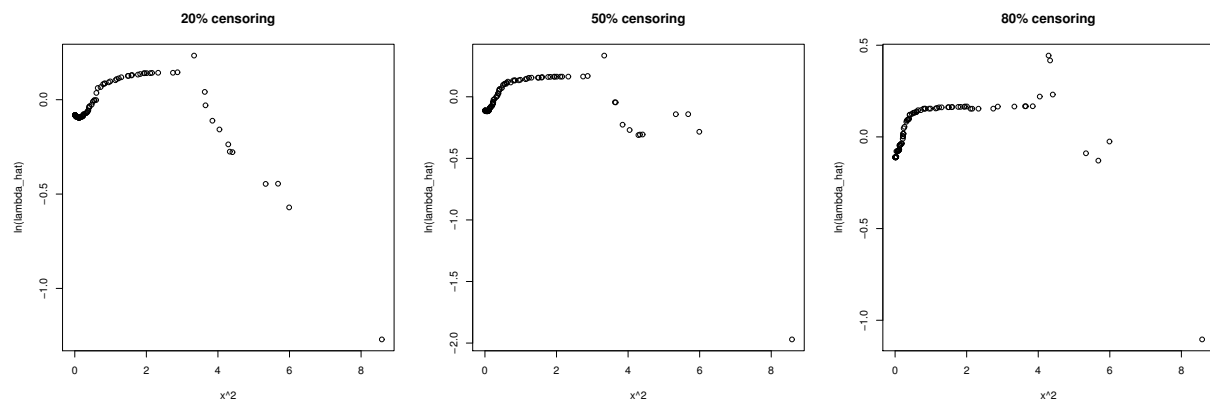


Figure 14: Plots of x_i^2 vs. $\ln(\hat{\lambda}(x_i^2))$ for 20%, 50% and 80% censoring (from left to right).

4.2.2.4 Statistical trend tests of the hazard rate

Three statistical tests of trend have been considered:

1. The Laplace test,
2. The military handbook test,
3. Trend test based on the Anderson–Darling statistic [14].

The Laplace test and the military handbook test are described in [6, p. 286-287]. These test the following null hypothesis, H_0 , against the alternative hypothesis, H_1 :

H_0 :“No trend”, i.e. that the artificial (point) poisson process is a homogeneous poisson process (HPP).

H_1 :“Monotonic trend”, i.e. that the artificial (point) poisson process is a non-homogeneous poisson process (NHPP) with *either* increasing or decreasing hazard rate.

The fact that these tests only test for monotonic trend is the reason why we decided not to use them. The increasing and decreasing characteristics in Figure 14 is clearly non-monotonic. The third test, the Anderson–Darling test, tests for both monotonic and non-monotonic trends as an alternative to the null hypothesis of no trend. The Anderson–Darling test used is described in [2]. The test statistic is given by

$$AD = -\frac{1}{\hat{k}} \left[\sum_{i=1}^{\hat{k}} (2i-1) \left(\ln \frac{S_i}{S} + \ln \left(1 - \frac{S_{\hat{k}+1-i}}{S} \right) \right) \right] - \hat{k}. \quad (4.31)$$

Here $S = \sum_{i=1}^n Y_i$ and \hat{k} is given by

$$\hat{k} = \begin{cases} k(n) & \text{if } S_{k(n)} < S, \\ k(n) - 1 & \text{if } S_{k(n)} = S. \end{cases}$$

This means that \hat{k} is dependent on whether or not the last interoccurrence time in the artificial point process is a censored or uncensored observation time, Y_i . The Anderson–Darling test statistic was calculated for each censoring level. Their values were found to be

$$\begin{aligned} AD_{x_i} &= 3.914077 \text{ at } 20\% \text{ censoring,} \\ AD_{x_i} &= 3.139462 \text{ at } 50\% \text{ censoring,} \\ AD_{x_i} &= 1.814149 \text{ at } 80\% \text{ censoring.} \end{aligned} \quad (4.32)$$

Our implementation of the Anderson–Darling statistic in R is found in Appendix D.2. The Anderson–Darling test is a one sided test, so we need to find our critical values. However, instead of using the critical value for the asymptotic null distribution of (4.31) given in [2], we used bootstrapping to estimate our own distribution for the Anderson–Darling test statistic. The reason for this is that we do not know if this asymptotic null distribution works for us, and using its critical value would therefore be erroneous.

4.2.2.5 Estimating the distribution of the Anderson–Darling statistic using bootstrapping

Before we do the actual bootstrapping, we give a brief description of the concept. Bootstrapping is a computer-intensive method used for statistical inference. It is used to

estimate properties of, in our case, the Anderson–Darling test statistic, but it can be used to estimate properties of other statistics as well. Our idea is to estimate the Anderson–Darling test statistic many times by using the observed data as basis for resampling. With resampling we *make a new* (i.e. bootstrap) data set from which we can calculate the Anderson–Darling test statistic. All these bootstrapped Anderson–Darling statistics estimate the distribution of the statistic. We will use this distribution to check if our statistics in (4.32) lead to the conclusion of significant trends in the hazard functions. Both a parametric and non-parametric bootstrapping method were used. The non-parametric bootstrapping method is obtained by resampling, with replacement, the observations in the data set. The parametric bootstrap samples are obtained by resampling residuals. This is done using the estimated $\hat{\sigma}$ from the fitted model in (4.29), and adding residuals, error terms, given by $\hat{\sigma}W$, to the observation times Y_i . Here W are drawn from the standard Gumbel distribution. We now have the bootstrapped observations $(Y_i^*, z_{i1}, z_{i2}, x_i)$, where Y_i^* denotes the bootstrapped observation time, for all i . For a more thorough description of bootstrapping, the reader is referred to [13]. We made, for each bootstrapping method, 10000 bootstrap samples. This to be sure that the critical values obtained would be sufficiently accurate. The critical values are found by sorting the six vectors of Anderson–Darling statistics, and then picking out the values in the 9500th position in each vector. We have one vector of bootstrapped Anderson–Darling statistics for each censoring level using both parametric and non-parametric bootstrapping. The values in the 9500th positions are the 95%-quantiles of the estimated distributions. Our code showing the implementation of the bootstrapping is included as Appendix D.3.

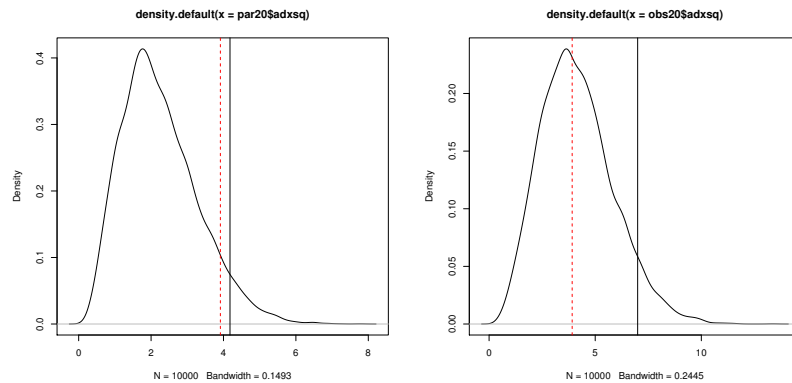
Censoring level	Statistic	95%-quantile	Observed statistic
20%	$AD_{x_i}^{par}$	4.17991	3.914077
	$AD_{x_i}^{non-par}$	7.20538	
50 %	$AD_{x_i}^{par}$	3.616888	3.139462
	$AD_{x_i}^{non-par}$	6.606602	
80%	$AD_{x_i}^{par}$	2.908031	1.814149
	$AD_{x_i}^{non-par}$	4.379813	

Table 3: Observed Anderson–Darling statistics and 95%-quantiles from the estimated distributions of the Anderson–Darling statistics.

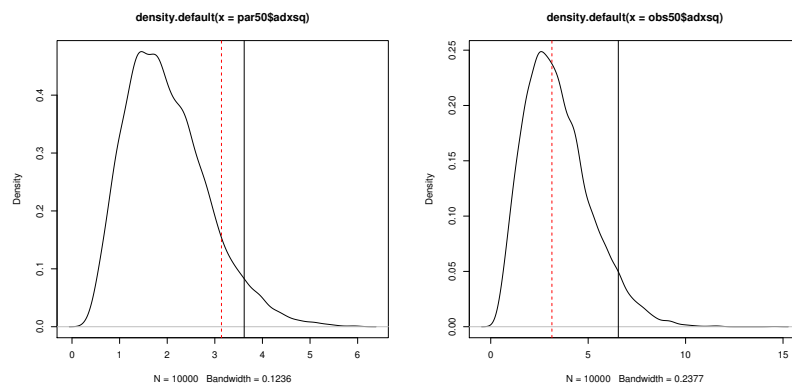
In Table 3 the critical values are listed. Here the superscript in AD^{sup} denotes if the estimated distribution came from the parametric or non-parametric bootstrapping. Comparing the critical values to the observed test statistics, we see that the observed test statistic for each covariate is always lower than the critical value, regardless of bootstrapping method. It is therefore safe to conclude that there is no significant trend in the hazard function. Which again means that the Cox–Snell residuals can be viewed as independent of the covariates. Thus we believe that the lifetime model in (4.29) is correct. However, we have to be careful not to be too obstinate here. Remember, it

was not obvious that the quadratic function was the underlying functional form of the x_i 's at 80% censoring (Figure 13). In Figure 15 the density plots of the bootstrapped distributions for the Anderson–Darling test statistics are displayed. The vertical lines are their respective observed test statistics. The distributions' scale is larger for the distributions from the non-parametric bootstrap. The reason for this is because of the censoring of the lifetimes. When adding “more residuals” to the censoring times causes the bootstrapped observation times to approach, in most cases, or maybe even surpass the real, unknown, lifetimes. This again causes the Cox–Snell residuals, in most cases, to be smaller, which again leads to smaller Anderson–Darling statistics.

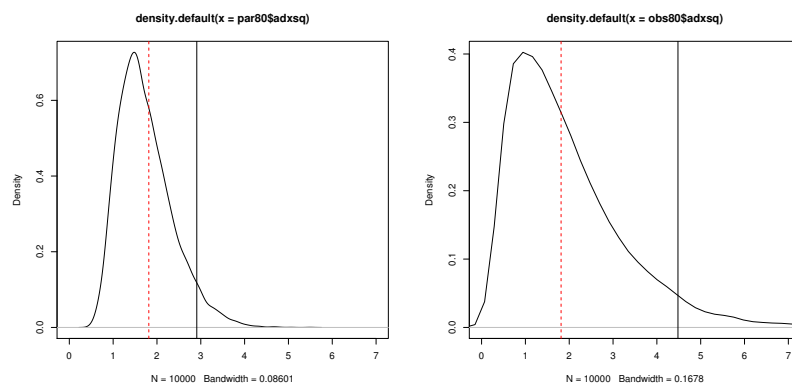
To conclude this section we plot the estimation of $f(x_i)$ for both methods, transformation and covariate order, along with the real function in Figure 16. Keep in mind, the *output* from using the covariate order method is the blue squares, while the *output* from the transformation method is the circles. Knowing this it is easy to say that the approach using the covariate order method is the best for retrieving the functional form of $f(x_i)$. Especially for high levels of censoring.



(a) Density plot of $AD_{x_i}^{par}$ for a censoring level of 20%. (b) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 20%.



(c) Density plot of $AD_{x_i}^{par}$ for a censoring level of 50%. (d) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 50%.



(e) Density plot of $AD_{x_i}^{par}$ for a censoring level of 80%. (f) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 80%.

Figure 15: Density plots of the six bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.

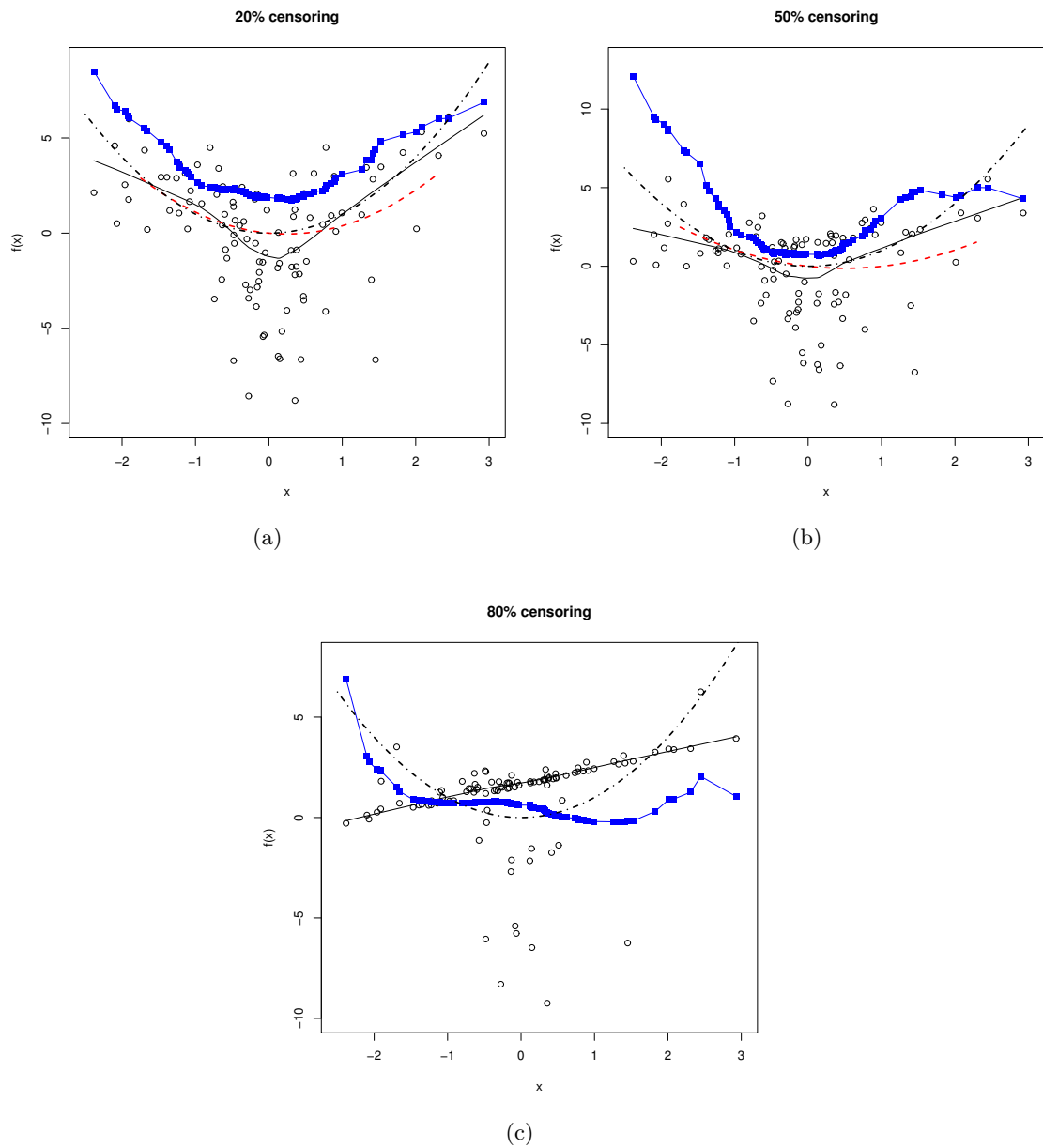


Figure 16: Plots of the estimation of $f(x_i)$ using the transformation method, covariate order method and the real function. The transformation method is illustrated by the circles, the black solid LOWESS line and the red dashed line (fitted quadratic function). The covariate order method is the blue squares connected by the blue line. The real function is the black dash-dot line.

5 Lognormally distributed lifetimes

Now we use lifetimes that are lognormally distributed with probability density function as in (2.19)

$$f_T(t) = \frac{1}{\sqrt{2\pi bt}} e^{-(\ln t - u)^2 / 2b^2}, \quad -\infty < w < \infty. \quad (5.1)$$

Here u is the log-scale parameter and b is the shape parameter. When simulating the model given by (3.1), we need to ensure that the T_i 's are lognormally distributed. This is done by using W_i 's which are from the standard normal distribution. The W_i 's have probability density function

$$f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \text{ for } -\infty < y < \infty. \quad (5.2)$$

By using the transformation formula for functions of random variables, we can show that the lifetime, T , in (3.1) is indeed lognormally distributed. Start by isolating W in (3.1) and transforming:

$$\begin{aligned} W &= \frac{\ln T}{\sigma} - \frac{\beta^T \mathbf{z} + f(x)}{\sigma} \\ \left| \frac{dW}{dT} \right| &= \frac{1}{bT} \\ f_T(t) &= f_W \left(\frac{\ln T - \beta^T \mathbf{z} - f(x)}{\sigma} \right) \cdot \left| \frac{dW}{dT} \right| \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2b^2} (\ln T - \beta^T \mathbf{z} - f(x))^2} \cdot \frac{1}{bT} \\ &= \frac{1}{\sqrt{2\pi bT}} e^{-\frac{1}{2b^2} (\ln T - (\beta^T \mathbf{z} + f(x)))^2}, \end{aligned} \quad (5.3)$$

which we recognize as a lognormally distributed variable, according to (5.1), with log-scale parameter $u = \beta^T \mathbf{z} + f(x)$ and shape parameter $b = \sigma$. The Cox-Snell residuals are given by

$$\hat{R}_i = -\ln \left[1 - F_W \left(\frac{\ln T_i - \hat{\beta}^T \mathbf{z}_i - \hat{\gamma} x_i}{\hat{\sigma}} \right) \right]. \quad (5.4)$$

Here $F_W(\cdot)$ is the CDF of lognormally distributed variables.

5.1 Coefficients minimizing the Kullback–Leibler distance

As for the Weibull distributed lifetimes, we want to find the coefficients minimizing the Kullback–Leibler distance for lognormally distributed lifetimes. For the Weibull distributed lifetimes we simulated from the model in (4.4):

$$\ln T_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + f(x_i) + \sigma W_i$$

The results from the data from this model showed that the linear terms, $\beta_1 z_{i1}$ and $\beta_2 z_{i2}$, were not important when trying to obtain the functional form of $f(x_i)$. In this section we will simulate lognormally distributed lifetimes from a model with only one covariate, which is misspecified.

$$\ln T_i = \beta_0 + f(x_i) + \sigma W_i, \quad (5.5)$$

with coefficients and function

$$\begin{aligned} \beta_0 &= 0, \\ f(x_i) &= x_i^2, \\ \sigma &= 2. \end{aligned} \quad (5.6)$$

Notice that σ and $f(x_i)$ are the same as for the censored Weibull data. The x_i 's are still standard normally distributed. We start with a complete data set with 1 million observations. When simulating these lifetimes from (5.5), we will again find, with almost exact precision, the coefficients minimizing the Kullback–Leibler, β_0^* , γ^* and σ^* , when fitting the misspecified model

$$\ln T_i = \beta_0^* + \gamma^* x_i + \sigma^* W_i. \quad (5.7)$$

5.1.1 Using Cox–Snell residuals to infer the form of $f(x_i)$

To find the functional form of $f(x_i)$ we want to use (3.7). However, as for the Weibull case, we will not know the values for β_0 and σ for real data set. So when solving these equations for $f(x_i)$, we will only be able to find an approximated expression, as we did in (4.10).

$$f(x_i) \sim \sigma^* \cdot \left[\Phi^{-1} \left(1 - e^{-\hat{R}_i^*} \right) \right] + \beta_0^* + \gamma^* x_i. \quad (5.8)$$

Here $\Phi^{-1}(\cdot)$ is the inverse of the standard normal CDF. 1 million lifetimes are simulated from the model in (5.5) and fitted as (5.7) with `survreg` in R. Cox–Snell residuals are calculated according to (5.4). These Cox–Snell residuals are used as \hat{R}_i^* in (5.8). The plot of the estimated $f(x_i)$'s are displayed in Figure 17(a). The function is blurred since because of the displacement of the points. The functional form is seen better, as for the Weibull lifetimes, in the smoothed line of the points. Still obtained by taking the mean value of all approximated values of $f(x_i)$ within intervals of length 0.1 on the x -axis.

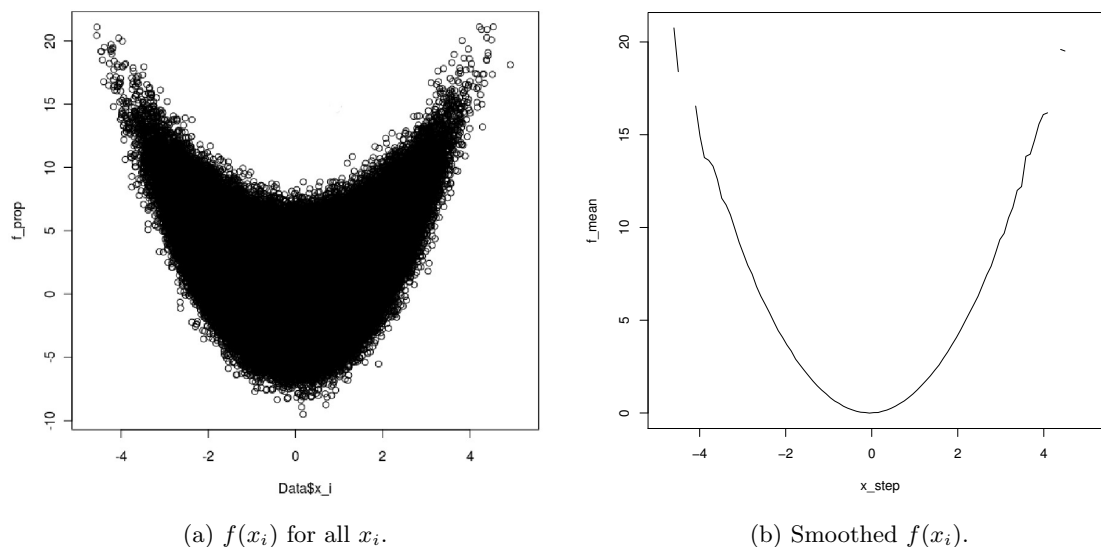


Figure 17: Plot of the estimation of the function $f(x_i)$ from (5.8).

5.1.2 Uncensored data set

As for the Weibull distributed lifetimes, we simulate 100 uncensored observations. This moderately sized, complete data set is used to see how the transformation method will work in practice for lognormally distributed lifetimes. The model is given by (5.5) and (5.6). Using `survreg` we fit the misspecified model

$$\ln T_i = \hat{\beta}_0 + \hat{\gamma}x_i + \hat{\sigma}W_i. \quad (5.9)$$

The summary from `survreg` was

```
> summary(lognormalMod)
```

Call:

```
survreg(formula = Surv(data$T_i) ~ data$x_i, data = data, dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	0.8109	0.2416	3.3564	7.90e-04
data\$x_i	0.0150	0.2478	0.0604	9.52e-01
Log(scale)	0.8777	0.0707	12.4119	2.25e-35

Scale= 2.41

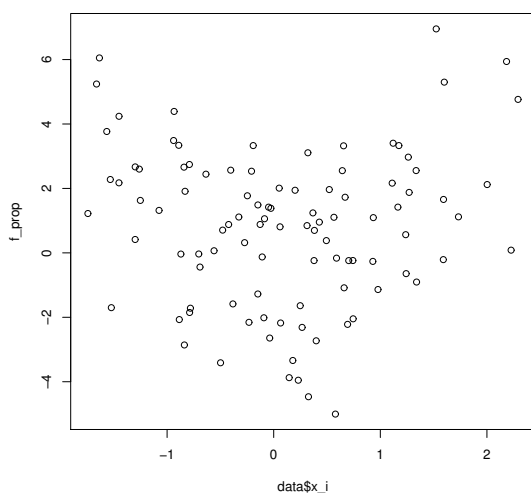
Log Normal distribution

```

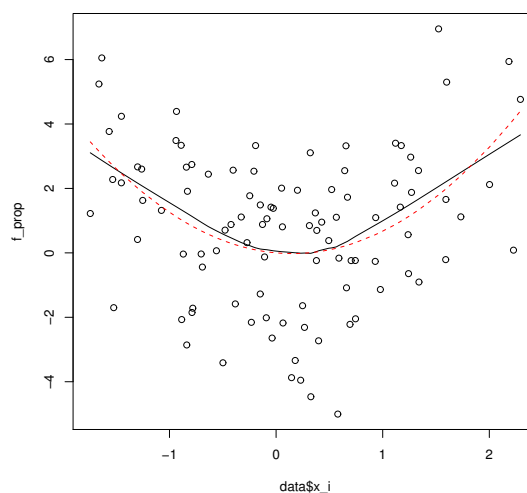
Loglik(model)= -310.9   Loglik(intercept only)= -310.9
Chisq= 0 on 1 degrees of freedom, p= 0.95
Number of Newton-Raphson Iterations: 1
n= 100

```

The linear term, `data$x_i`, is not significant. Which is the same observation as seen in the summary for the complete Weibull distributed data set. We calculate the Cox–Snell residuals according to (5.4), and use them to estimate $f(x_i)$ from (5.8). The results are displayed in Figure 18(a).



(a) Estimated $f(x_i)$.



(b) Estimated $f(x_i)$ with LOWESS and fitted quadratic function.

Figure 18: Plots of the estimation of the function $f(x_i)$ from (5.8) for the complete data set of lognormally distributed lifetimes. The solid black line in (b) is the LOWESS, the red dashed line is a fitted quadratic function.

By just looking at the dots, one observes the tendency of large values of $f(x_i)$ for large absolute values of x_i . By fitting the LOWESS line, as seen in Figure 18(b), we see tendencies of a “U”-shaped function characteristic for quadratic polynomials. A quadratic regression model was fitted with the function `lm` in R to see if this could be significant.

```
> summary(quadlognorm)
```

```
Call:
```

```
lm(formula = f_prop ~ data$x_i + I(data$x_i^2) - 1)
```

```
Residuals:
```

```
   Min       1Q   Median       3Q      Max
-5.158 -1.773  0.317  1.527  5.154
```

```
Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
data\$x_i	-0.2930	0.2280	-1.285	0.202
I(data\$x_i^2)	0.9662	0.1494	6.469	3.9e-09 ***

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.145 on 98 degrees of freedom
```

```
Multiple R-squared:  0.3001, Adjusted R-squared:  0.2858
```

```
F-statistic: 21.01 on 2 and 98 DF,  p-value: 2.547e-08
```

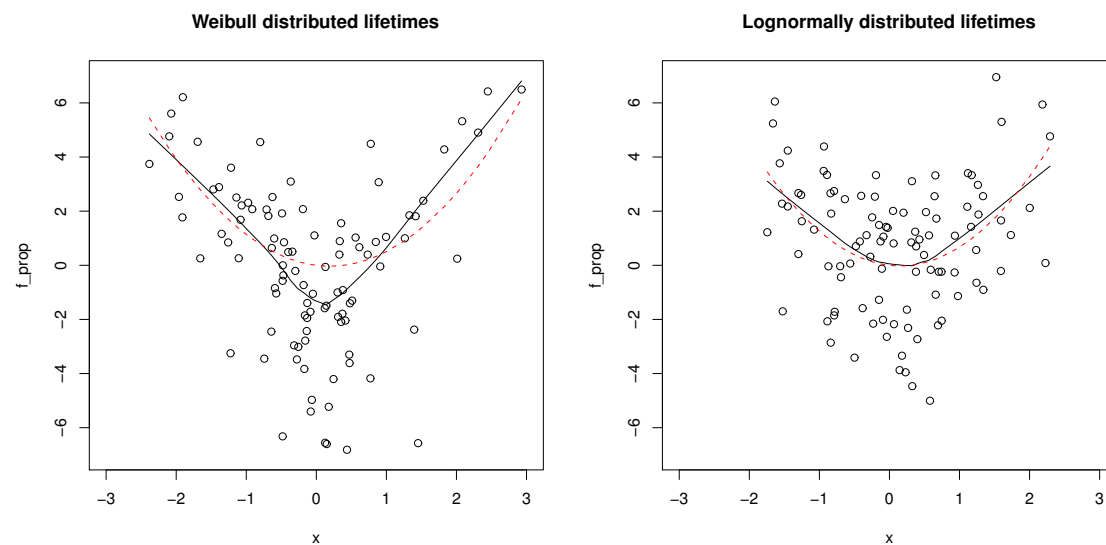


Figure 19: Plots comparing the estimation of the function $f(x_i)$ for Weibull and lognormally distributed lifetimes.

From the summary we see that a quadratic term is significant. Figure 18 is the lognormal equivalent to Figure 7 for Weibull distributed lifetimes. Immediately it does not look like one is any better the other, but if we plot them again with the same scale we see a difference. Figure 19 shows the two plots together. We see that the points are more scattered for the Weibull distributed lifetimes, and this leads to a poorer estimation of $f(x_i)$. However, this is believed to be caused by the variance of Cox–Snell residuals from the two different lifetime distributions. The variance is larger for the Cox–Snell residuals from the lifetimes from the Weibull distribution ($1.21 > 0.96$), since the variance of the actual lifetimes are alot larger ($3.3 \cdot 10^{16} \gg 1.4 \cdot 10^4$). Another moment is that the x 's for the Weibull case are a bit more spread than the for the lognormal case, even though they are drawn from the same distribution.

5.1.3 Censored data sets

We continue with the 100 observations from the complete data set, but we draw censoring times for approximately 20%, 50% and 80% censoring levels. The independent censoring times, C_i , were drawn from the exponential distribution with $\lambda_{20} = 3.0 \cdot 10^{-2}$, $\lambda_{50} = 1/4$ and $\lambda_{80} = 2$ for ascending censoring levels. The data set is listed in Appendix B.

As in Section 4.2.2.1 we use `survreg` to fit the misspecified model

$$\ln Y_i = \beta_0 + \gamma x_i + \sigma W_i, \quad (5.10)$$

for all three censoring levels. Here $Y_i = \min(T_i, C_i)$ The summaries from `survreg` were

```
> summary(lognormalMod20)
```

Call:

```
survreg(formula = Surv(data$Y20_i, data$s20_i) ~ data$x_i, data = data,
        dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	0.9063	0.2595	3.493	4.78e-04
data\$x_i	-0.0607	0.2655	-0.229	8.19e-01
Log(scale)	0.9113	0.0817	11.154	6.85e-29

Scale= 2.49

Log Normal distribution

Loglik(model)= -217.6 Loglik(intercept only)= -217.7

Chisq= 0.05 on 1 degrees of freedom, p= 0.82

Number of Newton-Raphson Iterations: 3

n= 100

```
> summary(lognormalMod50)
```

```
Call:
```

```
survreg(formula = Surv(data$Y50_i, data$s50_i) ~ data$x_i, data = data,
        dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	1.017	0.325	3.128	1.76e-03
data\$x_i	-0.196	0.314	-0.623	5.33e-01
Log(scale)	0.959	0.107	8.937	4.01e-19

```
Scale= 2.61
```

```
Log Normal distribution
```

```
Loglik(model)= -105.1 Loglik(intercept only)= -105.3
```

```
Chisq= 0.39 on 1 degrees of freedom, p= 0.53
```

```
Number of Newton-Raphson Iterations: 3
```

```
n= 100
```

```
> summary(lognormalMod80)
```

```
Call:
```

```
survreg(formula = Surv(data$Y80_i, data$s80_i) ~ data$x_i, data = data,
        dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	0.955	0.586	1.629	1.03e-01
data\$x_i	0.182	0.411	0.444	6.57e-01
Log(scale)	0.949	0.175	5.411	6.26e-08

```
Scale= 2.58
```

```
Log Normal distribution
```

```
Loglik(model)= -26.1 Loglik(intercept only)= -26.2
```

```
Chisq= 0.2 on 1 degrees of freedom, p= 0.66
```

```
Number of Newton-Raphson Iterations: 4
```

```
n= 100
```

We see the same “error” in the lognormal models as in the Weibull models. The linear term $\text{data}\$x_i = x_i$ is insignificant for all censoring levels. The models solution to cope with this is again to add a large $(\text{Intercept}) = \hat{\beta}_0$ value.

The Cox–Snell residuals for each model were calculated from

$$\hat{R}_i = -\ln \left[1 - \Phi \left(\frac{\ln Y_i - \hat{\beta}^T z_i - \hat{\gamma} x_i}{\hat{\sigma}} \right) \right], \quad (5.11)$$

for each censoring level. Here $\Phi(\cdot)$ is the CDF of the standard normal distribution. Assuming that the model is correct, we do the same as in Section 4.2.2.1 for the censored observations. We add 1 to the Cox–Snell residuals from censored observations to compensate for the observation times being censored. A plot of the logarithm of the Cox–Snell residuals versus x_i can be seen in Figure 20. The dots does not appear to be randomly scattered for any of the censoring levels. For 20% censoring the logarithm of the Cox–Snell residuals are particularly small for small values of x_i , indicating that the covariate may be misspecified. This can be seen for 50% as well. For 80% one cannot draw any solid conclusions for the dots coming from uncensored observations. Here the logarithm of the residuals from censored observations are close to zero. The cause of this is the same as for Weibull distributed lifetimes. The censoring times are small, and adding one to the residuals causes them to approximate to one.

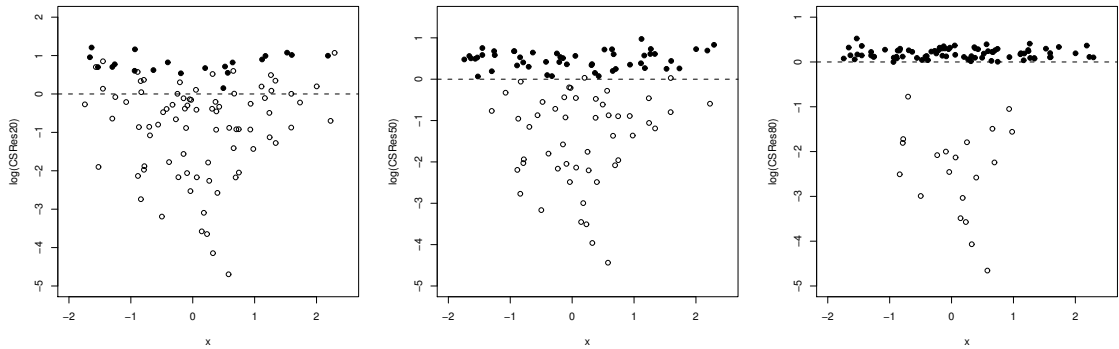


Figure 20: Plot of the logarithm of the Cox–Snell residuals versus x_i for censoring levels of 20%, 50% and 80% (left to right). Circles from uncensored observations, solid dots from censored observations.

We continue our investigation to find the functional form of $f(x_i)$ by using (5.8). We use the computed values for the Cox–Snell residuals for \hat{R}_i^* and the model parameters $\hat{\beta}_0$, $\hat{\gamma}$, $\hat{\sigma}$ for β_0^* , γ^* , σ^* . The results can be seen in Figure 21. The LOWESS line was fitted to each censoring level, suggesting a quadratic function for the two lowest censoring levels. For 80% censoring the plot is inconclusive, as it was for Weibull distributed lifetimes.

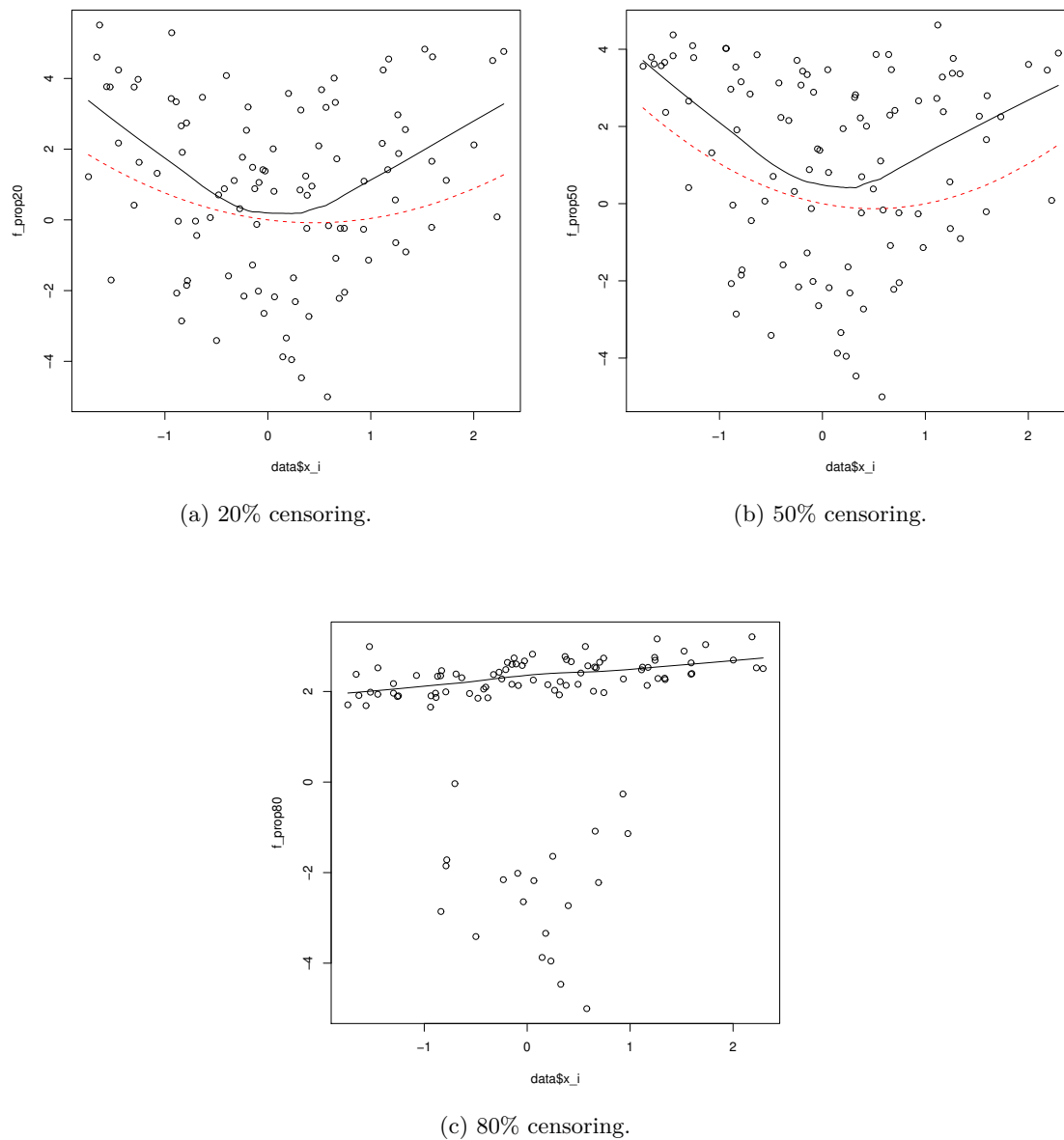


Figure 21: Plot of the estimation of the function $f(x_i)$ with LOWESS (black solid line) and quadratic regression (red dashed line) estimates for the different censoring levels for lognormally distributed data.

For 20% and 50% censoring we fit a quadratic regression model:

```
> summary(quad20lognorm)
```

Call:

```
lm(formula = f_prop20 ~ data$x_i + I(data$x_i^2) - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.119	-1.692	0.526	1.665	4.115

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
data\$x_i	-0.3599	0.2398	-1.501	0.137
I(data\$x_i^2)	0.9647	0.1571	6.140	1.77e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.257 on 98 degrees of freedom

Multiple R-squared: 0.2779, Adjusted R-squared: 0.2631

F-statistic: 18.85 on 2 and 98 DF, p-value: 1.182e-07

```
> summary(quad50lognorm)
```

Call:

```
lm(formula = f_prop50 ~ data$x_i + I(data$x_i^2) - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.0553	-1.5872	0.2982	2.1105	3.8897

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
data\$x_i	-0.5211	0.2406	-2.165	0.0328 *
I(data\$x_i^2)	1.0530	0.1576	6.679	1.46e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.264 on 98 degrees of freedom

Multiple R-squared: 0.3137, Adjusted R-squared: 0.2997

F-statistic: 22.4 on 2 and 98 DF, p-value: 9.745e-09

The summary of the quadratic regression model for the censoring level of 20% (`quad20lognorm`) suggests that only a quadratic term is significant. For 50% the linear term is also significant, which we know is not the case. This again indicates that the transformation method is not good for high levels of censoring. However, we continue as for the Weibull distributed lifetimes, and respecify the models for 20% and 50% censoring

$$\begin{aligned} \ln Y_i &= \beta_0 + \gamma_1 x_i^2 + \sigma W_i, \text{ for 20\% censoring,} \\ \ln Y_i &= \beta_0 + \gamma_2 x_i + \gamma_3 x_i^2 + \sigma W_i, \text{ for 50\% censoring.} \end{aligned} \quad (5.12)$$

The models were fitted with `survreg`:

```
> summary(lognormalMod20_2)
```

Call:

```
survreg(formula = Surv(data$Y20_i, data$s20_i) ~ I(data$x_i^2),
        data = data, dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	0.0256	0.3001	0.0852	9.32e-01
I(data\$x_i^2)	0.8990	0.2069	4.3454	1.39e-05
Log(scale)	0.8103	0.0814	9.9489	2.55e-23

Scale= 2.25

Log Normal distribution

Loglik(model)= -208.8 Loglik(intercept only)= -217.7

Chisq= 17.65 on 1 degrees of freedom, p= 2.7e-05

Number of Newton-Raphson Iterations: 4

n= 100

```
> summary(lognormalMod50_2)
```

Call:

```
survreg(formula = Surv(data$Y50_i, data$s50_i) ~ data$x_i + I(data$x_i^2),
        data = data, dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	-0.112	0.340	-0.329	7.42e-01
data\$x_i	-0.703	0.339	-2.074	3.81e-02
I(data\$x_i^2)	1.279	0.299	4.280	1.87e-05
Log(scale)	0.832	0.105	7.894	2.93e-15

Scale= 2.3

Log Normal distribution

Loglik(model)= -94.6 Loglik(intercept only)= -105.3

Chisq= 21.51 on 2 degrees of freedom, p= 2.1e-05

Number of Newton-Raphson Iterations: 4

n= 100

The coefficients are estimated better for the respecified models in (5.12) than for the initial misspecified models in (5.10). This is best seen in the (Intercept) and Scale coefficients. New Cox–Snell residuals were computed and log-plotted versus the x_i 's in Figure 22. It still looks like the logarithm of the residuals have large values for large absolute values of x_i , but perhaps not as clear as the plot from the original misspecified model in Figure 20.

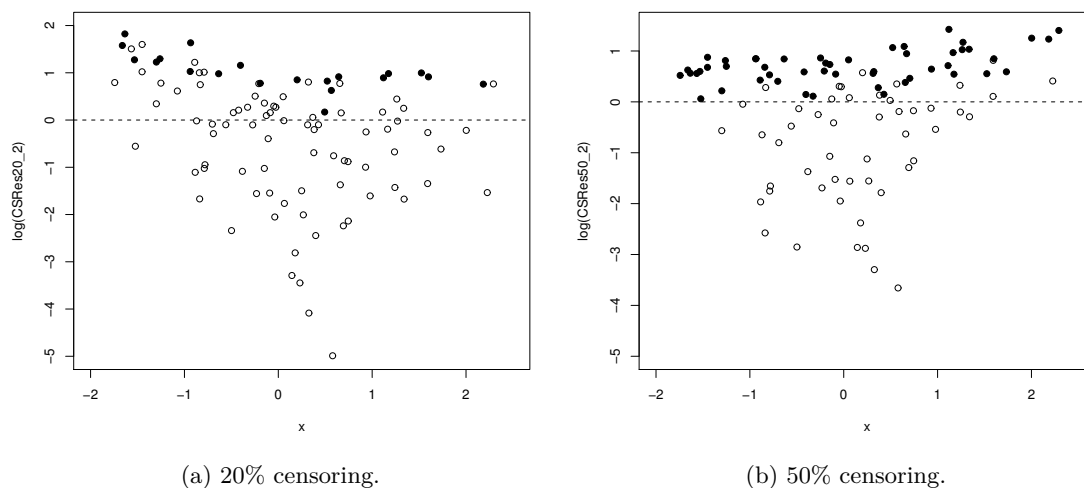
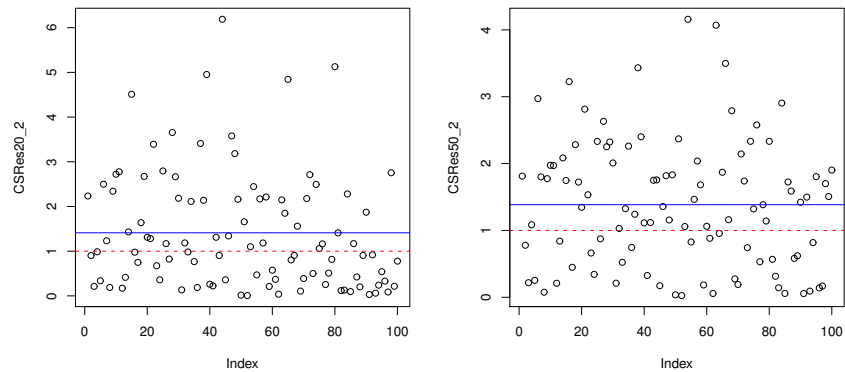


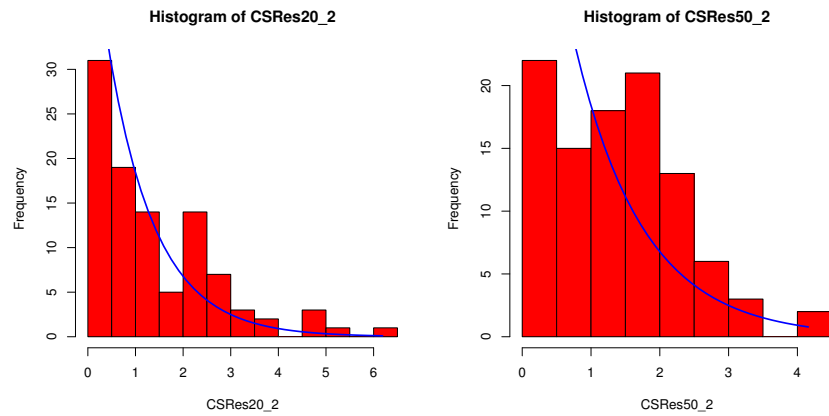
Figure 22: Plot of the logarithm of the Cox–Snell residuals versus x_i for censoring levels of 20% and 50% from the respecified models in (5.12). Circles from uncensored observations, solid dots from censored observations.

To check if our respecified models are correct we investigate the Cox–Snell residuals more formally. When a model is correct these residuals should be unit exponentially distributed. We start by plotting them along with their theoretical mean of one. This plot is displayed in part (a) of Figure 23. We see that the estimated mean value, the blue solid line, does not coincide with the theoretical one, the red dashed line. The difference is larger for the model with 50% censoring than for the one with 20%. In part (b) we have fitted the Cox–Snell residuals into histograms. If the residuals truly were unit exponentially distributed the histogram should follow the solid blue line, which is the theoretical density function. Neither of the plots fit, but the one for 20% censoring

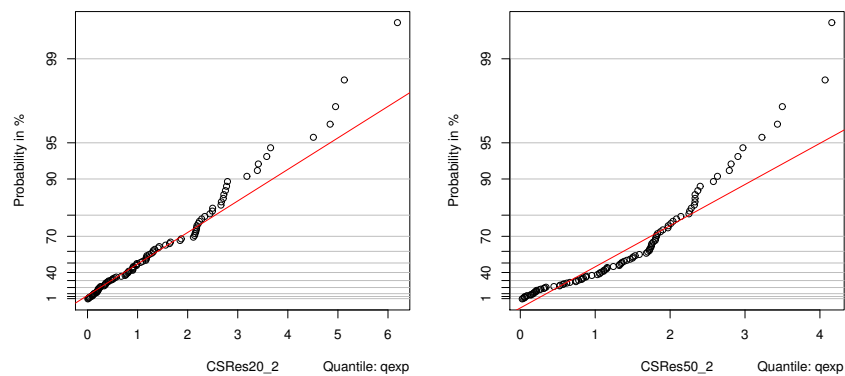
is remarkably better than the one for 50%. Part (c) shows the probability plots for both censoring levels. Again, the plot coming from the model with 20% censoring fits best, but it is not a very good fit. This leads to the conclusion that the residuals are not unit exponentially distributed, and hence the respecified model is not the correct one either. Then again, at least for the model with 20% censoring, we know that the model is correct. So the poor results could simply come from the fact that the fitted model does not fit the data as good as we could have hoped for.



(a) Mean plot.



(b) Density plot.



(c) Probability plot.

Figure 23: Plots checking if the Cox–Snell residuals are unit exponentially distributed. 20% censoring in the left column, 50% censoring in the right. (a) is a plot of the Cox–Snell residuals from the models in (5.12) with theoretical mean (red dashed line) and estimated mean (blue solid line). (b) is a histogram of the Cox–Snell residuals with theoretical density function. (c) is the unit exponential probability plot. 1 is added to the censored residuals.

5.1.3.1 Covariate order method

Now we turn to the covariate order method. Starting by fitting an empty model, a model without any covariates, using `survreg`, for all censoring levels.

$$\ln Y_i = \beta_0 + \sigma W_i. \quad (5.13)$$

From the regression we get estimates for $\hat{\beta}_0$ and $\hat{\sigma}$. The Cox–Snell residuals are for the empty model

$$\hat{R}_i = -\ln \left[1 - \Phi \left(\frac{\ln Y_i - \hat{\beta}_0}{\hat{\sigma}} \right) \right], \quad (5.14)$$

which, under the assumption that the empty model is correct, should be unit exponentially distributed. The residuals are used as the response variable when estimating the hazard function, $\hat{\lambda}(x)$, with the covariate order method. Suspecting that the empty model is not the correct model, we can assume that the correct model is given by

$$\ln Y_i = \beta_0 + f(x_i) + \sigma W_i. \quad (5.15)$$

The expected value of the Cox–Snell residuals, \hat{R}_i , in (5.14), are estimated as $1/\hat{\lambda}(x_i)$, where $\hat{\lambda}(x_i)$ is estimated by the covariate order method

$$E[\hat{R}_i] = E \left[-\ln \left[1 - \Phi \left(\frac{\ln Y_i - \hat{\beta}_0}{\hat{\sigma}} \right) \right] \right] \approx \frac{1}{\hat{\lambda}(x_i)}, \quad (5.16)$$

Estimating and solving for $f(x_i)$, gives us the relationship $f(x_i)$ has with the known coefficients and covariates.

$$f(x_i) \approx \hat{\sigma} \cdot \left[\Phi^{-1} \left(1 - e^{-1/\hat{\lambda}(x_i)} \right) \right] + \hat{\beta}_0. \quad (5.17)$$

Before using the covariate order method we need to find the optimal smoothing parameters. This is done in the same way as for Weibull distributed lifetimes in Section 4.2.2.3, i.e. using a likelihood cross-validation criterion. The optimal smoothing parameters for the different censoring levels are listed in Table 4, together with their corresponding maximal $lCV(h)$. We still smooth over the covariate axis.

Censoring level	h	$lCV(h)$
20%	1.399745	-76.09383
50%	1.351782	-44.84995
80%	2.600193	-19.53924

Table 4: Maximum value for $lCV(h)$ with corresponding h for the covariate x_i for censoring levels of 20%, 50% and 80%.

When using the optimal smoothing parameters for each censoring level with the covariate order method, we get the plots in Figure 24 estimating the functional form of $f(x_i)$. We see that the functional form is estimated better with this method than with the transformation method, albeit its accuracy deteriorates for increasing censoring levels. The function increases too rapidly when the absolute value of x_i increases.

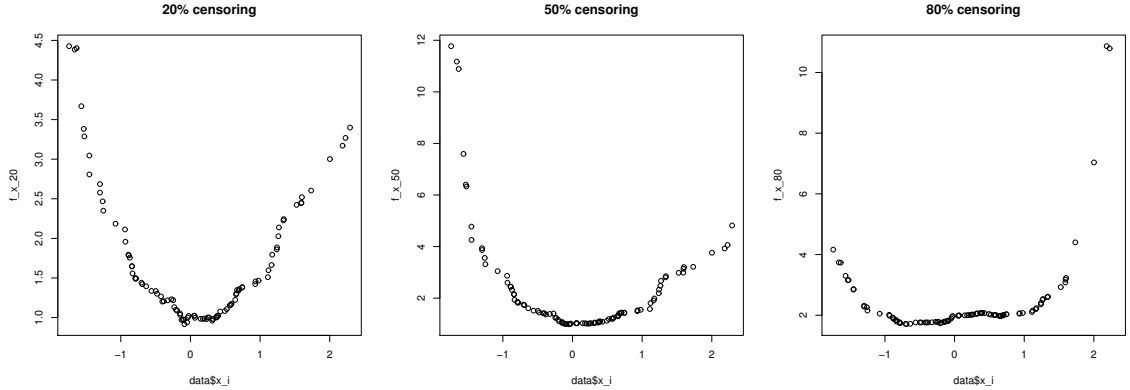


Figure 24: Estimation of the $f(x_i)$ using the covariate order method for censoring levels of 20%, 50% and 80%.

Assuming that $f(x_i) \propto x_i^2$, we fit new models with `survreg` where this covariate is included. We fit the model

$$\ln Y_i = \beta_0 + \gamma x_i^2 + \sigma W_i, \quad (5.18)$$

for all censoring levels. Their Cox–Snell residuals are estimated with the model coefficients

$$\hat{R}_i = -\ln \left[1 - \Phi \left(\frac{\ln Y_i - \hat{\beta}_0 - \hat{\gamma} x_i^2}{\hat{\sigma}} \right) \right] \quad (5.19)$$

To test if this model is correct, we check if the hazard rate estimated with the covariate order method has any trends. The model is believed to be correct if we cannot find any trends. We start by finding the optimal smoothing parameter using the same likelihood cross-validation criterion as before (Section 4.2.2.3). The optimal smoothing parameters, h , together with their maximal values for $lCV(h)$ are listed in Table 5.

Censoring level	h	$lCV(h)$
20%	4.882172	-78.61886
50%	12.74270	-49.50710
80%	12.74270	-21.60141

Table 5: Maximum value for $lCV(h)$ with corresponding h for the covariate x_i^2 for censoring levels of 20%, 50% and 80%.

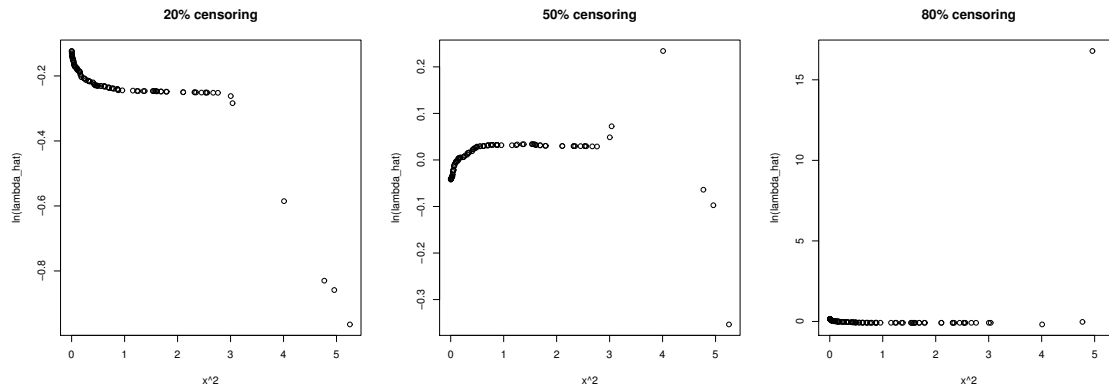


Figure 25: Plots of x_i^2 vs $\ln(\hat{\lambda}(x_i^2))$ for 20%, 50% and 80% censoring (from left to right).

The hazard rates, from the covariate order method, are displayed in Figure 25. The plot shows that the hazard rate is decreasing for small values of x_i^2 with 20% censoring and increasing hazard rate for small x_i^2 with 50% censoring. For 80% censoring it looks as there is no trend, except for the largest x_i^2 having an abnormally high hazard rate. When we look at the scale on the $\ln(\hat{\lambda}(x_i^2))$ -axis it is seen to be much larger for the 80% censoring plot than the two others. With this in mind it can look like this hazard rate is also decreasing for small values of x_i^2 .

We continue with the formal trend test used for the Weibull distributed data, the Anderson–Darling test. Using (4.31), the test statistics were found to be

$$\begin{aligned}
 AD_{x_i} &= 1.245602 \text{ at } 20\% \text{ censoring,} \\
 AD_{x_i} &= 0.4133454 \text{ at } 50\% \text{ censoring,} \\
 AD_{x_i} &= 0.9418295 \text{ at } 80\% \text{ censoring.}
 \end{aligned}
 \tag{5.20}$$

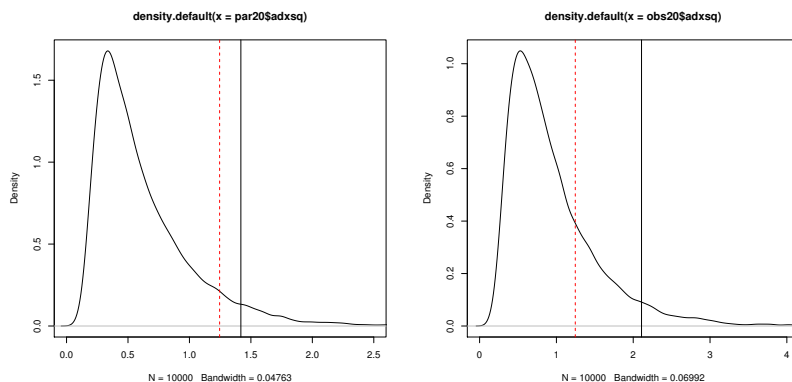
The next step is to bootstrap the observations to find distributions for the Anderson–Darling statistics. Then we can check if our observed statistics are smaller than the critical values. This is done analogously as in Section 4.2.2.5, with both parametric and non-parametric bootstrapping. The critical values, 95%-quantiles, are listed in Table

6, while the bootstrapped densities, along with the critical and observed values of the Anderson–Darling statistic, can be seen in Figure 26. From the table and figures we see that all the observed statistics are smaller than the corresponding critical values for both parametric and non-parametric bootstrapped densities. The critical values for the non-parametric bootstrap is seen to be higher than the values for the parametric, as it was also seen for the Weibull distributed lifetimes. This leads to the conclusion that the model in (5.18) is correct for all censoring levels.

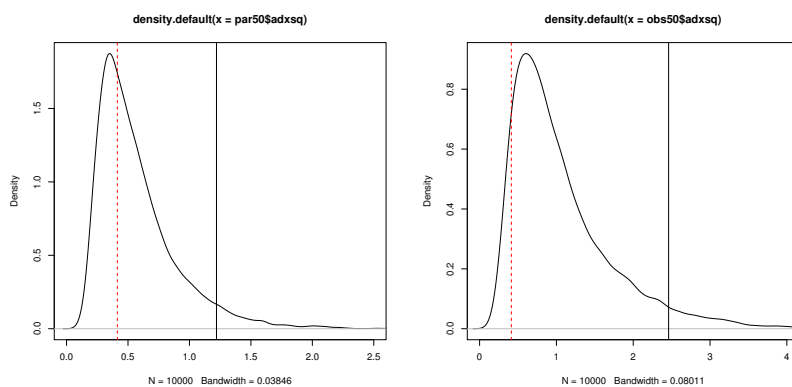
Censoring level	Statistic	95%-quantile	Observed statistic
20%	$AD_{x_i}^{par}$	1.418442	1.245602
	$AD_{x_i}^{non-par}$	2.108066	
50 %	$AD_{x_i}^{par}$	1.220073	0.4133454
	$AD_{x_i}^{non-par}$	2.460707	
80%	$AD_{x_i}^{par}$	1.240678	0.9418295
	$AD_{x_i}^{non-par}$	3.126138	

Table 6: Observed Anderson–Darling statistics and 95%-quantiles of the estimated distributions of the Anderson–Darling statistics.

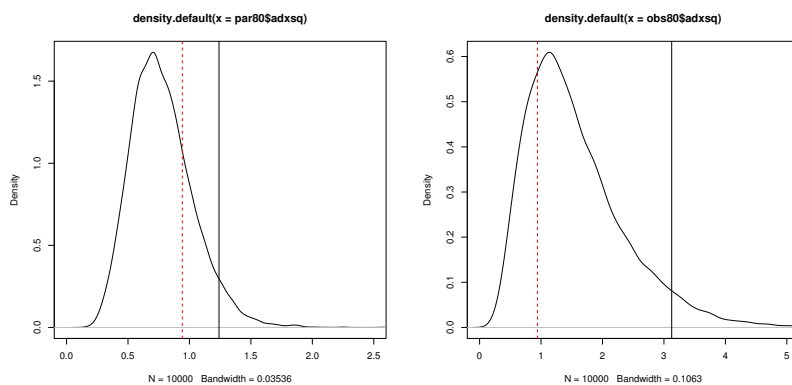
We conclude this section with a plot of the estimation of $f(x_i)$ for both methods, transformation and covariate order, along with the real function in Figure 27. The *output* from using the covariate order method is the blue squares, while the *output* from the transformation method is the circles. As it was for Weibull distributed lifetimes, the covariate order method is the best approach for retrieving the functional form of $f(x_i)$ when the lifetimes are lognormally distributed.



(a) Density plot of $AD_{x_i}^{par}$ for a censoring level of 20%. (b) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 20%.



(c) Density plot of $AD_{x_i}^{par}$ for a censoring level of 50%. (d) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 50%.



(e) Density plot of $AD_{x_i}^{par}$ for a censoring level of 80%. (f) Density plot of $AD_{x_i}^{non-par}$ for a censoring level of 80%.

Figure 26: Density plots of the six bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.

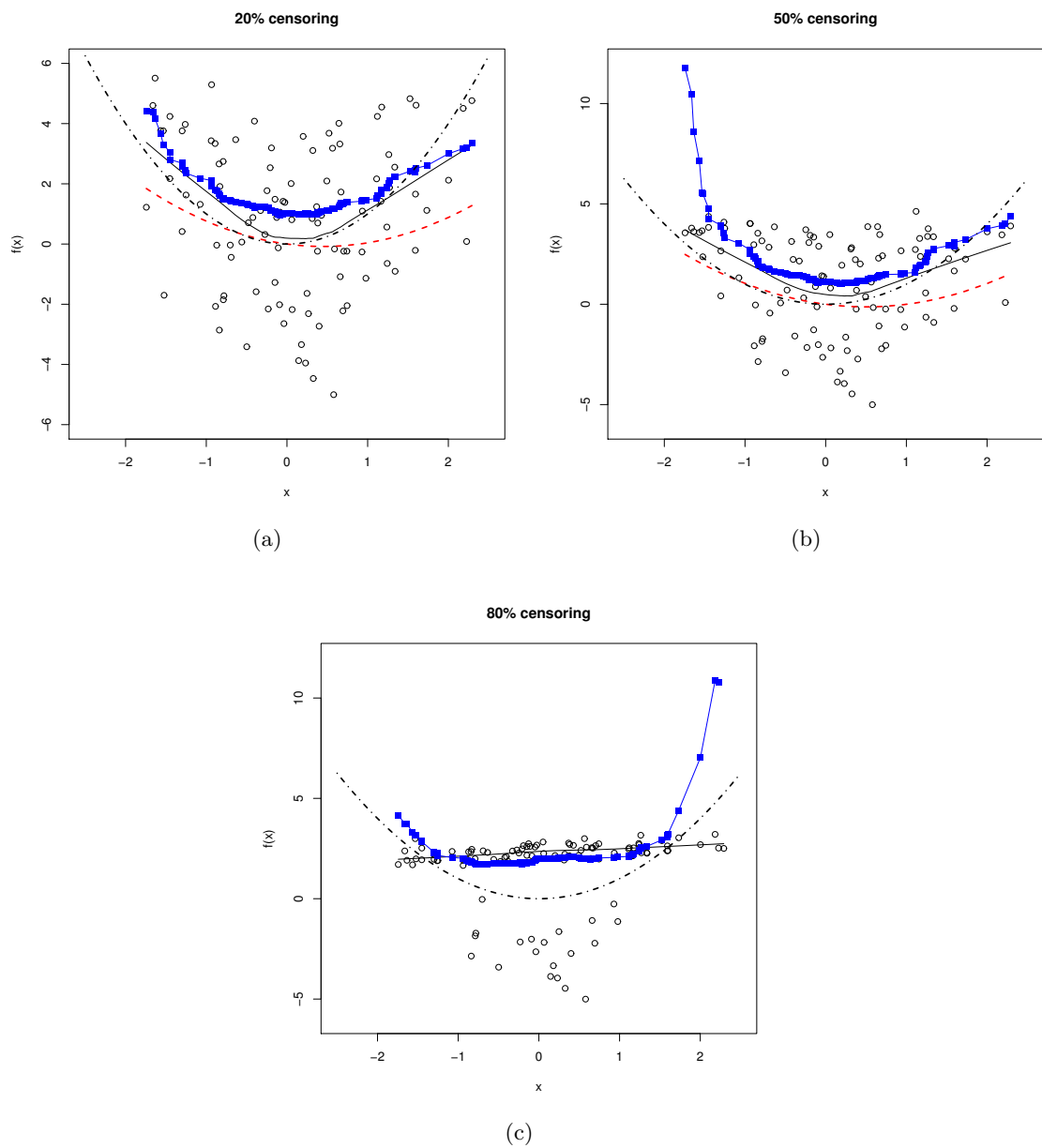


Figure 27: Plots of the estimation of $f(x_i)$ using the transformation method, covariate order method and the real function. The transformation method is illustrated by the circles, the black solid LOWESS line and the red dashed line (fitted quadratic function). The covariate order method is the blue squares connected by the blue line. The real function is the black dash-dot line.

6 Analytical Kullback–Leibler

In this section we will investigate the possibility of obtaining the coefficients minimizing the Kullback–Leibler distance analytically. Previously we have used a simulated data set with an unrealistic number of observations to find the coefficients. Finding the coefficients analytically is mainly a curiosity, but it would also be reassuring to see that the theory is consistent with what we have seen in practice. In general the expression in equation (3.3) can be written as

$$\begin{aligned}
 KL(\hat{\boldsymbol{\theta}}) &= E \left[\ln \frac{g(T_1, \dots, T_n)}{h(T_1, \dots, T_n, \hat{\boldsymbol{\theta}})} \right] = E \left[\ln g(T_1, \dots, T_n) - \ln h(T_1, \dots, T_n, \hat{\boldsymbol{\theta}}) \right], \\
 &= E \left[\ln [g(T_1) \cdot \dots \cdot g(T_n)] - \ln [h(T_1, \hat{\boldsymbol{\theta}}) \cdot \dots \cdot h(T_n, \hat{\boldsymbol{\theta}})] \right], \\
 &= E \left[\sum_{i=1}^n \left\{ \ln g(T_i) - \ln h(T_i, \hat{\boldsymbol{\theta}}) \right\} \right], \\
 &= \sum_{i=1}^n \left\{ E \left[\ln g(T_i) - \ln h(T_i, \hat{\boldsymbol{\theta}}) \right] \right\}.
 \end{aligned} \tag{6.1}$$

The expectation is taken with respect to the true model, $g(\cdot)$. $g(\cdot)$ is given by the lifetimes from the model

$$\ln T = f_1(x_1) + f_2(x_2) + \dots + f_p(x_p) + \sigma W. \tag{6.2}$$

Where the $f_i(x_i)$'s are unknown. $h(\cdot)$ is given by the lifetimes from the misspecified model where linearity is assumed.

$$\ln T = \hat{\boldsymbol{\beta}}^T \mathbf{x} + \hat{\sigma} W. \tag{6.3}$$

Here $\hat{\boldsymbol{\beta}} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p\}$ and $\mathbf{x} = \{1, x_1, x_2, \dots, x_p\}$. We will use the following true and misspecified model, respectively $h(\cdot)$ and $g(\cdot)$, in this section.

$$\ln T_i = \beta_0 + x_i^2 + \sigma W_i. \tag{6.4}$$

$$\ln T_i = \hat{\beta}_0 + \hat{\gamma} x_i + \hat{\sigma} W_i. \tag{6.5}$$

6.1 Lognormal

We start with lognormally distributed lifetimes. Here $g(\cdot)$ and $h(\cdot)$ are given, respectively, by

$$\begin{aligned} g(T_i) &= \frac{1}{\sqrt{2\pi}\sigma T_i} e^{-\frac{1}{2\sigma^2}(\ln T_i - (\beta_0 + x_i^2))^2}, \\ h(T_i, \hat{\boldsymbol{\theta}}) &= \frac{1}{\sqrt{2\pi}\hat{\sigma} T_i} e^{-\frac{1}{2\hat{\sigma}^2}(\ln T_i - (\hat{\beta}_0 + \hat{\gamma}x_i))^2}. \end{aligned} \quad (6.6)$$

Inserting this into (6.1) for a given i

$$\begin{aligned} KL(\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^n \left\{ E \left[\ln \frac{\hat{\sigma}}{\sigma} + (\ln T_i)^2 \cdot \left(\frac{1}{2\hat{\sigma}^2} - \frac{1}{2\sigma^2} \right) + \ln T_i \cdot \left(\frac{\beta_0 + x_i^2}{\sigma^2} - \frac{\hat{\beta}_0 + \hat{\gamma}x_i}{\hat{\sigma}^2} \right) \right. \right. \\ &\quad \left. \left. - \frac{(\beta_0 + x_i^2)^2}{2\sigma^2} + \frac{(\hat{\beta}_0 + \hat{\gamma}x_i)^2}{2\hat{\sigma}^2} \right] \right\} \end{aligned} \quad (6.7)$$

Here we need to find $E[\ln T_i]$ and $E[(\ln T_i)^2]$. For $E[\ln T_i]$, we know that W_i is standard normal distributed.

$$E[\ln T_i] = E[\beta_0 + x_i^2 + \sigma W_i] = \beta_0 + x_i^2 + \sigma E[W_i] = \beta_0 + x_i^2 + \sigma \cdot 0 = \beta_0 + x_i^2, \quad (6.8)$$

Then for $E[(\ln T_i)^2]$.

$$\begin{aligned} E[(\ln T_i)^2] &= (E[\ln T_i])^2 + Var[\ln T_i] = (\beta_0 + x_i^2)^2 + \sigma^2 Var[W_i] \\ &= (\beta_0 + x_i^2)^2 + \sigma^2. \end{aligned} \quad (6.9)$$

Here we use that $Var[X] = E[X^2] - (E[X])^2$ for a stochastic variable X . The expression for the Kullback–Leibler distance which needs to be minimized for $\hat{\boldsymbol{\theta}} = \{\hat{\beta}_0, \hat{\gamma}, \hat{\sigma}\}$ is with these results

$$\begin{aligned} KL(\hat{\boldsymbol{\theta}}) &= \sum_{i=1}^n \left[\ln \frac{\hat{\sigma}}{\sigma} + ((\beta_0 + x_i^2)^2 + \sigma^2) \cdot \left(\frac{1}{2\hat{\sigma}^2} - \frac{1}{2\sigma^2} \right) + (\beta_0 + x_i^2) \cdot \left(\frac{\beta_0 + x_i^2}{\sigma^2} - \frac{\hat{\beta}_0 + \hat{\gamma}x_i}{\hat{\sigma}^2} \right) \right. \\ &\quad \left. - \frac{(\beta_0 + x_i^2)^2}{2\sigma^2} + \frac{(\hat{\beta}_0 + \hat{\gamma}x_i)^2}{2\hat{\sigma}^2} \right]. \end{aligned} \quad (6.10)$$

Notice that the lifetimes, T_i 's, are eliminated from the function. As an example we now use a fixed design with only five covariate values $x_i \in \{-2, -1, 0, 1, 2\}$. To minimize

the function we take the partial derivative with respect to $\hat{\beta}_0$, $\hat{\gamma}$ and $\hat{\sigma}$. Setting the expressions equal to zero, and solving. Remember that $\beta_0 = 0$, $\sigma = 2$ and $n = 5$. We start with $\hat{\gamma}$.

$$\begin{aligned}\frac{\partial KL(\hat{\theta})}{\partial \hat{\gamma}} &= \sum_{i=1}^5 \left[-\frac{(\beta_0 + x_i^2)x_i}{\hat{\sigma}^2} + \frac{(\hat{\beta}_0 + \hat{\gamma}x_i)x_i}{\hat{\sigma}^2} \right] = 0 \\ &\Rightarrow \sum_{i=1}^5 [-x_i^3 + \hat{\gamma}x_i] = 0 \\ &\Rightarrow \hat{\gamma} = \frac{\sum_{i=1}^5 x_i^3}{\sum_{i=1}^5 x_i^2} = 0.\end{aligned}\tag{6.11}$$

The coefficient in front of the linear term in (6.5) is found to be zero. Then for β_0 .

$$\begin{aligned}\frac{\partial KL(\hat{\theta})}{\partial \hat{\beta}_0} &= \sum_{i=1}^5 \left[-\frac{\beta_0 + x_i^2}{\hat{\sigma}^2} + \frac{\hat{\beta}_0 + \hat{\gamma}x_i}{\hat{\sigma}^2} \right] = 0 \\ &\Rightarrow \sum_{i=1}^5 [-x_i^2 + \hat{\beta}_0] = 0 \\ &\Rightarrow \hat{\beta}_0 = \frac{1}{5} \sum_{i=1}^5 x_i^2 = 2.\end{aligned}\tag{6.12}$$

Finally, $\hat{\sigma}$.

$$\begin{aligned}\frac{\partial KL(\hat{\theta})}{\partial \hat{\sigma}} &= \sum_{i=1}^5 \left[\frac{1}{\hat{\sigma}} - \frac{(\beta_0 + x_i^2)^2 + \sigma^2}{\hat{\sigma}^3} + \frac{2(\beta_0 + x_i^2)(\hat{\beta}_0 + \hat{\gamma}x_i)}{\hat{\sigma}^3} - \frac{(\hat{\beta}_0 + \hat{\gamma}x_i)^2}{\hat{\sigma}^3} \right] = 0 \\ &\Rightarrow \sum_{i=1}^5 [\hat{\sigma}^2 - x_i^4 - 4 + 4x_i^2 - 4] = 0 \\ &\Rightarrow \hat{\sigma} = \sqrt{\frac{1}{5} \sum_{i=1}^5 [x_i^4 - 4x_i^2 + 8]} = \sqrt{34/5} \approx 2.60781.\end{aligned}\tag{6.13}$$

To check if these results are correct we simulate 1 million lifetimes from the model in (6.4). Here $\beta_0 = 0$, $\sigma = 2$ and W_i is standard normal distributed. For the x_i 's we use 200 000 observations with value -2, 200 000 with -1 and so on. The summary when fitting (6.5) gives us the estimated coefficients.

```
> summary(survLmod)
```

```
Call:
```

```
survreg(formula = Surv(fixedLData$T) ~ fixedLData$x, data = fixedLData,
        dist = "lognormal")
```

	Value	Std. Error	z	p
(Intercept)	2.00e+00	0.002608	7.68e+02	0.000
fixedLData\$x	1.44e-05	0.001844	7.81e-03	0.994
Log(scale)	9.58e-01	0.000707	1.36e+03	0.000

```
Scale= 2.61
```

```
Log Normal distribution
```

```
Loglik(model)= -4379657   Loglik(intercept only)= -4379657
```

```
Chisq= 0 on 1 degrees of freedom, p= 0.99
```

```
Number of Newton-Raphson Iterations: 1
```

```
n= 1000000
```

We see that all the coefficients from `survreg` match the analytically obtained ones. The coefficients minimizing the Kullback–Leibler distance are

$$\begin{aligned}
 \beta_0^* &= 2, \\
 \gamma_* &= 0, \\
 \sigma^* &= 2.61.
 \end{aligned}
 \tag{6.14}$$

6.2 Weibull

For Weibull distributed lifetimes. $g(\cdot)$ and $h(\cdot)$ are given, respectively, by

$$\begin{aligned} g(T_i) &= \frac{1}{\sigma} e^{-\frac{1}{\sigma}(\beta_0+x_i^2)} T_i^{1/\sigma-1} e^{-T_i^{1/\sigma}} e^{-(\beta_0+x_i^2)/\sigma}, \\ h(T_i, \hat{\theta}) &= \frac{1}{\hat{\sigma}} e^{-\frac{1}{\hat{\sigma}}(\hat{\beta}_0+\hat{\gamma}x_i)} T_i^{1/\hat{\sigma}-1} e^{-T_i^{1/\hat{\sigma}}} e^{-(\hat{\beta}_0+\hat{\gamma}x_i)/\hat{\sigma}}. \end{aligned} \quad (6.15)$$

Inserting this into (6.1) for a given i

$$\begin{aligned} KL(\hat{\theta}) &= \sum_{i=1}^n \left\{ E \left[-\ln \sigma - \frac{\beta_0 + x_i^2}{\sigma} + \frac{\ln T_i}{\sigma} - T_i^{1/\sigma} e^{-(\beta_0+x_i^2)/\sigma} \right. \right. \\ &\quad \left. \left. + \ln \hat{\sigma} + \frac{\hat{\beta}_0 + \hat{\gamma}x_i}{\hat{\sigma}} - \frac{\ln T_i}{\hat{\sigma}} + T_i^{1/\hat{\sigma}} e^{-(\hat{\beta}_0+\hat{\gamma}x_i)/\hat{\sigma}} \right] \right\}. \end{aligned} \quad (6.16)$$

Which means that we need to find $E[\ln T_i]$, $E[T_i^{1/\sigma}]$ and $E[T_i^{1/\hat{\sigma}}]$. Knowing that W_i is standard Gumbel distributed, $E[\ln T_i]$ is

$$E[\ln T_i] = E[\beta_0 + x_i^2 + \sigma W_i] = \beta_0 + x_i^2 + \sigma E[W_i] = \beta_0 + x_i^2 - \sigma \phi, \quad (6.17)$$

where ϕ is Euler's constant. For $E[T_i^{1/\sigma}]$ and $E[T_i^{1/\hat{\sigma}}]$ we use the standard approach for finding expected values. Starting with $E[T_i^{1/\sigma}]$.

$$\begin{aligned} E[T_i^{1/\sigma}] &= \int_0^\infty T_i^{1/\sigma} \frac{1}{\sigma} e^{-(\beta_0+x_i^2)/\sigma} T_i^{1/\sigma-1} e^{-T_i^{1/\sigma}} e^{-(\beta_0+x_i^2)/\sigma} dT_i \\ &= \frac{1}{\sigma} e^{-(\beta_0+x_i^2)/\sigma} \int_0^\infty T_i^{2/\sigma-1} e^{-T_i^{1/\sigma}} e^{-(\beta_0+x_i^2)/\sigma} dT_i \end{aligned} \quad (6.18)$$

Set $u = T_i^{1/\sigma} e^{-(\beta_0+x_i^2)/\sigma}$. Then $dT_i = e^{(\beta_0+x_i^2)/\sigma} \cdot \sigma \cdot T_i^{1-1/\sigma} du$. Inserting this in (6.18)

$$\begin{aligned} E[T_i^{1/\sigma}] &= \frac{1}{\sigma} e^{-(\beta_0+x_i^2)/\sigma} \int_0^\infty T_i^{2/\sigma-1} e^{-u} e^{(\beta_0+x_i^2)/\sigma} \cdot \sigma \cdot T_i^{1-1/\sigma} du \\ &= \int_0^\infty T_i^{1/\sigma} e^{-u} du = e^{(\beta_0+x_i^2)/\sigma} \int_0^\infty u e^{-u} du = e^{(\beta_0+x_i^2)/\sigma}. \end{aligned} \quad (6.19)$$

Here we have used that $T_i^{1/\sigma} = u \cdot e^{(\beta_0+x_i^2)/\sigma}$. Then we do much of the same for $E[T_i^{1/\hat{\sigma}}]$

$$\begin{aligned} E[T_i^{1/\hat{\sigma}}] &= \int_0^\infty T_i^{1/\hat{\sigma}} \frac{1}{\hat{\sigma}} e^{-(\beta_0+x_i^2)/\sigma} T_i^{1/\hat{\sigma}-1} e^{-T_i^{1/\hat{\sigma}}} e^{-(\beta_0+x_i^2)/\sigma} dT_i \\ &= \frac{1}{\hat{\sigma}} e^{-(\beta_0+x_i^2)/\sigma} \int_0^\infty T_i^{1/\hat{\sigma}+1/\hat{\sigma}-1} e^{-T_i^{1/\hat{\sigma}}} e^{-(\beta_0+x_i^2)/\sigma} dT_i \end{aligned} \quad (6.20)$$

Again, we use $u = T_i^{1/\sigma} e^{-(\beta_0+x_i^2)/\sigma}$ for substitution.

$$\begin{aligned} E[T_i^{1/\hat{\sigma}}] &= \frac{1}{\sigma} e^{-(\beta_0+x_i^2)/\sigma} \int_0^\infty T_i^{1/\sigma+1/\hat{\sigma}-1} e^{-u} e^{(\beta_0+x_i^2)/\sigma} \cdot \sigma \cdot T_i^{1-1/\sigma} du \\ &= \int_0^\infty T_i^{1/\hat{\sigma}} e^{-u} du = e^{(\beta_0+x_i^2)/\hat{\sigma}} \int_0^\infty u e^{-u} du = e^{(\beta_0+x_i^2)/\hat{\sigma}} \int_0^\infty u^{\sigma/\hat{\sigma}} e^{-u} du \quad (6.21) \\ &= e^{(\beta_0+x_i^2)/\hat{\sigma}} \cdot \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \end{aligned}$$

The last integral is recognized as the Gamma function. We also used that $T_i^{1/\hat{\sigma}} = \left(u \cdot e^{(\beta_0+x_i^2)/\sigma}\right)^{\sigma/\hat{\sigma}}$. The expression for the Kullback–Leibler distance which needs to be minimized for $\hat{\theta}$ is with these results

$$\begin{aligned} KL(\hat{\theta}) &= \sum_{i=1}^n \left[\ln \frac{\hat{\sigma}}{\sigma} + \frac{\hat{\beta}_0 + \hat{\gamma}x_i - \beta_0 - x_i^2 + \sigma\phi}{\hat{\sigma}} \right. \\ &\quad \left. - \phi - 1 + e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \cdot \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \right]. \quad (6.22) \end{aligned}$$

Again we use the fixed design with only five covariate values $x_i \in \{-2, -1, 0, 1, 2\}$. To minimize the function we take the partial derivative with respect to $\hat{\beta}_0$, $\hat{\gamma}$ and $\hat{\sigma}$. Setting the expressions equal to zero, and try to solve them. Remember that $\beta_0 = 0$, $\sigma = 2$ and $n = 5$.

$$\frac{\partial KL}{\partial \hat{\beta}_0} = \sum_{i=1}^n \left[\frac{1}{\hat{\sigma}} - \frac{1}{\hat{\sigma}} e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \cdot \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \right] = 0 \quad (6.23)$$

$$\frac{\partial KL}{\partial \hat{\gamma}} = \sum_{i=1}^n \left[\frac{x_i}{\hat{\sigma}} - \frac{x_i}{\hat{\sigma}} e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \cdot \Gamma\left[\frac{\sigma}{\hat{\sigma}} + 1\right] \right] = 0 \quad (6.24)$$

Finding the partial derivative for $\hat{\sigma}$ is not as straightforward as for the two other coefficients. To find the derivative of the gamma function we need to make use of the digamma function [15]. The digamma function is defined as the logarithmic derivative of the gamma function. For any x

$$\begin{aligned} \psi_0(x) &= \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}, \\ \Rightarrow \Gamma'(x) &= \psi_0(x)\Gamma(x). \end{aligned} \quad (6.25)$$

Then application of the chain rule gives the partial derivative for $\hat{\sigma}$.

$$\begin{aligned} \frac{\partial KL}{\partial \hat{\sigma}} = \sum_{i=1}^n \left[\frac{1}{\hat{\sigma}} - \frac{\hat{\beta}_0 - \beta_0 + \hat{\gamma}x_i - x_i^2 + \sigma\phi}{\hat{\sigma}^2} + e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \right. \\ \left. \cdot \left[-\frac{\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i}{\hat{\sigma}^2} \cdot \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] - \frac{\sigma \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \psi_0 \left[\frac{\sigma}{\hat{\sigma}} + 1 \right]}{\hat{\sigma}^2} \right] \right] = 0. \end{aligned} \quad (6.26)$$

These functions can not be solved analytically, so we try a numerical approach. We use the Newton-Rhapson method, found in any elementary calculus book, to find the values where a function $f(x)$ is zero, starting with an initial guess $x^{(0)}$

$$\begin{aligned} x^{(1)} &= x^{(0)} - \frac{f(x^{(0)})}{f'(x^{(0)})} \\ x^{(2)} &= x^{(1)} - \frac{f(x^{(1)})}{f'(x^{(1)})} \\ &\vdots \\ x^{(n)} &= x^{(n-1)} - \frac{f(x^{(n-1)})}{f'(x^{(n-1)})} \end{aligned} \quad (6.27)$$

According to (6.27) we need the second partial derivatives of $KL(\hat{\theta})$ for $\hat{\beta}_0$, $\hat{\gamma}$ and $\hat{\sigma}$.

$$\frac{\partial^2 KL}{\partial \hat{\beta}_0^2} = \sum_{i=1}^n \left[\frac{1}{\hat{\sigma}^2} e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \cdot \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \right] \quad (6.28)$$

$$\frac{\partial^2 KL}{\partial \hat{\gamma}^2} = \sum_{i=1}^n \left[\frac{x_i^2}{\hat{\sigma}^2} e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma}x_i)/\hat{\sigma}} \cdot \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \right] \quad (6.29)$$

The second partial derivative for $\hat{\sigma}$ utilizes the trigamma function [15]. The trigamma function is the second derivative of the logarithm of the gamma function. For any x

$$\begin{aligned} \psi_1(x) &= \frac{d^2}{dx^2} \ln \Gamma(x) = \frac{d}{dx} \psi_0(x), \\ \frac{d^2}{dx^2} \Gamma(x) &= \frac{d}{dx} (\psi_0(x)\Gamma(x)) = \frac{d}{dx} \psi_0(x)\Gamma(x) + \psi_0(x) \frac{d}{dx} \Gamma(x) \\ &= \psi_1(x)\Gamma(x) + \psi_0(x)\psi_0(x)\Gamma(x) = \Gamma(x) \left(\psi_0(x)^2 + \psi_1(x) \right). \end{aligned} \quad (6.30)$$

Applying this gives us the second partial derivative for $\hat{\sigma}$.

$$\begin{aligned}
\frac{\partial^2 KL}{\partial \hat{\sigma}^2} = \sum_{i=1}^n & \left[-\frac{1}{\hat{\sigma}^2} + 2 \frac{\hat{\beta}_0 + \hat{\gamma} x_i - \beta_0 - x_i^2 + \sigma \phi}{\hat{\sigma}^3} \right. \\
& + e^{(\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma} x_i)/\hat{\sigma}} \left[\left(\frac{\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma} x_i}{\hat{\sigma}^2} \right)^2 \cdot \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \right. \\
& + 2 \cdot \frac{\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma} x_i}{\hat{\sigma}^3} \cdot \Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \\
& + 4 \cdot \frac{\beta_0 - \hat{\beta}_0 + x_i^2 - \hat{\gamma} x_i}{\hat{\sigma}^2} \cdot \frac{\Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \psi_0 \left[\frac{\sigma}{\hat{\sigma}} + 1 \right]}{\hat{\sigma}^2} \\
& \left. \left. + \frac{4\Gamma \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \left(\psi_0 \left[\frac{\sigma}{\hat{\sigma}} + 1 \right]^2 + \hat{\sigma} \psi_0 \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] + \psi_1 \left[\frac{\sigma}{\hat{\sigma}} + 1 \right] \right)}{\hat{\sigma}^4} \right] \right]
\end{aligned} \tag{6.31}$$

The next step is to insert the expressions for the partial derivatives in (6.27) and continuously update to the newest estimated coefficient before the next calculation. It is easily explained by looking at the first iterations. We start with initial guesses for each parameter: $\hat{\beta}_0^{(0)}$, $\hat{\gamma}^{(0)}$ and $\hat{\sigma}^{(0)}$

First iteration:

$$\begin{aligned}
\hat{\beta}_0^{(1)} &= \hat{\beta}_0^{(0)} - \frac{\frac{\partial KL}{\partial \hat{\beta}_0}(\hat{\beta}_0^{(0)}, \hat{\gamma}^{(0)}, \hat{\sigma}^{(0)})}{\frac{\partial^2 KL}{\partial \hat{\beta}_0^2}(\hat{\beta}_0^{(0)}, \hat{\gamma}^{(0)}, \hat{\sigma}^{(0)})}, \\
\hat{\gamma}^{(1)} &= \hat{\gamma}^{(0)} - \frac{\frac{\partial KL}{\partial \hat{\gamma}}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(0)}, \hat{\sigma}^{(0)})}{\frac{\partial^2 KL}{\partial \hat{\gamma}^2}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(0)}, \hat{\sigma}^{(0)})}, \\
\hat{\sigma}^{(1)} &= \hat{\sigma}^{(0)} - \frac{\frac{\partial KL}{\partial \hat{\sigma}}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(0)})}{\frac{\partial^2 KL}{\partial \hat{\sigma}^2}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(0)})}.
\end{aligned}$$

Second iteration:

$$\begin{aligned}
\hat{\beta}_0^{(2)} &= \hat{\beta}_0^{(1)} - \frac{\frac{\partial KL}{\partial \hat{\beta}_0}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(1)})}{\frac{\partial^2 KL}{\partial \hat{\beta}_0^2}(\hat{\beta}_0^{(1)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(1)})}, \\
\hat{\gamma}^{(2)} &= \hat{\gamma}^{(1)} - \frac{\frac{\partial KL}{\partial \hat{\gamma}}(\hat{\beta}_0^{(2)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(1)})}{\frac{\partial^2 KL}{\partial \hat{\gamma}^2}(\hat{\beta}_0^{(2)}, \hat{\gamma}^{(1)}, \hat{\sigma}^{(1)})}, \\
\hat{\sigma}^{(2)} &= \hat{\sigma}^{(1)} - \frac{\frac{\partial KL}{\partial \hat{\sigma}}(\hat{\beta}_0^{(2)}, \hat{\gamma}^{(2)}, \hat{\sigma}^{(1)})}{\frac{\partial^2 KL}{\partial \hat{\sigma}^2}(\hat{\beta}_0^{(2)}, \hat{\gamma}^{(2)}, \hat{\sigma}^{(1)})}.
\end{aligned}$$

Continuing until it converges. This kind of numerical scheme is known as the Gauss-Seidel iterative method [16, p. 440]. This technique converges faster than the Jacobi iterative method, which does not update the coefficients immediately. It updates them after all three new estimates has been found. We set our initial values approximately to what `survreg` believes to be the correct coefficients. Again we use 1 million observations from (6.4). Here $\beta_0 = 0$, $\sigma = 2$ and W_i are from the standard Gumbel distribution of the smallest extreme. Also here we use 200 000 observations for each value of x_i . The summary when fitting (6.5) with `survreg` was

```
> summary(survWmod)
```

Call:

```
survreg(formula = Surv(fixedWData$T) ~ fixedWData$x, data = fixedWData,
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	2.30362	0.002816	818.042	0.000
fixedWData\$x	0.00123	0.001535	0.799	0.425
Log(scale)	0.97970	0.000755	1297.863	0.000

Scale= 2.66

Weibull distribution

Loglik(model)= -3375706 Loglik(intercept only)= -3375706

Chisq= 0.64 on 1 degrees of freedom, p= 0.42

Number of Newton-Raphson Iterations: 5

n= 1000000

We set the initial values to

$$\begin{aligned}\hat{\beta}_0^{(0)} &= 2.3, \\ \hat{\gamma}^{(0)} &= 0, \\ \hat{\sigma}^{(0)} &= 2.6.\end{aligned}\tag{6.32}$$

After 7 iterations we have convergence for the values

$$\begin{aligned}\hat{\beta}_0^{(7)} &= \beta_0^* = 2.298448, \\ \hat{\gamma}^{(7)} &= \gamma^* = -1.044555 \cdot 10^{-16}, \\ \hat{\sigma}^{(7)} &= \sigma^* = 2.66316.\end{aligned}\tag{6.33}$$

Which is approximately the same as `survreg`'s coefficients. These are the coefficients minimizing the Kullback–Leibler distance. The same coefficients are found using cruder initial values, but then we need more iterations.

7 Investigation of a real data set

In this section we investigate the functional form of a covariate in a real data set. The data set is censored, so we use the covariate order method. With a real data set, if there is a model to be found, we will not know the true functional form of the covariate. The data we look at are of low-cycle fatigue life of nickel-base superalloy found in [5]. We have 26 observations, of which four are censored. We have included the data set as Appendix C. Our goal is to find the relationship between the number of cycles to failure and pseudostress.

7.1 Distribution of the data

In order to use the covariate order method, we need to suggest the distribution of the data. In [5] they use a Weibull model with the natural logarithm of pseudostress as the covariate. To make our results comparable we use this as the covariate as well, denoting it x_i . We start with a scatter plot of x_i versus k-Cycles, denoted Y_i .

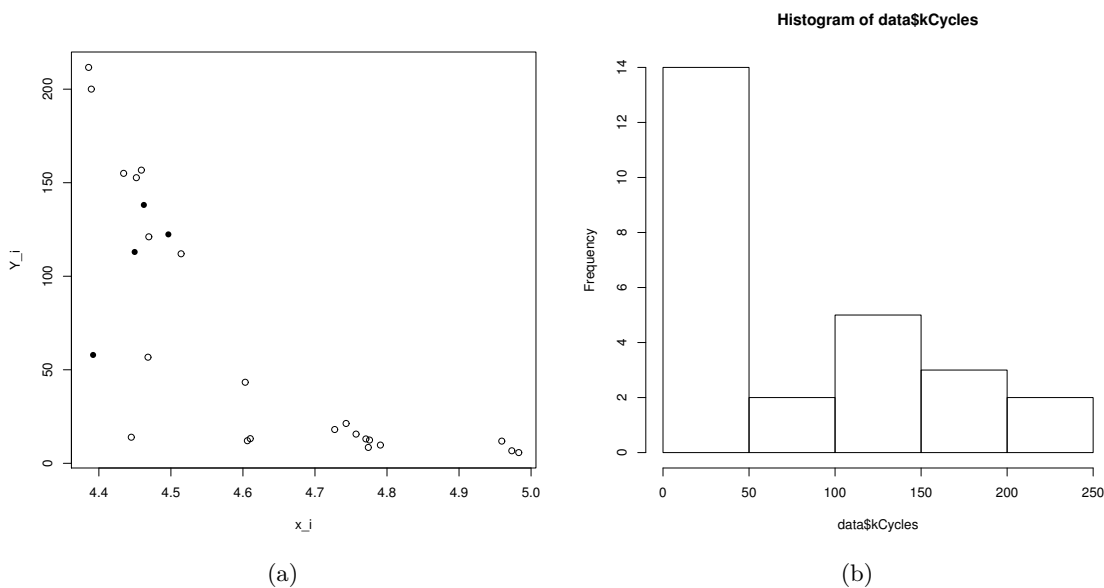


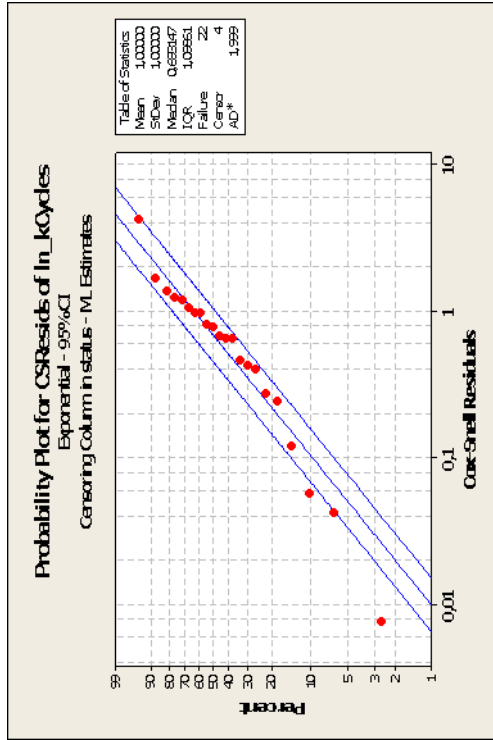
Figure 28: (a) Scatter plot of the covariate x_i , $\ln(\text{pseudostress})$, versus the response Y_i , k-Cycles. Circles represent failures, while dots are censored events. (b) Histogram of the lifetimes, k-Cycles, from the real data set.

In Figure 28(a) we see that the long lifetimes, i.e. large number of cycles, are found amongst the observations with low pseudostress. This is logical, since less stress means less strain on the items. However, there are some shorter lifetimes among the items with little pseudostress as well. This could be *infant mortality* failures, known from the well

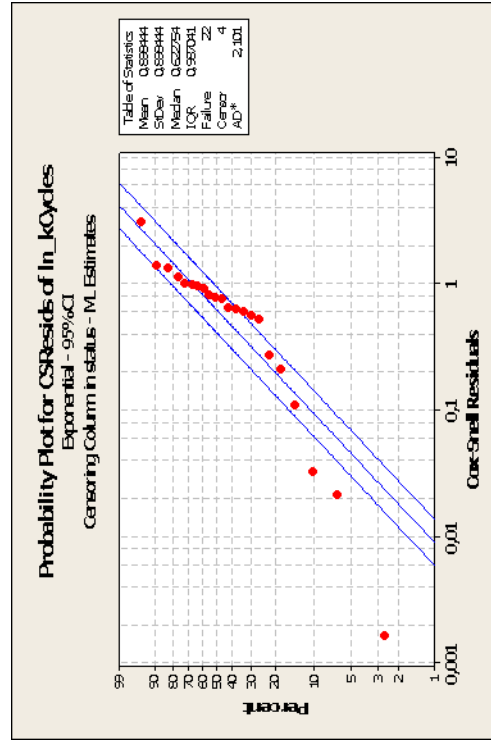
known *bathtub curve* in survival analysis. A plot of the histogram of the lifetimes can be seen in Figure 28(b). This histogram might suggest an exponential distribution, but we need more results to conclude. The histogram does not take any covariate dependency into consideration, so its relevance is questionable. Using MINITAB [17], a statistical software, we find probability plots for different distributions of $\ln Y_i$ as a linear function of the x_i 's:

$$\ln Y_i = \beta_0 + \beta_1 x_i + \sigma W_i. \quad (7.1)$$

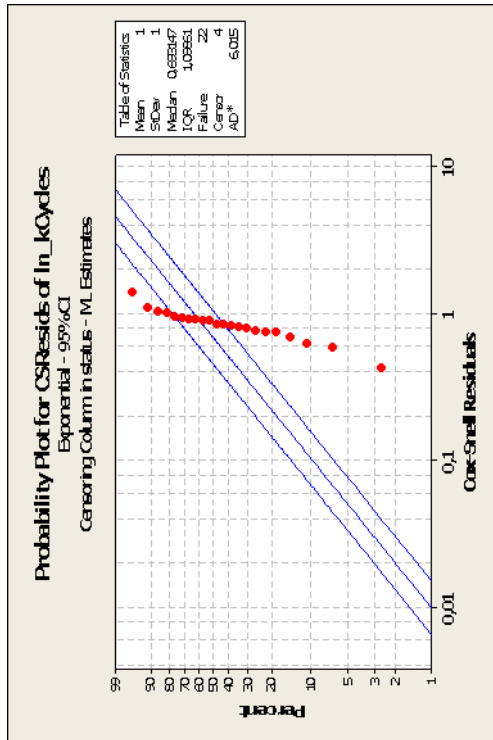
The probability plots can be seen in Figure 29. We see that the Weibull distribution is the one that fits the data best. This is also the distribution used in [5]. In the continuation of our investigation to find the functional form of the covariate, we assume $\ln Y_i$ to be Weibull distributed conditional on x_i .



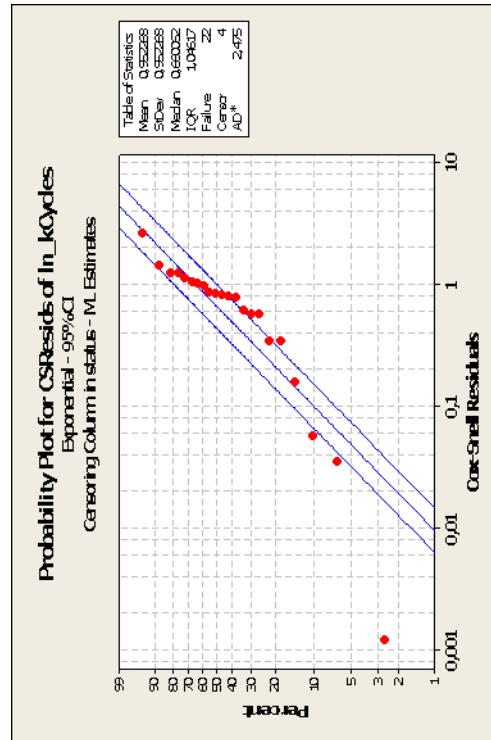
(a) Exponential.



(b) Weibull.



(c) Lognormal.



(d) Loglogistic.

Figure 29: Probability plots from the model in (7.1).

7.2 Covariate order method

Following the procedure used for with the simulated data sets with the covariate order method, we start with fitting the empty Weibull model with `survreg`.

$$\ln Y_i = \beta_0 + \sigma W_i. \quad (7.2)$$

From the regression summary in R the coefficients $\hat{\beta}_0 = \text{(Intercept)}$ and $\hat{\sigma} = \text{Scale}$ are found.

```
> summary(weibMod)
```

```
Call:
```

```
survreg(formula = Surv(Y, s) ~ 1, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	4.358	0.257	16.94	2.20e-64
Log(scale)	0.174	0.173	1.01	3.14e-01

```
Scale= 1.19
```

```
Weibull distribution
```

```
Loglik(model)= -118.4 Loglik(intercept only)= -118.4
```

```
Number of Newton-Raphson Iterations: 6
```

```
n= 26
```

With these coefficients the Cox–Snell residuals are calculated according to (4.3)

$$\hat{R}_i = e^{\frac{\ln Y_i - \hat{\beta}_0}{\hat{\sigma}}}. \quad (7.3)$$

Under the assumption that the empty model is correct, these should be approximately unit exponentially distributed. We use the Cox–Snell residuals as basis for the artificial point process in the covariate order method. The hazard rate is estimated for covariate x_i . If a trend is seen, the covariate is not modelled correctly. Before using the covariate order method, we must find the optimal smoothing parameter. Still using the likelihood cross-validation criterion from Section 4.2.2.3, the smoothing parameter, h is found to be

$$h = 0.2824328.$$

With this smoothing parameter the covariate order method estimates the hazard rate of the point process using the Cox–Snell residuals in (7.3). The hazard rate is plotted in Figure 30(a). Here a decreasing trend is clearly seen, so the functional form of the x_i 's needs more attention.

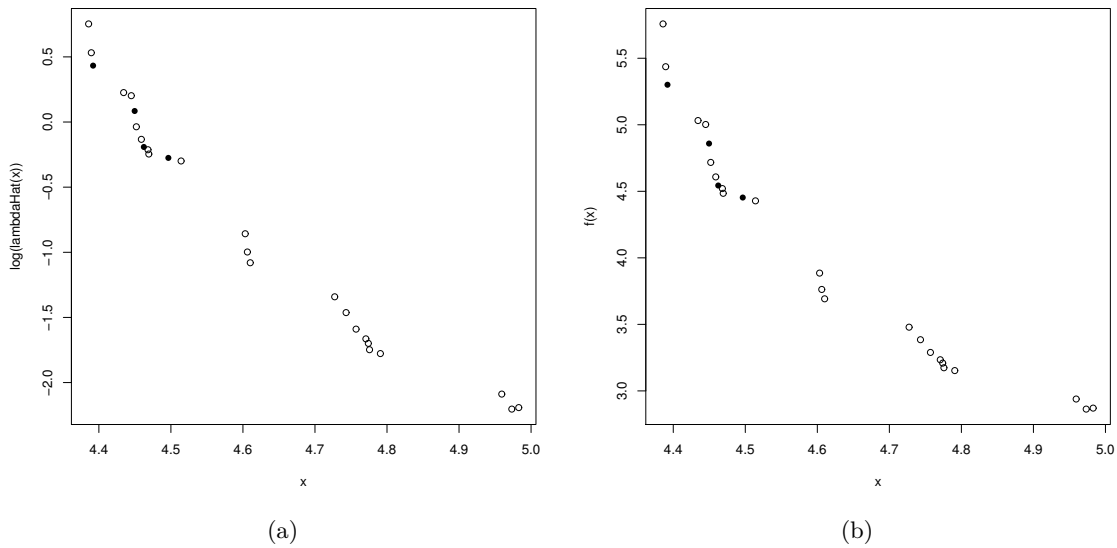


Figure 30: (a) Plot of the covariate x plotted against the estimated log-hazard rate $\ln(\hat{\lambda}(x))$. Circles represent failures, while dots are censored events. (b) Plot of x against $f(x)$ from (7.5). Circles represent failures, dots are censored events.

We assume that the covariate, x_i , should be included in the model.

$$\ln Y_i = \beta_0 + f(x_i) + \sigma W_i, \quad (7.4)$$

Then, according to (4.25) in Section 4.2.2.2, $f(x_i)$ can be seen by calculating

$$f(x_i) = \hat{\beta}_0 - \hat{\sigma} \left[\hat{\lambda}(x_i) \right], \quad (7.5)$$

and plotting it against the x_i 's. This is done in Figure 30(b). From the plot $f(x)$ is seen to decrease rapidly for small values of x , and more moderately for large values. Because of an apparent non-linear $\ln Y_i$, we try a quadratic polynomial, setting $f(x) = \beta_1 x + \beta_2 x^2$ in (7.4),

$$\ln Y_i = \beta_0 + \beta_1 x + \beta_2 x^2 + \sigma W_i. \quad (7.6)$$

One could as well suspect the model to be linear, but we choose to add the quadratic term since the summary from the fitted model will tell us if this term is significant or not. From the following summary, obtained when fitting the model in (7.6), we see that the quadratic term is significant. The values in the summary are the same as the ones found in [5, p. 440].

```
> summary(weibModNew)
```

```
Call:
```

```
survreg(formula = Surv(Y, s) ~ x + x2, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	217.611	62.132	3.50	4.61e-04
x	-85.522	26.546	-3.22	1.27e-03
x2	8.483	2.831	3.00	2.73e-03
Log(scale)	-0.982	0.179	-5.48	4.18e-08

```
Scale= 0.375
```

```
Weibull distribution
```

```
Loglik(model)= -93.4   Loglik(intercept only)= -118.4
```

```
Chisq= 50.01 on 2 degrees of freedom, p= 1.4e-11
```

```
Number of Newton-Raphson Iterations: 8
```

```
n= 26
```

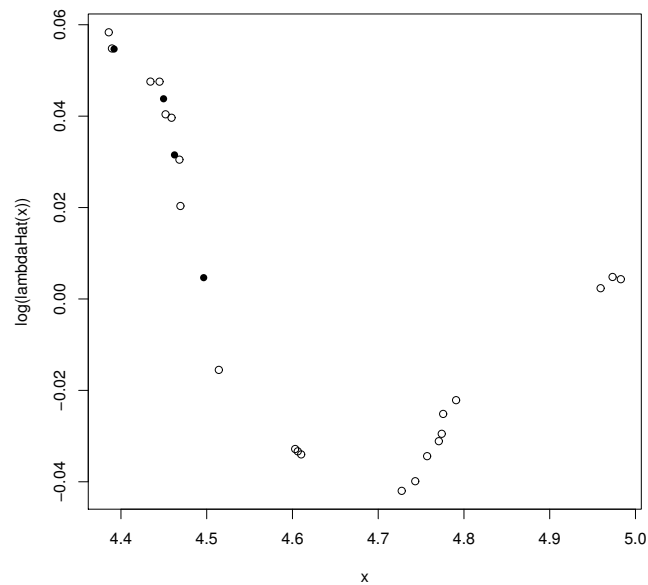


Figure 31: The covariate x plotted against the estimated log-hazard rate $\ln(\hat{\lambda}(1/x))$ from the model in (7.6)

New Cox–Snell residuals are computed

$$\hat{R}_i = e^{\frac{\ln T_i - \hat{\beta}_0 - \hat{\beta}_1 x - \hat{\beta}_2 x^2}{\hat{\sigma}}}. \quad (7.7)$$

If the new model in (7.6) is correct, there should be no trend in the hazard rate estimated with the covariate order method. We start by finding the new optimal smoothing parameter for the quadratic $f(x_i)$ using the likelihood cross-validation criterion. The new h is found to be

$$h = 3.999942.$$

The logarithm of the estimated hazard rate can be seen in Figure 31. A shape is seen, but when looking at the scale on the $\ln(\hat{\lambda}(x))$ -axis, we see that all the points are close to zero. This suggests that there is no trend.

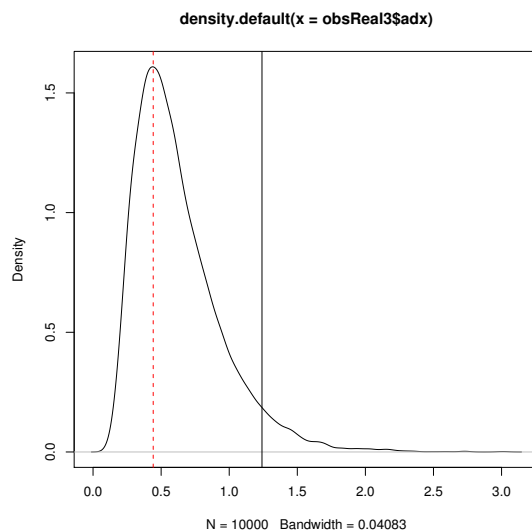


Figure 32: Density plots of the two bootstrapped distributions for the Anderson–Darling test statistic. The vertical, red, dashed line is the observed statistic, while the vertical, solid, black line is the 95%-quantile.

To formally check for trend, we make use of the Anderson–Darling statistic from (4.31). We bootstrap, non-parametrically, the data set to find our own distribution for the Anderson–Darling statistic. The observed statistic and 95%-quantile are listed in Table 7. We see that the observed statistic is smaller than the critical value. This is also illustrated in Figure 32. Hence there is no evidence of trend, and the model in (7.6) is acceptable.

95%-quantile	Observed statistic
1.240545	0.4418252

Table 7: Observed Anderson–Darling statistic and 95%-quantile from the estimated distribution for the superalloy data.

Another way to investigate if the model is correct, is to check if the Cox–Snell residuals can be assumed to be unit exponentially distributed, as we did for the transformation method in previous sections. A mean plot, density plot and probability plot are displayed in Figure 33. Here we have added 1 to the residuals from censored observations, as explained in Section 4.2.2.1. The plots are not perfect fits, but they indicate that the unit exponential might be the distribution of the residuals.

There might be other models which fit the data better. We tried fitting $f(x_i)$ with only a linear function as well. This model’s hazard rate was also found to have no significant trend. However, it had an Anderson–Darling statistic of approximately 0.67, which is higher than the statistic from the model with the quadratic polynomial. This is an indication that the quadratic model we have chosen is a better fit than the linear. The quadratic model was also the preferred model in [5]. Here they also tried a model where σ varies with x_i . This approach is described in Chapter 17.5.2, p. 439-441, in [5].

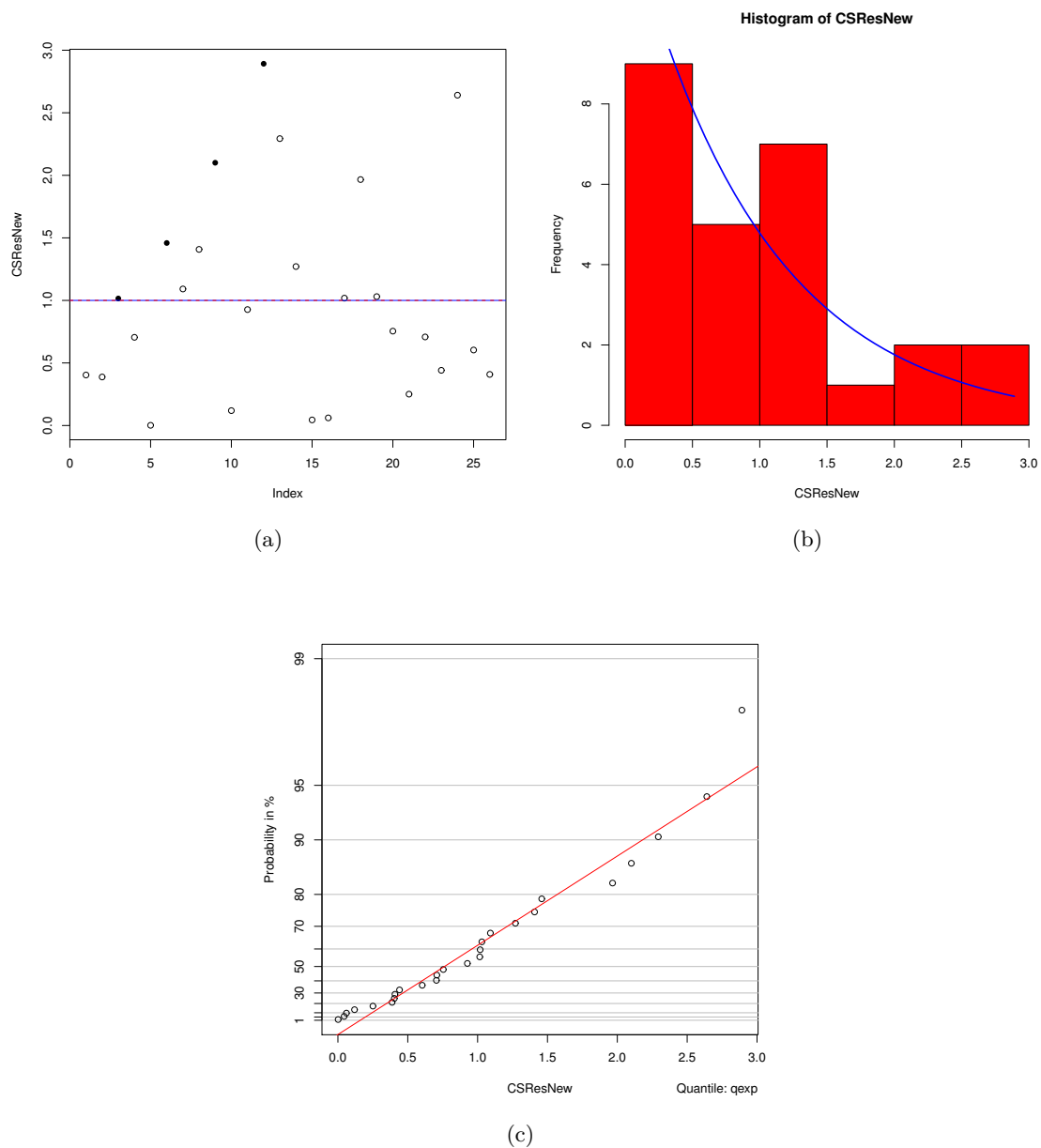


Figure 33: Plots verifying that the Cox-Snell residuals truly are unit exponentially distributed. (a) is a plot of the Cox-Snell residuals from the model (7.6) with theoretical mean (red dashed line) and estimated mean (blue solid line). Circles represent failures, solid squares are censored events. (b) is a histogram of the Cox-Snell residuals with theoretical density function. (c) is the exponential probability plot. 1 is added to the censored residuals.

8 Discussion and conclusions

In this thesis we have shown that the functional form of covariates in misspecified survival regression models can be retrieved with a transformation of the model's Cox–Snell residuals, as well as estimating and investigating the hazard rate functions of the Cox–Snell residuals using the covariate order method.

In Section 4 and 5 we see that when using the transformation method, the functional form is clearly seen for complete data sets. However, when censoring is introduced the functional form is more blurred. The covariate order method retrieves the functional form better than the transformation method for censored data sets, and especially for high levels of censoring (50% and 80%). This is because of the transformation method's naïve approach of adding 1 to the Cox–Snell residuals of censored observations justified by the memoryless property of the exponential distribution. This leads to many Cox–Snell residuals approximating to 1, thus causing the method to fail. In the project [1] we could not conclude which of the methods that was preferred because of an insufficient basis of results. Now however, we can say that the covariate order method clearly is better for high levels of censoring. This is seen for two different probability distributions, the Weibull and lognormal distribution. In our search for $f(x)$, in (3.1), we start by assuming it is linear. Although we know that it is a quadratic function. This is done throughout the simulation parts of the thesis. It would have been interesting to test the methods on other polynomials, as well as logarithmic and exponential functions. For the real data set we did not know the original functional form, but it turned out to be quadratic as well. For the simulated data sets, if we know the true functional form, we might subconsciously recognize its form easier. For real data sets we will not know if the chosen model is correct. There are benefits and liabilities for both real data and simulated data, like a double-edged sword. Another moment is that we have used standard normally distributed covariates, which are for the most part clustered around zero. With uniformly distributed variables it was seen in the project that the functional form was retrieved clearer for complete data sets. However, in the project, uniform covariates from -5 to 5 led to unrealistically large lifetimes.

In Section 6, it has been shown that the maximum likelihood approach to finding the coefficients in a misspecified model, which minimizes the Kullback–Leibler distance, is consistent with the coefficients obtained analytically. It is reassuring to know that the methods implemented in the software used give the same results as a theoretical approach does.

After completing the project it was believed that recovering the functional form in other lifetime distributions than Weibull would lead to a blurring effect. The blurring would be caused by approximations of the cumulative distribution function of the error terms, W , in (3.1). While working with this thesis, it was seen that what was thought to be unique for the elegant Cox–Snell residuals of Weibull distributed lifetimes, seen in (4.3),

in fact could be generalized to any other lifetime distribution as well.

References

- [1] Stein Aaserud. *Inferring the functional form of covariates in misspecified survival regression models*. TMA4500 Industrial Mathematics, Specialization Project, December 2010.
- [2] Jan Terje Kvaløy. Covariate order tests for covariate effect. *Lifetime Data Analysis*, 8:35–51, 2002.
- [3] Halbert White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):pp. 1–25, 1982.
- [4] Robert Gentleman and Ross Ihaka et al. The R Project for Statistical Computing. <http://www.r-project.org/>. 22.05.2011.
- [5] William Q. Meeker and Luis A. Escobar. *Statistical Methods for Reliability Data*. John Wiley & Sons, 1998.
- [6] Marvin Rausand and Arnljot Høyland. *System Reliability Theory - Models, Statistical Methods and Applications*. John Wiley & Sons, second edition, 2004.
- [7] Oxford Dictionary of Statistics: Cox–Snell residuals on Answers.com, *A Dictionary of Statistics*, 2008.
- [8] B. W. Silverman. Density estimation for statistics and data analysis. *Monographs on Statistics and Applied Probability*, 1986.
- [9] Jan Terje Kvaløy and Bo Henry Lindqvist. The covariate order method for censored exponential regression. Preprint S10-1998, Department of Mathematical Sciences Norwegian University of Science and Technology, October 1998.
- [10] NIST/SEMATECH. *e-Handbook of Statistical Methods*, chapter 4.1.4.4. LOESS (aka LOWESS). NIST, 2003. <http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm>, 22.05.2011.
- [11] NIST/SEMATECH. *e-Handbook of Statistical Methods*, chapter 1.3.3.22. Probability Plot. NIST, 2003. <http://www.itl.nist.gov/div898/handbook/eda/section3/probplot.htm>, 22.05.2011.
- [12] Alvin C. Rencher and G. Bruce Schaalje. *Linear Models in Statistics*, chapter 9.3. John Wiley & Sons, second edition, 2008.
- [13] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1994.
- [14] NIST/SEMATECH. *e-Handbook of Statistical Methods*, chapter 1.3.5.14. Anderson–Darling Test. NIST, 2003. <http://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm>, 22.05.2011.

- [15] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, 1964.
- [16] Richard L. Burden and J. Douglas Faires. *Numerical Analysis*. Thomson Brooks/Cole, 8th edition, 2005.
- [17] Brian L. Joiner, Barbara F. Ryan, Thomas A. Ryan et al. Minitab Software for Quality Improvement. <http://www.minitab.com>. 22.05.2011.

A Simulated data set of Weibull distributed lifetimes with censoring

Table of simulated Weibull distributed lifetimes. Y_{20_i} is the observation time for observation i with 20% censoring. $Y_{20_i} = \min(T_i, C_{20_i})$, where T_{20_i} is the lifetime and C_{20_i} is the censoring time. s_{20_i} is the censoring status, 1 if the observation is a failure and 0 if it is censored. Analogous for 50% and 80% censoring.

	Y20_i	Y50_i	Y80_i	T_i	C20_i	C50_i	C80_i	s20_i	s50_i	s80_i	z_i1	z_i2	x_i	W_i
1	0.96	0.45	0.00	0.96	57.54	0.45	0.00	1	0	0	-0.43	-0.71	-0.63	0.93
2	0.00	0.00	0.00	0.00	561.28	2.61	0.02	1	1	1	-2.86	1.91	-0.06	-2.32
3	0.08	0.08	0.00	0.08	38.44	0.91	0.00	1	1	0	-0.40	-1.97	-0.40	-0.17
4	138.20	1.89	0.04	138.20	258.77	1.89	0.04	1	0	0	0.15	-0.60	0.78	1.85
5	0.00	0.00	0.00	0.00	15.47	2.10	0.00	1	1	1	0.10	1.51	0.36	-4.24
6	0.01	0.01	0.01	0.01	394.60	0.38	0.02	1	1	1	-2.13	-0.13	2.45	0.09
7	23.80	0.14	0.00	23.80	90.06	0.14	0.00	1	0	0	0.48	1.50	0.13	0.22
8	0.01	0.01	0.00	0.01	119.62	0.88	0.00	1	1	0	-0.39	0.07	-0.16	-1.41
9	410.28	1.63	0.01	23074.89	410.28	1.63	0.01	0	0	0	1.04	1.90	-2.38	-0.62
10	462.51	4.01	0.04	11003.12	462.51	4.01	0.04	0	0	0	2.12	-0.15	0.48	-0.74
11	22.56	0.20	0.00	22.56	217.84	0.20	0.00	1	0	0	0.12	-0.04	-1.96	-0.66
12	1.41	1.41	0.05	1.41	334.49	1.65	0.05	1	1	0	0.05	-1.50	0.33	0.15
13	0.00	0.00	0.00	0.00	942.97	0.70	0.01	1	1	1	-0.43	-1.31	1.45	-4.57
14	2.13	0.23	0.02	2.13	60.85	0.23	0.02	1	0	0	-0.17	-0.16	-1.08	0.22
15	15.43	2.42	0.01	15.43	220.99	2.42	0.01	1	0	0	0.25	-1.40	-1.07	0.33
16	297.05	3.08	0.00	6679.99	297.05	3.08	0.00	0	0	0	0.63	-1.75	2.93	-1.29
17	0.01	0.01	0.01	0.01	319.53	0.23	0.01	1	1	1	-0.65	-0.03	0.14	-0.79

18	0.01	0.01	0.00	0.01	0.00	0.01	800.10	0.59	0.00	1	1	0	-0.79	0.44	-0.13	-0.67
19	5.47	3.02	0.01	5.47	3.02	0.01	11.50	3.02	0.01	1	0	0	0.10	-1.08	-1.91	-1.11
20	614.77	1.27	0.00	916.12	614.77	1.27	614.77	1.27	0.00	0	0	0	0.95	-0.64	-0.97	0.63
21	0.00	0.00	0.00	0.00	222.94	1.95	222.94	1.95	0.00	1	1	1	-1.85	1.42	0.12	-0.66
22	0.00	0.00	0.00	0.00	439.09	1.52	439.09	1.52	0.00	1	1	0	-0.78	0.95	0.35	-1.00
23	0.00	0.00	0.00	0.00	306.69	0.94	306.69	0.94	0.00	1	1	1	-1.91	-0.27	-0.57	-0.82
24	0.00	0.00	0.00	0.00	130.81	0.89	130.81	0.89	0.04	1	1	1	-1.06	-0.78	-0.47	-0.47
25	0.29	0.29	0.02	0.29	471.50	1.12	471.50	1.12	0.02	1	1	0	-0.52	2.23	-1.11	-0.15
26	10.69	2.75	0.01	10.69	83.91	2.75	83.91	2.75	0.01	1	0	0	0.74	-0.76	0.30	-0.63
27	0.00	0.00	0.00	0.00	66.27	0.69	66.27	0.69	0.01	1	1	1	-1.23	-0.14	0.51	-0.87
28	8.21	0.74	0.00	8.21	368.49	0.74	368.49	0.74	0.00	1	0	0	0.04	1.38	-1.35	-0.11
29	32.24	0.03	0.01	32.24	195.39	0.03	195.39	0.03	0.01	1	0	0	-0.22	1.90	-1.22	1.35
30	98.00	0.72	0.00	359.94	98.00	0.72	98.00	0.72	0.00	0	0	0	0.46	1.96	-1.14	0.94
31	0.00	0.00	0.00	0.00	340.95	3.43	340.95	3.43	0.01	1	1	1	-1.48	0.12	0.56	0.30
32	0.58	0.58	0.00	0.58	61.20	2.11	61.20	2.11	0.00	1	1	0	0.30	-0.52	0.37	-1.03
33	287.40	2.58	0.01	287.40	536.65	2.58	536.65	2.58	0.01	1	0	0	1.66	0.27	-0.32	-1.41
34	107.85	2.08	0.00	107.85	345.34	2.08	345.34	2.08	0.00	1	0	0	0.41	-0.54	-1.39	0.41
35	0.77	0.77	0.02	0.77	14.29	1.42	14.29	1.42	0.02	1	1	0	0.01	-1.11	2.01	-2.08
36	0.20	0.20	0.01	0.20	112.05	0.99	112.05	0.99	0.01	1	1	0	-0.15	-0.05	-0.59	-0.61
37	169.76	1.09	0.01	169.76	540.14	1.09	540.14	1.09	0.01	1	0	0	0.02	-0.58	2.08	0.41
38	52.27	0.80	0.01	1823419297.39	52.27	0.80	52.27	0.80	0.01	0	0	0	3.12	-0.35	-2.07	0.75
39	138.31	8.58	0.01	5606.40	138.31	8.58	138.31	8.58	0.01	0	0	0	0.79	-0.62	2.31	-0.27
40	1.33	0.44	0.01	1.33	17.25	0.44	17.25	0.44	0.01	1	0	0	-0.62	0.73	0.89	1.22
41	150.73	1.85	0.00	150.73	759.12	1.85	759.12	1.85	0.00	1	0	0	1.42	-1.00	-0.08	-0.95
42	13.64	3.94	0.01	13.64	240.93	3.94	240.93	3.94	0.01	1	0	0	0.95	-0.62	0.31	-1.05
43	1.48	0.21	0.01	1.48	30.34	0.21	30.34	0.21	0.01	1	0	0	0.25	0.06	0.38	-0.51
44	0.00	0.00	0.00	0.00	284.85	1.11	284.85	1.11	0.02	1	1	1	-2.43	-1.79	-0.48	-3.68

45	7.15	1.62	0.02	9.15	7.15	1.62	0.02	0	0	0	0.23	-1.52	-0.68	0.45
46	0.14	0.14	0.01	0.14	41.91	1.14	0.01	1	1	0	0.09	1.12	-0.26	-1.36
47	317.65	6.68	0.01	894.32	317.65	6.68	0.01	0	0	0	0.41	-0.06	-2.10	0.19
48	0.00	0.00	0.00	0.00	126.45	5.74	0.00	1	1	1	0.06	-0.45	0.15	-3.38
49	2.68	2.68	0.01	2.68	92.56	4.19	0.01	1	1	0	0.04	-0.37	-0.60	0.26
50	0.03	0.03	0.02	0.03	165.91	2.02	0.02	1	1	0	-0.60	-0.69	-0.48	-0.25
51	0.29	0.19	0.00	0.29	276.75	0.19	0.00	1	0	0	-1.63	1.66	-1.91	1.47
52	0.15	0.15	0.00	0.15	82.17	3.51	0.00	1	1	0	0.23	-1.11	-0.64	-1.60
53	0.52	0.52	0.00	0.52	63.37	0.55	0.00	1	1	0	1.17	0.35	0.44	-3.39
54	4.77	0.17	0.01	6160.94	4.77	0.17	0.01	0	0	0	2.09	-2.58	-0.48	-0.71
55	250.92	1.64	0.01	463.59	250.92	1.64	0.01	0	0	0	1.18	-1.07	0.61	0.03
56	0.02	0.02	0.00	0.02	63.96	1.50	0.00	1	1	0	-0.70	-0.39	-0.30	-0.24
57	0.02	0.02	0.01	0.02	41.85	0.28	0.01	1	1	0	-0.08	-0.47	0.47	-1.84
58	0.53	0.27	0.00	0.53	23.58	0.27	0.00	1	0	0	0.72	-1.34	0.47	-2.10
59	2.69	0.17	0.00	2.69	283.41	0.17	0.00	1	0	0	1.05	-0.29	0.25	-2.13
60	65.99	0.58	0.02	65.99	133.98	0.58	0.02	1	0	0	0.64	0.86	-0.34	0.36
61	0.05	0.05	0.00	0.05	114.33	0.99	0.00	1	1	0	0.47	-0.21	0.18	-2.64
62	0.00	0.00	0.00	0.00	87.93	1.41	0.01	1	1	1	-2.10	0.59	-0.19	1.01
63	0.00	0.00	0.00	0.00	223.99	0.92	0.01	1	1	1	-0.82	0.32	-0.13	-0.97
64	424.29	3.25	0.03	22909.50	424.29	3.25	0.03	0	0	0	1.67	-0.59	1.42	-0.11
65	15.46	4.64	0.01	15.46	77.78	4.64	0.01	1	0	0	-0.25	-0.48	1.83	0.39
66	4.09	2.60	0.00	10.84	4.09	2.60	0.00	0	0	0	0.84	-0.10	-0.16	-0.92
67	0.07	0.07	0.01	0.07	102.19	3.84	0.01	1	1	0	-1.49	0.65	-1.69	0.88
68	0.00	0.00	0.00	0.00	154.30	0.31	0.01	1	1	1	-2.13	-0.92	0.35	0.47
69	0.90	0.90	0.01	0.90	17.21	1.24	0.01	1	1	0	-0.33	1.09	1.26	-0.14
70	428.06	2.01	0.01	492.12	428.06	2.01	0.01	0	0	0	0.96	0.90	-1.26	-0.18
71	0.20	0.20	0.02	0.20	290.27	0.47	0.02	1	1	0	0.32	0.53	-0.28	-1.68
72	0.00	0.00	0.00	0.00	177.88	3.42	0.01	1	1	1	-0.76	-0.78	-0.28	-4.42

73	0.89	0.89	0.02	0.89	0.02	0.89	0.02	165.04	0.93	0.02	1	1	0	0.12	0.58	-0.05	-0.43
74	0.01	0.01	0.01	0.01	0.01	0.01	0.01	31.56	0.31	0.01	1	1	0	-0.19	0.02	-0.17	-1.94
75	0.74	3.01	0.01	3.01	0.01	47.33	0.74	0.74	3.01	0.01	0	0	0	1.47	-0.76	-1.22	-2.42
76	140.02	1.43	0.01	140.02	0.01	140.02	853.50	1.43	1.43	0.01	1	0	0	0.94	-1.86	-0.03	0.30
77	238.37	3.91	0.01	238.37	0.01	2118.34	238.37	3.91	3.91	0.01	0	0	0	1.01	0.21	1.53	0.11
78	0.15	0.15	0.02	0.15	0.02	0.15	366.22	0.94	0.94	0.02	1	1	0	0.39	0.52	0.77	-2.29
79	4.69	1.02	0.00	4.69	0.00	4.69	230.86	1.02	1.02	0.00	1	0	0	0.06	0.40	0.99	0.10
80	13.18	0.34	0.01	13.18	0.01	13.18	293.77	0.34	0.34	0.01	1	0	0	-0.13	-2.52	-0.80	1.55
81	0.00	0.00	0.00	0.00	0.00	0.00	34.93	11.94	11.94	0.01	1	1	1	-1.36	1.30	-0.46	0.46
82	201.04	0.48	0.02	1281.08	0.48	1281.08	201.04	0.48	0.48	0.02	0	0	0	0.85	0.01	-1.47	0.37
83	0.00	0.00	0.00	0.00	0.00	0.00	164.44	0.23	0.23	0.05	1	1	1	-0.22	-0.09	-0.08	-2.73
84	22.60	1.63	0.00	22.60	0.00	22.60	293.06	1.63	1.63	0.00	1	0	0	0.53	-0.85	0.85	-0.04
85	64.53	0.05	0.01	69.27	0.05	69.27	64.53	0.05	0.05	0.01	0	0	0	0.77	-0.11	0.33	0.16
86	0.01	0.01	0.01	0.01	0.01	0.01	24.96	1.73	1.73	0.01	1	1	0	-0.33	-1.59	1.40	-2.43
87	0.12	0.12	0.00	0.12	0.00	0.12	41.98	2.22	2.22	0.00	1	1	0	-0.60	1.28	-1.66	-1.07
88	0.00	0.00	0.00	0.00	0.00	0.00	3.15	3.15	0.25	0.00	1	1	1	-1.76	-2.08	-0.49	0.42
89	237.10	0.72	0.00	105499.72	0.72	105499.72	237.10	0.72	0.72	0.00	0	0	0	1.77	0.92	-0.91	0.85
90	101.65	1.43	0.02	101.65	0.02	101.65	555.46	1.43	1.43	0.02	1	0	0	0.91	0.08	0.91	-0.38
91	495.40	1.67	0.01	495.40	0.01	495.40	513.05	1.67	1.67	0.01	1	0	0	1.35	0.12	-0.18	-0.30
92	0.00	0.00	0.00	0.00	0.00	0.00	77.24	4.26	4.26	0.01	1	1	1	-0.59	-1.26	0.42	-1.34
93	221.50	1.58	0.02	1042.78	1.58	1042.78	221.50	1.58	1.58	0.02	0	0	0	0.96	0.38	1.33	0.15
94	0.02	0.02	0.01	0.02	0.01	0.02	401.52	0.68	0.68	0.01	1	1	0	-0.09	-0.08	-0.74	-2.02
95	72.70	4.70	0.02	4868.43	4.70	4868.43	72.70	4.70	4.70	0.02	0	0	0	1.05	0.08	-0.37	1.54
96	10.43	4.86	0.01	10.43	0.01	10.43	47.91	4.86	4.86	0.01	1	0	0	0.33	0.57	0.73	0.03
97	0.00	0.00	0.00	0.00	0.00	0.00	62.72	1.08	1.08	0.01	1	1	1	-0.77	0.70	-0.14	-1.15
98	0.00	0.00	0.00	0.00	0.00	0.00	147.98	0.59	0.59	0.00	1	1	0	-1.61	1.12	-0.64	0.22
99	2.44	2.44	0.02	2.44	0.02	2.44	143.69	4.08	4.08	0.02	1	1	0	-0.40	1.68	-0.71	1.03
100	0.04	0.04	0.03	0.04	0.03	0.04	170.58	0.14	0.14	0.03	1	1	0	0.72	-0.88	0.13	-3.39

B Simulated data set of lognormally distributed lifetimes with censoring

Table of simulated lognormally distributed lifetimes. $Y20_i$ is the observation time for observation i with 20% censoring. $Y20_i = \min(T_i, C20_i)$, where $T20_i$ is the lifetime and $C20_i$ is the censoring time. $s20_i$ is the censoring status, 1 if the observation is a failure and 0 if it is censored. Analogous for 50% and 80% censoring.

	T.i	Y20.i	Y50.i	Y80.i	x.i	W.i	C20.i	C50.i	C80.i	s20.i	s50.i	s80.i
1	22.35	22.35	0.99	0.15	0.32	1.50	81.45	0.99	0.15	1	0	0
2	1.37	1.37	1.37	0.41	-0.27	0.12	23.54	11.46	0.41	1	1	0
3	0.13	0.13	0.13	0.13	-0.09	-1.01	20.78	2.42	2.15	1	1	1
4	2.24	2.24	2.24	0.20	0.06	0.40	69.12	2.75	0.20	1	1	0
5	0.20	0.20	0.20	0.03	-0.38	-0.87	90.11	4.09	0.03	1	1	0
6	12.83	10.11	6.87	0.02	0.64	1.07	10.11	6.87	0.02	0	0	0
7	2.41	2.41	1.62	0.11	-0.42	0.35	5.24	1.62	0.11	1	0	0
8	0.06	0.06	0.06	0.06	-0.84	-1.78	8.11	7.82	0.45	1	1	1
9	6.98	5.16	6.98	0.12	0.20	0.95	5.16	7.85	0.12	0	1	0
10	14.30	14.30	3.31	0.40	-0.84	0.98	36.84	3.31	0.40	1	0	0
11	8.79	8.79	5.05	0.13	-1.45	0.03	70.02	5.05	0.13	1	0	0
12	0.11	0.11	0.11	0.11	0.06	-1.09	121.88	2.69	0.37	1	1	1
13	0.79	0.79	0.79	0.70	0.74	-0.39	1.47	3.85	0.70	1	1	0
14	4.43	4.43	2.59	0.63	-0.15	0.73	29.37	2.59	0.63	1	0	0
15	43.28	43.28	3.13	0.03	-1.57	0.66	56.72	3.13	0.03	1	0	0
16	6.53	6.53	6.29	0.10	1.27	0.13	80.69	6.29	0.10	1	0	0
17	0.64	0.64	0.64	0.42	-0.69	-0.46	63.69	11.40	0.42	1	1	0

18	7.46	7.46	3.35	1.03	0.05	1.00	8.82	3.35	1.03	1	0	0
19	27.94	21.62	0.43	0.33	1.17	0.98	21.62	0.43	0.33	0	0	0
20	3.99	3.99	3.99	0.74	-0.03	0.69	92.84	15.23	0.74	1	1	0
21	12.86	12.86	3.34	0.10	1.34	0.39	20.92	3.34	0.10	1	0	0
22	28.18	28.18	1.01	0.10	-0.89	1.27	56.76	1.01	0.10	1	0	0
23	0.88	0.88	0.88	0.64	-0.11	-0.07	67.97	8.59	0.64	1	1	0
24	0.28	0.28	0.28	0.15	-0.15	-0.65	16.58	4.14	0.15	1	1	0
25	32.68	3.90	7.38	0.00	-0.94	1.30	3.90	7.38	0.00	0	0	0
26	2.03	2.03	2.03	0.03	-0.48	0.24	139.17	7.40	0.03	1	1	0
27	4.14	4.14	2.83	0.03	1.16	0.03	38.20	2.83	0.03	1	0	0
28	13.47	9.08	7.95	0.09	-1.26	0.51	9.08	7.95	0.09	0	0	0
29	11.50	4.19	5.87	0.33	-0.64	1.02	4.19	5.87	0.33	0	0	0
30	5.09	5.09	4.78	0.10	-1.25	0.03	13.80	4.78	0.10	1	0	0
31	0.10	0.10	0.10	0.05	0.27	-1.19	0.65	3.23	0.05	1	1	0
32	1.46	0.18	1.46	0.10	0.49	0.07	0.18	3.79	0.10	0	1	0
33	0.96	0.96	0.96	0.39	-0.87	-0.40	25.67	3.98	0.39	1	1	0
34	6.76	6.76	6.76	0.54	-0.83	0.61	153.69	8.12	0.54	1	1	0
35	5.25	5.25	5.25	0.38	1.59	-0.44	132.93	8.44	0.38	1	1	0
36	0.40	0.40	0.40	0.08	1.34	-1.35	75.85	6.93	0.08	1	1	0
37	14.41	6.52	0.36	0.13	-1.30	0.49	6.52	0.36	0.13	0	0	0
38	379.98	20.86	4.33	1.50	2.18	0.59	20.86	4.33	1.50	0	0	0
39	69.35	69.35	12.01	0.74	-1.45	1.07	183.07	12.01	0.74	1	0	0
40	0.81	0.81	0.81	0.13	1.59	-1.37	132.11	6.42	0.13	1	1	0
41	0.19	0.19	0.19	0.19	0.25	-0.85	12.79	4.36	0.19	1	1	1
42	3.04	3.04	0.07	0.36	-0.33	0.50	26.81	0.07	0.36	1	0	0
43	2.33	2.33	0.84	0.01	0.31	0.37	19.15	0.84	0.01	1	0	0
44	424.51	70.92	3.37	0.13	-1.63	1.69	70.92	3.37	0.13	0	0	0

73	0.79	0.79	0.79	0.79	0.09	0.38	-0.19	3.74	5.12	0.09	1	1	0
74	200.24	23.70	1.24	0.14	1.60	1.37	23.70	23.70	1.24	0.14	0	0	0
75	3.46	3.46	0.17	0.85	0.37	0.55	4.24	4.24	0.17	0.85	1	0	0
76	5.62	5.62	3.67	0.39	0.67	0.64	91.55	91.55	3.67	0.39	1	0	0
77	0.34	0.34	0.34	0.34	0.66	-0.76	23.78	23.78	1.69	1.03	1	1	1
78	1.76	1.76	1.76	0.63	1.24	-0.48	35.42	35.42	2.17	0.63	1	1	0
79	2.01	2.01	2.01	0.72	0.38	0.28	23.54	23.54	7.62	0.72	1	1	0
80	80.64	54.61	7.37	0.08	-0.93	1.76	54.61	54.61	7.37	0.08	0	0	0
81	1.51	1.51	1.51	0.28	-1.30	-0.64	78.83	78.83	2.50	0.28	1	1	0
82	0.13	0.13	0.13	0.00	0.75	-1.30	49.24	49.24	2.05	0.00	1	1	0
83	0.07	0.07	0.07	0.07	-0.04	-1.32	31.58	31.58	3.78	0.89	1	1	1
84	7.13	6.16	6.79	0.28	0.52	0.85	6.16	6.16	6.79	0.28	0	0	0
85	0.03	0.03	0.03	0.03	-0.50	-1.83	6.49	6.49	0.11	0.20	1	1	1
86	2.88	2.88	1.04	0.13	-0.09	0.53	5.51	5.51	1.04	0.13	1	0	0
87	0.79	0.79	0.40	0.55	0.70	-0.37	50.81	50.81	0.40	0.55	1	0	0
88	0.32	0.32	0.32	0.32	0.98	-1.05	90.23	90.23	0.63	1.08	1	1	1
89	1.07	1.07	1.07	0.07	-0.56	-0.12	7.83	7.83	14.72	0.07	1	1	0
90	3.02	2.71	3.02	1.34	0.56	0.39	2.71	2.71	8.01	1.34	0	1	0
91	0.02	0.02	0.02	0.02	0.23	-2.00	2.42	2.42	0.78	0.94	1	1	1
92	0.97	0.97	0.78	0.97	-0.70	-0.26	4.86	4.86	0.78	1.29	1	0	1
93	0.04	0.04	0.04	0.04	0.18	-1.69	41.21	41.21	1.85	1.34	1	1	1
94	0.52	0.52	0.52	0.52	1.24	-1.09	21.22	21.22	5.22	0.52	1	1	0
95	3.06	3.06	0.36	1.13	1.73	-0.94	49.68	49.68	0.36	1.13	1	0	0
96	0.13	0.13	0.13	0.06	-0.89	-1.43	41.14	41.14	5.14	0.06	1	1	0
97	0.07	0.07	0.07	0.07	0.40	-1.44	14.23	14.23	1.64	0.64	1	1	1
98	15.53	15.53	1.59	0.11	-0.79	1.06	16.16	16.16	1.59	0.11	1	0	0
99	1.09	1.09	1.09	0.17	2.23	-2.44	14.02	14.02	11.12	0.17	1	1	0
100	2.99	2.99	0.80	0.13	0.93	0.11	53.64	53.64	0.80	0.13	1	0	0

C Real data set with censoring

	Pseudostress	k-Cycles	Status
1	80.3	211.629	F
2	80.6	200.027	F
3	80.8	57.923	C
4	84.3	155.000	F
5	85.2	13.949	F
6	85.6	112.968	C
7	85.8	152.680	F
8	86.4	156.725	F
9	86.7	138.114	C
10	87.2	56.723	F
11	87.3	121.075	F
12	89.7	122.372	C
13	91.3	112.002	F
14	99.8	43.331	F
15	100.1	12.076	F
16	100.5	13.181	F
17	113.0	18.067	F
18	114.8	21.300	F
19	116.4	15.616	F
20	118.0	13.030	F
21	118.4	8.489	F
22	118.6	12.434	F
23	120.4	9.750	F
24	142.5	11.865	F
25	144.5	6.705	F
26	145.9	5.733	F

Low-cycle fatigue life of nickel-base superalloy specimens. Lifetimes are measured in k-Cycles, i.e. units of thousands of cycles. In the status column, cases marked with “F” stands for failures, while the ones marked with “C” are censored.

D R-code

All the code in this section is the version used with the Weibull distributed lifetimes in Section 4. Code for lognormally distributed lifetimes and real data are analogous.

D.1 Covariate order method

Code for the covariate order method. This code was provided by Jan Terje Kvaløy.

```
#covorder.R

sx <- function(xgrid, Xdata, V){
  ngrid <- length(xgrid)
  sx <- vector(length=ngrid)
  for(i in 1:ngrid){
    sx[i] <- tail(c(0, V[Xdata<xgrid[i]]), 1)
  }
  sx
}

epK <- function(u){
  ifelse(abs(u)<1, (3/4)*(1-u^2), 0)
}

# The boundary kernel of Zhang and Karunamuni (JSPI, 1998)
epbK <- function(u, c){
  ifelse((abs(u)<1 & u<c), ((12/((1+c)^4))*(1+u)*((1-2*c)*u+(3*c^2-2*c+1)/2)
), 0)
}

covordn <- function(Xdata, Tdata, ind, h, xgrid=1, glatt="x", splot=F, edge="BK"){
# Sort data
  orden <- order(Xdata)
  ind <- ind[orden]
  Tdata <- Tdata[orden]
  Xdata <- Xdata[orden]

# Calculate basic quantities
  n <- length(Xdata)
  V <- Tdata/n # scaling observations for the constructed poisson process.
  V <- cumsum(V)
  Vmax <- V[n]
  S <- V[ind==1] # S is the observations in the poisson process. The
process consists only of the non-censored observations from Tdata
  X_s <- Xdata[ind==1]
  r <- length(V)
# Make plot of the constructed poisson process.
  if(splot)
  {
    plot(X_s, S, type="s")
    points(X_s, S)
  }
# Calculate sx (the correspondance function)
}
```

```

if(length(xgrid)>1){
  xorder <- order(xgrid)
  invorden <- order(xorder)
  xgrid <- xgrid[xorder]
  tildesx <- sx(xgrid,Xdata,V)
}
if(length(xgrid)==1){
  xgrid <- Xdata
  invorden <- order(orden)
  tildesx <- V
}
nxgrid <- length(xgrid)
# Smooth
lambdaest <- vector(length=nxgrid)
# Boundary kernel
if(edge!="R"){
  if(glatt=="s"){ # smoothing the s-axis
    for(i in 1:nxgrid){
      if(h>Vmax/2)
        h=Vmax/2
      if(tildesx[i]<h)
        lambdaest[i] <- sum(epbK((tildesx[i]-S)/h,tildesx[i]/h))/(n*h)
      if(tildesx[i]>=h & tildesx[i]<=Vmax-h)
        lambdaest[i] <- sum(epK((tildesx[i]-S)/h))/(n*h)
      if(tildesx[i]>Vmax-h)
        lambdaest[i] <- sum(epbK(-(tildesx[i]-S)/h,(Vmax-tildesx[i])/h))/(n*h)
    }
  }
  if(glatt!="s"){ # smoothing the x-axis (covariate axis)
    for(i in 1:nxgrid){
      hsx <- sx(xgrid[i]+h/2,Xdata,V)-sx(xgrid[i]-h/2,Xdata,V)
      if(hsx>Vmax/2)
        hsx=Vmax/2
      if(tildesx[i]<hsx){
        lambdaest[i] <- sum(epbK((tildesx[i]-S)/hsx,tildesx[i]/hsx))/(n*hsx)
      }
      if(tildesx[i]>=hsx & tildesx[i]<=Vmax-hsx){
        lambdaest[i] <- sum(epK((tildesx[i]-S)/hsx))/(n*hsx)
      }
      if(tildesx[i]>Vmax-hsx){
        lambdaest[i] <- sum(epbK(-(tildesx[i]-S)/hsx,(Vmax-tildesx[i])/hsx))/(n*hsx)
      }
    }
  }
}
# Reflection
if(edge=="R"){
  if(glatt=="s"){
    for(i in 1:nxgrid){
      lambdaest[i] <- (sum(epK((tildesx[i]-S)/h))+sum(epK((tildesx[i]+S)/h))+sum(epK((tildesx[i]+S-2*Vmax)/h)))/(n*h)
    }
  }
}

```



```

if(glatt!="s"){
  for(i in 1:nxgrid){
    hsx <- sx(xgrid[i]+h/2,Xdata,V)-sx(xgrid[i]-h/2,Xdata,V)
    lambdaest[i] <- (sum(epK((tildesx[i]-S)/hsx))+sum(epK((tildesx[i]+S
      )/hsx))+sum(epK((tildesx[i]+S-2*Vmax)/hsx)))/(n*hsx)
  }
}
}
list(x=xgrid, lambdaest=lambdaest, x_nosort=xgrid[invorden], lambdaest_
  nosort=lambdaest[invorden], s=S)
}

```

D.2 Anderson-Darling trend test

Code for calculating the Anderson-Darling trend test statistic.

```
#trendtest.R
```

```

AD <- function(x,T,status){
  # Sorting the data
  orden <- order(x)
  status <- status[orden]
  T <- T[orden]
  x <- x[orden]
  n <- length(T)
  S <- cumsum(T)[status==1]
  S_max <- cumsum(T)[n]
  k <- sum(status[1:n-1])

  ln_1 <- 0
  ln_2 <- 0

  for(i in 1:k){
    ln_1 <- ln_1 + (2*i-1)*log(S[i]/S_max)
    ln_2 <- ln_2 + (2*i-1)*log(1-(S[k+1-i]/S_max))
  }

  ad <- -k-(1/k)*(ln_1 + ln_2)

  return(ad)
}

```

D.3 Bootstrapping the Anderson-Darling statistics

Code for non-parametric and parametric bootstrapping to find the distribution of the Anderson-Darling statistics.

```

#bootstrapAD.R
library(survival)
library(VGAM)
library(fields)
library(e1071)

source('covorder.R')
source('trendtest.R')

data <- read.table('weibullDataCens.csv', header= TRUE, sep=',',
                  row.names=1)
data$x_i_sq <- (data$x_i)^2

#Non-parametric bootstrap
obsBoot <- function(data, bs){

  ADz1 <- c()
  ADz2 <- c()
  ADx_sq <- c()
  muVec <- c()
  beta1Vec <- c()
  beta2Vec <- c()
  beta3Vec <- c()
  sigmaVec <- c()

  for(i in 1:bs){
    bsData <- data[sample(100,replace=TRUE),]
    mod <- survreg(Surv(bsData$Y_i,bsData$s_i) ~ bsData$z_i1+
                  bsData$z_i2+bsData$x_i_sq, bsData, dist='weibull')

    # Extracting model coefficients
    muHat <- as.numeric(mod$coeff[1]) #beta_0
    beta1Hat <- as.numeric(mod$coeff[2])
    beta2Hat <- as.numeric(mod$coeff[3])
    beta3Hat <- as.numeric(mod$coeff[4])
    sigmaHat <- as.numeric(mod$scale)

    muVec <- c(muVec, muHat)
    beta1Vec <- c(beta1Vec, beta1Hat)
    beta2Vec <- c(beta2Vec, beta2Hat)
    beta3Vec <- c(beta3Vec, beta3Hat)
    sigmaVec <- c(sigmaVec, sigmaHat)

    # Computing Cox-Snell residuals
    CSRes <- exp((log(bsData$Y_i)-muHat-beta1Hat*bsData$z_i1
                    -beta2Hat*bsData$z_i2-beta3Hat*bsData$x_i_sq)/sigmaHat)

    lambda_z1 <- covordn(bsData$z_i1, CSRes,
                        bsData$s_i, h=opt_h_z1_20, glatt='x', edge='R')
    lambda_z2 <- covordn(bsData$z_i2, CSRes,
                        bsData$s_i, h=opt_h_z2_20, glatt='x', edge='R')
    lambda_x_sq <- covordn(bsData$x_i_sq, CSRes,

```

```

        bsData$s_i, h=opt_h_x_20, glatt='x', edge='R')

        # Computing Anderson-Darling test statistic
        ADz1 <- c(ADz1, AD(bsData$z_i1, CSRes, bsData$s_i))
        ADz2 <- c(ADz2, AD(bsData$z_i2, CSRes, bsData$s_i))
        ADx_sq <- c(ADx_sq, AD(bsData$x_i_sq, CSRes, bsData$s_i))
    }
    list(adz1 = ADz1, adz2 = ADz2, adxsq = ADx_sq, mu=muVec,
        b1=beta1Vec, b2=beta2Vec, b3=beta3Vec, sigma=sigmaVec)
}

#Parametric bootstrap
parBoot <- function(data, bs){
  ADz1 <- c()
  ADz2 <- c()
  ADx_sq <- c()
  muVec <- c()
  beta1Vec <- c()
  beta2Vec <- c()
  beta3Vec <- c()
  sigmaVec <- c()

  Mod <- survreg(Surv(data$Y_i, data$s_i) ~ data$z_i1+
                data$z_i2+data$x_i_sq, data, dist='weibull')
  MuHat <- as.numeric(Mod$coeff[1]) #beta_0
  Beta1Hat <- as.numeric(Mod$coeff[2])
  Beta2Hat <- as.numeric(Mod$coeff[3])
  Beta3Hat <- as.numeric(Mod$coeff[4])
  SigmaHat <- as.numeric(Mod$scale)

  for(i in 1:bs){
    bsData <- data
    u <- runif(100)
    W <- log(-log(u))
    bs_res <- SigmaHat*W
    yTmp <- log(bsData$Y_i)+bs_res
    bsData$Y_i <- exp(yTmp)
    bsMod <- survreg(Surv(bsData$Y_i, bsData$s_i) ~ bsData$z_i1+
                    bsData$z_i2+bsData$x_i_sq, bsData, dist='weibull')

    # Extracting model coefficients
    muHat <- as.numeric(bsMod$coeff[1]) #beta_0
    beta1Hat <- as.numeric(bsMod$coeff[2])
    beta2Hat <- as.numeric(bsMod$coeff[3])
    beta3Hat <- as.numeric(bsMod$coeff[4])
    sigmaHat <- as.numeric(bsMod$scale)

    muVec <- c(muVec, muHat)
    beta1Vec <- c(beta1Vec, beta1Hat)
    beta2Vec <- c(beta2Vec, beta2Hat)
    beta3Vec <- c(beta3Vec, beta3Hat)
    sigmaVec <- c(sigmaVec, sigmaHat)
  }
}

```

```

# Computing Cox-Snell residuals
CSRes <- exp((log(bsData$Y_i)-muHat-beta1Hat*bsData$z_i1
-beta2Hat*bsData$z_i2-beta3Hat*bsData$x_i_sq)/sigmaHat)

lambda_z1 <- covordn(bsData$z_i1, CSRes,
bsData$s_i, h=opt_h_z1_20, glatt='x', edge='R')
lambda_z2 <- covordn(bsData$z_i2, CSRes,
bsData$s_i, h=opt_h_z2_20, glatt='x', edge='R')
lambda_x_sq <- covordn(bsData$x_i_sq, CSRes,
bsData$s_i, h=opt_h_x_20, glatt='x', edge='R')

# Computing Anderson-Darling test statistic
ADz1 <- c(ADz1, AD(bsData$z_i1, CSRes, bsData$s_i))
ADz2 <- c(ADz2, AD(bsData$z_i2, CSRes, bsData$s_i))
ADx_sq <- c(ADx_sq, AD(bsData$x_i_sq, CSRes, bsData$s_i))
}
list(adz1 = ADz1, adz2 = ADz2, adxsq = ADx_sq, mu=muVec,
b1=beta1Vec, b2=beta2Vec, b3=beta3Vec, sigma=sigmaVec)
}

```

D.4 Likelihood cross-validation

Code for using likelihood cross-validation to find the optimal smoothing parameters.

```
# lcv.R
```

```
library(splines)
```

```
lcv <- function(H, cov){
  data <- read.table('weibullDataCens.csv', header=TRUE, sep=',',
    row.names=1)
  x_i_sq <- data$x_i^2
  data$x_i_sq <- x_i_sq
  n <- length(x_i_sq)

  Model <- survreg(Surv(data$Y_i, data$s_i)~data$z_i1+data$z_i2+data$x
    _i_sq, dist='weibull')
  MuHat <- as.numeric(Model$coeff[1])
  Beta1Hat <- as.numeric(Model$coeff[2])
  Beta2Hat <- as.numeric(Model$coeff[3])
  Beta3Hat <- as.numeric(Model$coeff[4])
  SigmaHat <- as.numeric(Model$scale)

  CSRES <- exp((log(data$Y_i) - MuHat - Beta1Hat*data$z_i1 - Beta2Hat
    *data$z_i2 - Beta3Hat*data$x_i_sq)/SigmaHat)

  if(cov=='z1'){
    x <- data$z_i1
  }
}

```

```

else if(cov=='z2'){
  x <- data$z_i2
}
else if(cov=='c'){
  x <- data$x_i
}

ICV <- 0

for(i in 1:n){
  # Using cov.ord. to estimate lambda without observation no.
  # 'i'
  lambdaEst <- covordn(x[-i], CSRES[-i], data$s_i[-i], H, glatt='
s', edge='R')

  # Fitting a spline function to the estimated lambdas. This
  # in order to predict the lambdaHat for observation 'i'.
  spline <- polySpline(interpSpline(lambdaEst$x, lambdaEst$
lambdaest))
  predSpline <- predict(spline, x[i])
  lHat_i <- predSpline$y

  # Adding to the ICV
  ICV <- ICV + (data$s_i[i]*log(lHat_i)-lHat_i*CSRES[i])
}
return(ICV)
}

```