An Evaluation of Methods for Assessing the Functional Form of Covariates in the Cox Model

By Linnea Karlsson

Master Thesis 15 cr
Department of Statistics
Uppsala University

Supervisor: Inger Persson

2016
Abstract

In this thesis, two methods for assessing the functional form of covariates in the Cox proportional hazards model are evaluated. The methods include one graphical check based on martingale residuals and one graphical check, together with a Kolmogorov-type supremum test, based on cumulative sums of martingale residuals. The methods are evaluated in a simulation study under five different covariate misspecifications with varying sample sizes and censoring degrees. The results from both methods indicate that the type of covariate misspecification, sample size and censoring degree affect the ability to detect and identify the misspecification. The procedure based on smoothed scatterplots of martingale residuals reveals difficulties with assessing whether the behaviour of the smoothed curve in the plot is an indication of a misspecification or a phenomenon that can occur in a correctly specified model. The graphical check together with the test procedure based on cumulative sums of martingale residuals is shown to successfully detect and identify three out of five covariate misspecifications for large sample sizes. For small sample sizes, especially combined with a high censoring degree, the power of the supremum test is low for all covariate misspecifications.

KEYWORDS: Cox model, martingale residuals, cumulative sums of martingale residuals, functional form, misspecification, smoothed scatterplots, supremum test.
Contents

1 Introduction 1

2 Theoretical Background 3
   2.1 The Cox Proportional Hazards Model 3
   2.2 Martingale Residuals 4
   2.3 Cumulative Sums of Martingale Residuals 6

3 Simulation Results 8
   3.1 Simulation Set-up 8
   3.2 Martingale Residuals 9
   3.3 Cumulative Sums of Martingale Residuals 17

4 Discussion 24

5 Appendix 29
1 Introduction

The Cox proportional hazards (PH) model (1972) is widely used in survival analysis to investigate how survival time depends on a set of explanatory covariates. In survival analysis focus lies on the risk, or hazard, of death at any time after the time of origin in the study. When including a continuous covariate, \( Z \), in the model it is important to consider whether it is appropriate to include it as a linear term or if a transformation, for instance \( Z^2 \) or \( \sqrt{Z} \), is a better fit. If the functional form of a covariate is misspecified it can lead to a loss in efficiency, biased estimators and detrimental effects on validity (Lagakos & Schoenfeld, 1984; Lagakos, 1988, Lin & Wei, 1989; Gers & Schumacher, 2001). It is therefore important to have reliable procedures for detecting possible covariate misspecifications in the model.

A number of methods for assessing the functional form of covariates in the Cox model have been proposed in the literature. Therneau et al. (1990) suggested a graphical approach, where the martingale residuals are computed from the null Cox model and then plotted against each covariate in the model separately with a superimposed scatterplot smooth. A couple of refinements of mentioned procedure were introduced by Grambsch et al. (1995), who suggested diagnostic plots based on a Poisson regression approach. Lin et al. (1993, 2002) developed a graphical check based on cumulative sums of martingale residuals plotted over the values of each covariate separately. This graphical check was introduced together with a test procedure for detecting covariate misspecifications. Other forms of test procedures have been developed by authors such as Verweij et al. (1998), who proposed a goodness-of-fit test based on martingale residuals for the presence of an extra random effect in the exponential part of the hazard function. León & Tsai (2004) introduced a test procedure based on a new class of residuals, called censoring consistent residual, for detection of non-linearity of the functional form of a covariate.

The aim of this thesis is to evaluate two of the existing methods for assessing the functional forms of covariates in the Cox model. Focus will be on the procedures based on martingale residuals proposed by Therneau et al. (1990) and Lin et al. (1993). These methods are evaluated under a set of covariate misspecifications with varying sample sizes and censoring degrees using simulations. The analysis also includes the correctly specified covariate as a point of reference. This thesis seeks to assess to what extent the methods are able to detect and identify different covariate misspecifications in the Cox.
PH model.

The paper begins with a brief introduction of the theoretical background behind the Cox PH model and the martingale residuals in section 2. The simulation set-up is described in section 3, together with a presentation of the results. In section 4 a discussion of the results will be presented.
2 Theoretical Background

In this section a description of the theoretical background behind the Cox PH model and martingale residuals is presented.

2.1 The Cox Proportional Hazards Model

The Cox model lets the investigator explore how survival time, or equally the risk of experiencing an event, depends on a set of covariates whose values can be either time-fixed at the start of the study or time-dependent. Let \( X \) denote the random variable "survival time", or equally "time to failure". If the event of interest has not been observed for an individual at the end of the study, the survival time is censored. The censoring time is denoted by \( C \). The only information available for a censored observation is the last date when the individual was known to be alive. A crucial assumption is that \( X \) and \( C \) are independent, i.e. the censoring is random, or non-informative. For this thesis, only type-I censoring is considered, which occurs when the censoring time takes place after entering the study (Collett, pp.2-3).

If an individual experiences the event during the study time, then the event indicator \( \delta \) is set to unity and if the observation is censored, then \( \delta \) is zero. Let \( X_i \) and \( C_i \) be the survival and censoring times and \( Z_i = (Z_{i1}, ..., Z_{ip}) \) be a time-fixed \((p \times 1)\) covariate vector for individual \( i = 1, ..., n \). The survival data then consists of \((T_i, \delta_i, Z_i)\), where \( T_i = min(X_i, C_i) \) is follow-up time.

The hazard rate at time \( t \) for individual \( i \) with covariate vector \( Z_i \) is expressed with the following Cox model:

\[
h(t \mid Z_i) = h_0(t) \exp(\beta'Z_i) \tag{1}\]

where \( h_0(t) \) is an arbitrary baseline hazard rate. The risk vector \( Z_i \) can consist of both factors or continuous covariates. The \( \beta \)-coefficients in the model are unknown parameters that can be estimated by the "method of maximum likelihood" (Collett, p.65). The Cox model does not assume any functional form for the baseline hazard, instead it is estimated from the data. However, if the failure time distribution is known, the investigator can allow the baseline hazard of the model to follow some probability distribution, for instance an Exponential, Weibull or Gamma distribution (Klein & Moeschberger, pp. 244; 393-404).
The logarithm of \( h(t \mid Z_i)/h_0(t) \) is equal to \( \sum_{j=1}^{p} \hat{\beta}_j Z_{ij} \), which formulates the effects of the covariates, in the same way as for usual linear models. One important assumption of the Cox PH model is that the ratio of two individuals' hazard rates are proportional over time (Klein & Moeschberger, pp. 244-245).

2.2 Martingale Residuals

Due to the censored observations present in survival data, the computation of residuals is not as straightforward as for regular linear regression models. Different types of residuals have therefore been developed to investigate various kinds of model misspecifications. Among the major ones are the martingale, deviance, score, and Schoenfeld residuals. In this thesis, focus will be on the martingale residuals that mainly have two fields of application. The first one is to reveal possible "excess events" in the data, i.e. individuals that poorly fit the model, and the second one is to assess whether a specific covariate in the model has an appropriate functional form or not (Therneau & Grambsch, pp. 79-80; 87-88).

The martingale residuals derive from a counting process approach, which counts the occurrences of some event over the time period \([0, t)\) for \( i = 1, \ldots, n \) individuals. The process is defined as the sequence of the random variable \( N_i(t), t \geq 0 \), which is a stochastic process where \( N_i(0) = 0 \) and \( N_i(t) < \infty \) with probability 1. In a right-censored sample, the counting process \( N_i(t) = I[T_i \geq t, \delta_i = 1] \), is zero until the observed event occur for individual \( i \) and then increases to one. The counting process \( N(t) = \sum_{i=1}^{n} N_i(t) = \sum_{t_i \geq t} \delta_i \) can be regarded as a sum of events in the sample at or prior to time \( t \) (Klein & Moeschberger, section 3.6; Collett, pp. 429-430).

To define the martingale residual, it is assumed that for the \( i \)th individual in the sample, there is a risk vector \( Z_i(t) \) with corresponding vector of regression coefficients \( \beta' \). The counting process \( N_i(t) \) starts at zero, since the individual is at risk at the beginning of the study and increases to unity at the event time for the \( i \)th individual. If the individual has a censored survival time then the counting process remains zero. Let \( Y_i(t), t \geq 0 \) be defined as an at-risk process, where \( Y_i(t) = 1 \) until the event occurs or the survival time is censored, and zero thereafter (Collett, pp. 430-431). Following Klein & Moeschberger (pp. 359-360) the martingale residuals are then defined as the difference between the
counting process and the integrated intensity function:

\[ \hat{M}_i = N_i(\infty) - \int_0^\infty Y_i(t) \exp(\hat{\beta}'Z_i(t))d\hat{H}_0(t) \]  

(2)

where \( \hat{H}_0(t) \) is the Breslow estimator of the cumulative baseline hazard rate defined as:

\[ \hat{H}_0(t) = \sum_{i \leq t} \frac{d_i}{\sum_{j \in R(t_i)} \exp(\hat{\beta}'Z_i)} \]  

(3)

where \( d_i \) is the number of deaths at time \( t_i \) (Klein & Moeschberger, p. 283). Given that the data is right-censored and that there are no time-dependent covariates, the martingale residuals are reduced to:

\[ \hat{M}_i = \delta_i - \hat{H}_0(T_i) \exp(\sum_{j=1}^{p} \hat{\beta}_j Z_{ij}), \quad i = 1, \ldots, n \]  

(4)

The martingale residuals can be interpreted as the difference between the observed number of events for the \( i \)th individual over the interval \([0, t_i]\) and the expected number of events based on the estimated model. This can also be expressed as the excess number of events seen in the data but not predicted by the model. The residuals are highly skewed since they can take the values \((-\infty, 1]\), where the residuals from censored observations \((\delta_i = 0)\) are negative or zero. The martingale residuals have properties similar to the residuals in ordinary linear models, such as \( \sum_{i=1}^{n} \hat{M}_i(t) = 0 \) for any \( t \) and \( E(\hat{M}_i(t)) = 0 \) and \( \text{cov}(\hat{M}_i(t), \hat{M}_j(t)) = 0 \) \((i \neq j)\) for large samples (Klein & Moeschberger, p. 360; Therneau & Grambsch, pp. 80-81; Lin et al., 1993).

In the graphical checking of the functional form of a covariate originally proposed by Therneau et al. (1990), the martingale residuals are obtained by fitting a null Cox model, i.e. a model without any covariates. The residuals are then plotted against the covariates separately with a superimposed smoothed curve. An alternative approach can be used if the functional form of some covariates in the model are assumed to be known. In such case the time-fixed covariate vector \( Z \) is partitioned into a vector \( Z^* \) for which we know the proper functional form of and a single covariate \( Z_1 \), assumed to be independent of \( Z^* \), for which we are unsure of what functional form to use. If \( f(Z_1) \) is the best function of \( Z_1 \) to explain its effect on survival time, then the optimal Cox model is:

\[ h(t \mid Z^*, Z_1) = h_0(t) \exp(\hat{\beta}'Z^*) \exp(\hat{\beta}_1 f(Z_1)) \]  

(5)

To find the function \( f(\cdot) \), the martingale residuals \( \hat{M}_i, i = 1, \ldots, n \) are computed after fitting a Cox model to the data based on \( Z^* \) (Klein & Moeschberger, pp. 359-360). Therneau & Grambsch (2000) show that under certain conditions, such as low to moderately
censored data, a smoothed plot of the martingale residuals against the covariate values will reveal the functional form of $Z_1$. Potential problems with the smoothed scatterplots have been pointed out, such as biasedness for large covariate effects and misleading results if the covariates in the model are correlated (Therneau & Grambsch, 2000).

### 2.3 Cumulative Sums of Martingale Residuals

As a tool for model checking of the Cox model presented in equation (1), Lin et al. (1993) introduced the multiparameter stochastic process:

$$W_z(t, z) = \sum_{i=1}^{n} f(Z_i)I(Z_{ij} \leq z)\hat{M}_i, (j = 1, \ldots, p)$$  \hspace{1cm} (6)

where $f(\cdot)$ is a known smooth function and all the $p$ components of $Z_i$ are no larger than the respective components of $z = (z_1, \ldots, z_p)$. If equation (6) holds, then the processes $W_z(\cdot)$ will fluctuate randomly around zero.

For checking the accuracy of the functional form of covariates in the Cox model, a special case of equation (6) has been developed. The partial-sum processes of the martingale residuals are then defined as:

$$W_j(x) = \sum_{i=1}^{n} I(Z_{ij} \leq x)\hat{M}_i, (-\infty < x < \infty, j = 1, \ldots, p)$$  \hspace{1cm} (7)

The stochastic processes under the model are assumed to follow the distribution of a zero-mean Gaussian process, which are realized through simulations. The observed processes $w_j(\cdot)$ are compared both graphically and numerically to assess whether a trend in the residual plot reflects a misspecification of the functional form or if it is due to natural variation. The extremity of the observed processes is measured as $s_j = sup_x|w_j(x)|$. If the value of $s_j$ is unusually large it suggests that the functional form of the covariate of interest may be inappropriate.

In the graphical checking the observed processes are plotted together with a number of simulated processes from the null distribution in order to detect atypical patterns in the observed processes. A Kolmogorov-type supremum test, in this thesis called "supremum test", tests whether the extremity of the observed process differs significantly from the simulated processes. In the test procedure, the p-value $Pr(S_j \geq s_j)$, which is approximated by $Pr(\hat{S}_j \geq s_j)$ where $\hat{S}_j = sup_x|\hat{W}_j(x)|$, assesses how unusual the observed residual pattern is under the null hypothesis that it follows a zero-mean Gaussian
process. The authors show that $S_j$ is consistent against incorrect functional forms for the covariate at interest provided that there is no model misspecification and that the covariate under study is independent of all other covariates included in the model.
3 Simulation Results

The first subsection, 3.1, gives a description of the simulation set-up, followed by subsection 3.2 which contains plots generated from the raw martingale residuals. In subsection 3.3 the results from the graphical check and test procedure using cumulative martingale residuals are presented. Due to the large amount of output from the simulations, parts of the results are put in appendix and the rest is omitted.

3.1 Simulation Set-up

The procedures for assessing the functional form of covariates, described in section 2, are evaluated under different covariate misspecifications using a simulation study. For this purpose a two-covariate Cox PH model is applied with the following hazard function:

\[
h(t \mid Z_1, Z_2) = \lambda \exp(\alpha t) \exp(\beta_1 Z_1 + \beta_2 Z_2)
\]

where \( h_0(t) = \lambda \exp(\alpha t) \) is the baseline hazard function. Following the model proposal by Bender et al. (2005), the scale parameter is set to \( \lambda = 7/(10^8) \) and the shape parameter is set to \( \alpha = 0.2138 \). The time-fixed covariate values for \( Z = (Z_1, Z_2) \) are independently and randomly generated from a normal distribution with parameters \( \mu = (24.3, 266.8) \) and \( \sigma = (8.4, 507.8) \). \( Z_1 \) is intended to resemble an age covariate and \( Z_2 \) is designed to reflect radon exposure. The coefficient values are set to \( \beta = (0.15, 0.001) \), generating hazard ratios of approximately 1.16 and 1.001 respectively. Random survival times, \( X_i \), are generated from a Gompertz distribution, given the set of covariates and parameter values stated above. The Gompertz distribution is chosen since it is an appropriate distribution for modeling human mortality (Bender et al., 2005). The censoring times, \( C_i \), are generated randomly and independent of \( X_i \) over the interval \([0, \tau]\), where \( \tau \) is the longest survival time in the generated datasets, to fulfill the assumption of non-informative type-I censoring.

1000 datasets are simulated with sample sizes of \( n = (50, 200, 500) \) together with censoring rates of 20 and 60 per cent. The sample sizes are chosen in order to represent small, medium and large samples in medical studies respectively and the censoring rates are chosen to reflect mild and heavy censoring.

The evaluation of the two methods is conducted under five arbitrarily chosen covariate misspecifications presented in table 1. The raw martingale residuals are obtained from fit-
ting model (8) excluding covariate $Z_1$, as described in section 2.2, and then plotted against the each chosen covariate misspecifications separately. A LOWESS (locally weighted scatterplot smoothing) scatterplot smoother is used, which uses a locally-weighted polynomial regression (Cleveland, 1979, 1981). The smoothing parameter is set to $2/3$, which is the default value for lowess() in R.

### Table 1: Covariate misspecifications used in the simulations. The column named ‘Data’ represents the functional form of the covariate used when the survival times were generated, i.e. the correct functional form.

<table>
<thead>
<tr>
<th>Data</th>
<th>Misspecification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$Z_1$</td>
</tr>
<tr>
<td>2</td>
<td>$Z_1^2$</td>
</tr>
<tr>
<td>3</td>
<td>$\sqrt{Z_1}$</td>
</tr>
<tr>
<td>4</td>
<td>$1/\sqrt{Z_1}$</td>
</tr>
<tr>
<td>5</td>
<td>$I(Z_1 &gt; 26)$</td>
</tr>
</tbody>
</table>

The plots based on cumulative sums of martingale residuals are accompanied with p-values from a Kolmogorov-type supremum test as presented in section 2.3. From a total of 1 000 simulated paths, the p-value gives the proportion of these paths that has an extreme point (suprema) exceeding the most extreme point in the observed path (Allison, pp. 173-174). As an evaluation of the supremum test, the empirical power and size of the test are calculated from generating 1 000 p-values for each covariate misspecification under the different sample sizes and censoring degrees. A standard significance level of 5% is used, i.e. the null hypothesis is rejected when the p-value falls below 0.05.

### 3.2 Martingale Residuals

Following the simulation scheme, 1000 plots are generated for each covariate misspecification presented in table 1. Results are also presented for the correct functional form of the covariate ($Z_1$), as a point of reference. The blue dots in the scatterplots represent the computed martingale residuals and the red line represents the LOWESS curve, also called “smoothed curve” or “scatterplot smoother” in this thesis.

The plots presented in figure 1 to 12 and in appendix (figure 22 to 57) reveal some
general patterns. The smoothed curve appears to behave differently depending on how
the covariate is misspecified. The behaviour of the scatterplot smoother is generally more
stable over the generated samples when the sample size is large (n=500). Here, stability
refers to a curve that does not show considerable differences in behaviour between the
plots generated from the simulated samples. For the smallest sample size (n=50) the
scatterplot smoother has a more fluctuating behaviour. Comparing the plots generated
from data with mild (20%) and heavy (60%) censoring, there is a difference in how the
residuals are grouped in the plots. However, there is no apparent differences regarding
the stability of the smoothed curve’s behaviour.

The plots presented in figure 1 and 2 show that when the martingale residuals are
plotted against the correctly specified covariate (Z₁) the smoothed curve does not appear
completely linear for the majority of the plots. The upward slope corresponds to the posi-
tive coefficient in the fitted model in equation (8) (Collett, p. 150). The smoothed curve
has a fairly similar appearance for covariate misspecification log(Z₁), presented in figures
3 and 4, and √Z₁, presented in figures 7 and 8. When the residuals are plotted against
misspecification Z₁², see figures 5 and 6, the behaviour of the smoothed curve is relatively
stable over the studied samples for all sample sizes and censoring degrees. Figures 9 and 10
demonstrate that the misspecification 1/√Z₁ deviates the most from the correctly speci-

fied covariate with a negative slope of the smoothed curve. Figures 11 and 12 show that
the scatterplot smoother has a roughly constant slope when the residuals are incorrectly
plotted against Z₁, that is when a discretized version of the covariate should be used.
Figure 1: Plots of martingale residuals against the correctly specified covariate with 20% censoring in the data.

Figure 2: Plots of martingale residuals against the correctly specified covariate with 60% censoring in the data.
Figure 3: Plots of martingale residuals against the covariate misspecified as $\log(Z_1)$ with 20% censoring in the data.

Figure 4: Plots of martingale residuals against the covariate misspecified as $\log(Z_1)$ with 60% censoring in the data.
Figure 5: Plots of martingale residuals against the covariate misspecified as $Z^2_i$ with 20% censoring in the data.

Figure 6: Plots of martingale residuals against the covariate misspecified as $Z^2_i$ with 60% censoring in the data.
Figure 7: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z_1}$ with 20% censoring in the data.

Figure 8: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z_1}$ with 60% censoring in the data.
Figure 9: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 20% censoring in the data.

Figure 10: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 60% censoring in the data.
Figure 11: Plots of martingale residuals against the covariate misspecified as $Z_1$ with 20% censoring in the data.

Figure 12: Plots of martingale residuals against the covariate misspecified as $Z_1$ with 60% censoring in the data.
3.3 Cumulative Sums of Martingale Residuals

In table 2 the rejection rates for the supremum test are presented together with a graphical summary in figure 13. When the covariate is misspecified a high rejection rate is desired since it implies that the test is able to detect incorrect functional forms. In other words, it indicates a high power of the test. Some general tendencies can be seen from table 2 and figure 13. For example, larger sample sizes generally increases the power of the test, whereas a higher censoring degree decreases the power. However, the results vary a lot depending on the covariate misspecification which is illustrated in figure 13. The fourth misspecification \((1/\sqrt{Z_1})\) shows the best results in terms of power of the test for all sample sizes and censoring degrees.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>20 % censoring</th>
<th>60 % censoring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=50</td>
<td>n=200</td>
</tr>
<tr>
<td>log((Z_1))</td>
<td>0.065</td>
<td>0.463</td>
</tr>
<tr>
<td>(Z_1^2)</td>
<td>0.078</td>
<td>0.503</td>
</tr>
<tr>
<td>(\sqrt{Z_1})</td>
<td>0.034</td>
<td>0.104</td>
</tr>
<tr>
<td>(1/\sqrt{Z_1})</td>
<td>0.138</td>
<td>0.787</td>
</tr>
<tr>
<td>(I(Z_1 &gt; 26))</td>
<td>0.034</td>
<td>0.057</td>
</tr>
<tr>
<td>(Z_1)</td>
<td>0.038</td>
<td>0.043</td>
</tr>
</tbody>
</table>

For misspecification \(Z_1\), i.e. when the correct functional form is \(I(Z_1 > 26)\), the supremum test has the lowest power for all sample sizes and censoring degrees. This result is consistent with the scatterplots shown in figure 11 and 12, where the smoothed curve gave no clear indication of any misspecification. The last row, where the rejection rates for the correctly specified covariate are presented, shows that the empirical size of the test is smaller than the pre-stated significance level at 5 % for most cases.
Figure 13: The power of supremum test for different covariate misspecifications.

Figure 14 on the following page presents plots of 50,000 simulated martingale residuals grouped cumulatively over the covariate values in order to illustrate the pattern of the observed processes, presented in equation (7). Results are presented for each covariate misspecification as well as the correctly specified covariate with 20% censoring in the data. The corresponding results from simulations with 60% censoring in the data can be found in figure 21 in the appendix. The plots show that each covariate misspecification exhibits a unique pattern. However, comparing plot a) and c) in figure 14, one can conclude that the pattern from the misspecification \( \sqrt{Z_1} \) does not differ much from \( \log(Z_1) \) and mentioned plots have a similar appearance to plot f) for the correctly specified covariate. The same similarities was also noticed in the smoothed scatterplots. Plot e) shows a sharp drop at the indicator value \( z_1 = 26 \) for the fifth misspecification \( Z_1 \) where the correct functional form is \( I(Z_1 > 26) \). However, the difference between largest and smallest values on the y-axis is considerably smaller compared with the other covariate misspecifications, which shows that the extremity of the observed processes is relatively low.
The figures on the following pages contain plots of the observed processes of cumulative sums of martingale residuals together with the first 20 simulated realizations. Since the graphical appearance does not differ substantially between the plots generated from datasets with mild and heavy censoring, only the results from data with 20% censoring are presented. If the functional form of a covariate has been correctly specified then the observed processes, presented by the solid black line, should lie within the range of the simulated realizations fluctuating randomly around zero. This is illustrated in figure 15 below, where the processes are plotted against the correctly specified covariate.
Figure 15: Cumulative sums of martingale residuals plotted against the correctly specified covariate. Sample size is $n=200$ with 20% censoring. The p-value from supremum test is 0.9010.

In figures 16 to 20, the observed and simulated processes are plotted against each covariate misspecification. When the observed processes deviate significantly from the simulated realizations, which is the case when the p-value is less than 5%, they generally display a similar appearance to the corresponding plots in figure 14. As an example, compare plot a) in figure 14 and plot b) in figure 16. The similarities are however less apparent for misspecification $\sqrt{Z_1}$ in figure 18, compared to plot c) in figure 14, and misspecification $I(Z_1 > 26)$ in figure 19, compared to plot e) in figure 14. Some of the plots accompanied with p-values greater than 5% still exhibit a pattern somewhat similar to the corresponding plots in figure 14. This can be seen in plot c) in figure 17, plot e) in figure 18, plot a) and c) in figure 19.
Figure 16: Cumulative martingale residuals plotted against misspecification $\text{log}(Z_i)$ with 20% censoring. The first two plots are generated from sample size $n=50$, the following two plots are generated from sample size $n=200$ and the last two plots are generated from sample size $n=500$. The $p$-values from the supremum test are given from top left to bottom right: 0.021, 0.01, 0.241, 0.006, 0.001, 0.004.

Figure 17: Cumulative martingale residuals plotted against misspecification $Z_i^2$ with 20% censoring. The first two plots are generated from sample size $n=50$, the following two plots are generated from sample size $n=200$ and the last two plots are generated from sample size $n=500$. The $p$-values from the supremum test are given from top left to bottom right: 0.028, 0.009, 0.144, 0.017, 0.019, <0.0001.
Figure 18: Cumulative martingale residuals plotted against misspecification $\sqrt{Z_1}$ with 20% censoring. The first two plots are generated from sample size $n=50$, the following two plots are generated from sample size $n=200$ and the last two plots are generated from sample size $n=500$. The p-values from the supremum test are given from top left to bottom right: 0.125, 0.018, 0.009, 0.001, 0.197, 0.042

Figure 19: Cumulative martingale residuals plotted against misspecification $1/\sqrt{Z_1}$ with 20% censoring. The first two plots are generated from sample size $n=50$, the following two plots are generated from sample size $n=200$ and the last two plots are generated from sample size $n=500$. The p-values from the supremum test are given from top left to bottom right: 0.135, 0.039, 0.087, 0.023, <0.0001, <0.0001
Figure 20: Cumulative martingale residuals plotted against misspecification $Z_1$ with 20% censoring. The first two plots are generated from sample size $n=50$, the following two plots are generated from sample size $n=200$ and the last two plots are generated from sample size $n=500$. The $p$-values from the supremum test are given from top left to bottom right: 0.009, 0.036, 0.002, 0.048, 0.225, 0.091.
4 Discussion

In this thesis two methods for assessing the functional form of covariates in the Cox model have been evaluated. The aim was to assess the ability of these methods to detect and identify different covariate misspecifications under varying sample sizes and censoring degrees.

According to Therneau & Grambsch (2000), a smoothed scatterplot of the martingale residuals should reveal the functional form of a covariate provided that the censoring degree is low to moderate. The simulation results in this thesis however raises questions regarding the reliability of such smoothed scatterplots. In this analysis, a reliable method refers to one where the behaviour of the smoothed scatterplots does not change substantially between the generated samples, but have a fairly stable behaviour. Furthermore, the smoothed curve should have a unique behaviour for each covariate misspecification in order for the researcher to successfully identify which transformation is needed. When the simulations are conducted with small sample sizes, the generated smoothed scatterplots exhibit a fairly unstable behaviour for the majority of the covariate misspecifications. For the largest sample size (n=500) the behaviour of the smoothed curve is more stable.

The smoothed scatterplots for covariate misspecification $Z_1^2$ generated a fairly stable and unique behaviour for medium and large sample sizes, indicating an ability of the method to detect and identify this misspecification. The same conclusion holds for misspecification $1/\sqrt{Z_1}$. However, some covariate misspecifications, such as $\log(Z_1)$ and $\sqrt{Z_1}$ might be difficult to identify, since the plots exhibit similar trends. Furthermore, the scatterplots of mentioned misspecifications might be hard to distinguish from the scatterplots of the correctly specified covariate. The results also indicate that it might be difficult to detect covariate misspecification $Z_1$, where the correct functional form is $I(Z_1 > 26)$, since the smoothed curve has a roughly constant slope in the majority of the plots when sample size is medium or large. Since the smoothed scatterplots for the correctly specified covariate is not completely linear, there is also a risk of incorrectly concluding that the covariate should be transformed. These results shed light upon difficulties with assessing whether the behaviour of the smoothed scatterplot is an indication of an actual misspecification or a phenomenon that can occur in a correctly specified model. Similar drawbacks with the smoothed scatterplots have been pointed out by other authors, such as Lin et al. (1993).
The plots based on cumulative sums of martingale residuals accompanied with a Kolmogorov-type supremum test is intended to provide a more objective check of the functional form. However, the results show that the supremum test has low power for small sample sizes, especially combined with a high censoring degree. The rejection rates for the correctly specified covariate were in most cases under the significance level 0.05, which indicates that the test is conservative. One general pattern seen from the supremum test is that the power increases with larger sample size and decreases when censoring is high. However large differences can be noticed between the covariate misspecifications. The power of the test is for example surprisingly low when the covariate is incorrectly included as a linear term, with the correct functional form defined as $I(Z_1 > 26)$. For this misspecification both the supremum test and the graphical check were in most cases not able to detect the misspecification, which was also the case for the procedure based on scatterplots.

When the supremum test yields significant results ($p$-value < 0.05), the observed process exhibit a specific pattern for each misspecification. This may provide the investigator with a useful tool for identifying which transformation is needed. Exceptions are misspecifications $\log(Z_1)$ and $\sqrt{Z_1}$, where the observed processes show fairly similar trends, in line with the results from the smoothed scatterplots. In summary, the procedure proposed by Lin et al. (1993) can be seen as a reliable method when sample size is large for misspecifications $Z_1^2$, $\log(Z_1)$ and $1/\sqrt{Z_1}$, since the test has high power irrespective of censoring degree.

The simulation study conducted in this thesis has some limitations. The conclusions are to some degree based on a subjective evaluation due to the lack of objective measures in the smoothed scatterplots. Moreover, since the simulation study was limited to only one Cox model together with five arbitrarily chosen covariate misspecifications, one should guard against generalizing the results too broadly. A quick test with a couple of different models, not presented in this paper, indicated that other Cox models might yield different results. For further research, it might be interesting to extend the evaluation to other Cox models and also to test if time-dependent and correlated covariates together with other levels of the smoothing parameter affect the results. Despite these limitations, the results show some general tendencies of how the methods work in different situations. The results presented in this paper show that the type of covariate misspecification,
sample size and censoring degree have an impact on the detectability and identification of covariate misspecifications. One might therefore want to complement these methods with other checks before conclusions are drawn regarding the functional form of covariates in the Cox model.
References


Figure 21: Cumulative sums of martingale residuals with 60% censoring in the data
Figure 22: Plots of martingale residuals against the correctly specified covariate with 20% censoring presented in the data
Figure 23: Plots of martingale residuals against the correctly specified covariate with 60% censoring presented in the data.
Figure 24: Plots of martingale residuals against the correctly specified covariate with 20% censoring presented in the data.
Figure 25: Plots of martingale residuals against the correctly specified covariate with 20% censoring presented in the data.
Figure 26: Plots of martingale residuals against the correctly specified covariate with 20% censoring presented in the data.
Figure 27: Plots of martingale residuals against the correctly specified covariate with 20% censoring presented in the data.
Figure 28: Plots of martingale residuals against the covariate misspecified as $\log(Z)$ with 20% censoring presented in the data.
Figure 29: Plots of martingale residuals against the covariate misspecified as \( \log(Z) \) with 60% censoring present in the data.
Figure 30: Plots of martingale residuals against the covariate misspecified as $\log(Z)$ with 20% censoring presented in the data.
Figure 31: Plots of martingale residuals against the covariate misspecified as $\log(Z)$ with 60% censoring presented in the data
Figure 32: Plots of martingale residuals against the covariate unspecified as $\log(Z)$ with 60% censoring presented in the data
Figure 33: Plots of martingale residuals against the covariate misspecified as \( \log(Z) \) with 60% censoring presented in the data.
Figure 34: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 20% censoring presented in the data.
Figure 35: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 60% censoring presented in the data
Figure 36: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 20% censoring presented in the data.
Figure 37: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 60% censoring presented in the data.
Figure 38: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 20% censoring presented in the data
Figure 39: Plots of martingale residuals against the covariate misspecified as $Z^2$ with 60% censoring present in the data.
Figure 40: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 20% censoring presented in the data.
Figure 41: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 60% censoring presented in the data.
Figure 42: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 20% censoring presented in the data.
Figure 43: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 60% censoring presented in the data.
Figure 44: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 20% censoring presented in the data.
Figure 45: Plots of martingale residuals against the covariate misspecified as $\sqrt{Z}$ with 60\% censoring presented in the data.
Figure 46: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 20% censoring presented in the data.
Figure 47: Plots of martingale residuals against the covariate unspecified as $1/\sqrt{Z}$ with 60% censoring presented in the data.
Figure 48: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 20% censoring presented in the data.
Figure 49: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 60% censoring presented in the data.
Figure 50: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 20% censoring presented in the data.
Figure 51: Plots of martingale residuals against the covariate misspecified as $1/\sqrt{Z}$ with 60% censoring presented in the data.
Figure 52: Plots of martingale residuals against the covariate misspecified as $Z$ with 20% censoring presented in the data.
Figure 53: Plots of martingale residuals against the covariate misspecified as $Z$ with 60% censoring presented in the data.
Figure 54: Plots of martingale residuals against the covariate misspecified as $Z$ with 20% censoring presented in the data.
Figure 55: Plots of martingale residuals against the covariate misspecified as $Z$ with 60% censoring presented in the data.
Figure 56: Plots of martingale residuals against the covariate misspecified as $Z$ with 20% censoring presented in the data.
Figure 57: Plots of martingale residuals against the covariate misspecified as $Z$ with 60% censoring presented in the data