

Survival Analysis
with
Multivariate adaptive
Regression Splines

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig-Maximilians-Universität München

zur Erlangung des Grades Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Monika Kriner

11. Mai 2007

Tag des Rigorosums: 12. Juli 2007

1. Gutachter: Prof. Dr. Kurt Ulm
2. Gutachter: Prof. Dr. Ludwig Fahrmeir
3. Prüfer: Prof. Dr. Thomas Augustin
4. Prüfer: Prof. Dr. Helmut Küchenhoff

Dankeschön

Diese Dissertation entstand im Rahmen meiner Arbeit am Institut für Medizinische Statistik und Epidemiologie am Klinikum rechts der Isar der Technischen Universität München. Finanziell gefördert wurde meine Arbeit durch die Deutsche Forschungsgesellschaft im Rahmen des Sonderforschungsbereichs 386 (Analyse diskreter Strukturen).

Mein besonderer Dank gilt Prof. Dr. Ludwig Fahrmeir und Prof. Dr. Kurt Ulm, die mir eine optimale Förderung zukommen ließen. Sie hatten stets ein offenes Ohr für die kleinen und größeren Nöte, Sorgen und Probleme, die während der Erstellung der Arbeit aufgetreten sind. Immer wenn ich das Gefühl hatte, dass es kein Fortkommen mehr gab, war Prof. Ulm zur Stelle und half mir mit viel Ausdauer, Freundlichkeit und der nötigen Ironie dabei, wieder irgendwo ein Schlupfloch zu finden, durch das es weitergehen konnte.

Ein weiterer Dank geht an meinen Kollegen Tibor Schuster, der mir immer mit Rat und Tat zur Seite gestanden hat, wenn ich mit R oder LaTeX meine liebe Not hatte. Oft hat mir schon geholfen, ihm mein Problem zu schildern, schon kam mir die Lösung von ganz allein.

Unser Administrator Andreas Enterrottacher hat stets dafür gesorgt, dass mein Computer reibungslos funktioniert und war immer da, wenn es gebrannt hat. Stundenlang saßen wir zusammen und haben technische Probleme mit viel Energie und auch Spaß gelöst. Dafür möchte ich ihm danken.

Den Kardiologen Prof. Georg Schmidt und Dr. Axel Bauer danke ich, dass sie mir die Herzdaten zur Verfügung gestellt haben, die in Kapitel 5 analysiert wurden.

Meinen Eltern Maria und Johann Seebauer danke ich für ihre immerwährende Unterstützung. Ohne sie wäre ich heute sicherlich nicht hier.

Und nicht zuletzt gilt großer Dank meinem Mann Fabian, der immer an mich geglaubt und mir dabei geholfen hat, nie den Boden unter den Füßen zu verlieren.

Allen meinen lieben Freunden hier auch ein herzliches Dankeschön, denn Freunde machen das Leben reicher und lebenswerter.

Monika Kriner

Zusammenfassung

Multivariate adaptive Regressions-Splines (MARS) sind eine geeignete Methode zur Identifizierung von linearen und nichtlinearen Effekten und Interaktionen zwischen Kovariablen.

In der vorliegenden Arbeit wird ein neuer Ansatz zur Modellierung von Überlebenszeitdaten mit Hilfe von MARS vorgestellt. Martingal- und Devianzresiduen eines Cox-PH-Modells werden als Response in einem herkömmlichen MARS-Modell verwendet. Damit lassen sich sowohl funktionale Formen der Kovariableninflüsse als auch mögliche Interaktionen datengesteuert modellieren.

Simulationsstudien zeigen, dass die neue Methode eine bessere Anpassung an die Daten liefert als der traditionelle Cox-PH-Ansatz.

Die Analyse reeller Daten des Deutschen Herzzentrums bestätigt ebenfalls die Güte des neuen Verfahrens.

Abstract

Multivariate adaptive regression splines (MARS) are a useful tool to identify linear and nonlinear effects and interactions between two covariates.

In this dissertation a new proposal to model survival type data with MARS is introduced. Martingale and deviance residuals of a Cox PH model are used as response in a common MARS approach to model functional forms of covariate effects as well as possible interactions in a data-driven way.

Simulation studies prove that the new method yields a better fit to the data than the traditional Cox PH approach.

The analysis of real data of the German Heart Center on survivors of an acute myocardial infarction also documents the good performance of the method.

Contents

Introduction	ix
1 Survival time data and the counting process theory	1
1.1 Basics	2
1.2 The hazard function and its estimation	4
1.3 The Cox PH model	6
1.4 Martingale basics	9
2 Multivariate adaptive Regression Splines	11
2.1 The approach	12
2.2 Goodness-of-Fit criterion	15
2.3 Candidates and Knots	15
3 MARS in Survival time data	20
3.1 Basic assumptions	20
3.2 Enhancements	27
4 Simulation studies	36

4.1	Monograph of the approach	39
4.2	Interaction studies	50
4.3	Simulations on survival time data	53
4.3.1	Simulation studies with exponential distribu- tion	55
4.3.2	Simulation studies with Weibull distribution .	65
5	Application on real data: prediction of mortality af- ter myocardial infarction	76
5.1	Data preparation	79
5.2	The classical Cox PH method	80
5.3	A "snapshot"	82
5.4	Stepwise analysis	88
5.5	Final Results	94
6	Conclusions	104
6.1	Summary	104
6.2	Outlook	108
A	Notation	110
B	R functions	112
B.1	Simulations on a continuous response	112
B.2	Interaction of a binary and a continuous covariate . .	122
B.3	Survival MARS simulations	126

B.4 Function for Survival MARS 135

List of Figures

2.1	MARS approximation of a cubic relationship, small error variance of $\sigma^2 = 0.25$	19
2.2	MARS approximation of a cubic relationship, large error variance of $\sigma^2 = 4$	19
3.1	Example from Klein & Moeschberger with martingale residuals	26
3.2	Example from Klein & Moeschberger with Deviance residuals	27
4.1	Scenario A, 500 observations, $\sigma^2 = 0.1$, $d = 3$	46
4.2	Scenario B, 1000 observations, $\sigma^2 = 1$, $d = 4$	47
4.3	Scenario A, 1000 observations, $\sigma^2 = 2$, $d = 5$	48
4.4	Scenario C, 1000 observations, $\sigma^2 = 1$, $d = 4$	49
4.5	MARS results for an interaction between a binary and a continuous covariate	52
4.6	Martingale residuals, $d = 2$, 25% censoring rate	59
4.7	Deviance residuals, $d = 3$, 50% censoring rate	60

4.8	Martingale residuals, $d = 4$, 50% censoring rate . . .	61
4.9	Deviance residuals, $d = 5$, 85% censoring rate	62
4.10	Comparison of the variances of MARS with martingale residuals as response and the classical Cox approach with exponential distributed survival times and a censoring rate of 50% ($d = 4$)	65
4.11	Comparison of the variances of MARS with deviance residuals as response and the classical Cox approach with exponential distributed survival times and a censoring rate of 50% ($d = 4$)	66
4.12	Density functions of the Weibull distribution	67
4.13	Martingale and deviance residuals for several parameters	68
4.14	Comparison of the variances of MARS with martingale residuals as response and the classical Cox approach with Weibull distributed survival times ($v = 2, \lambda = 5$) and a censoring rate of 50% ($d = 4$)	71
4.15	Comparison of the variances of MARS with deviance residuals as response and the classical Cox approach with Weibull distributed survival times ($v = 2, \lambda = 5$) and a censoring rate of 50% ($d = 4$)	72
5.1	Histograms of deceleration capacity for all data and trimmed by extreme values of DC	79

5.2	Basis functions of AGE and SDNN in the martingale model	86
5.3	Basis functions of AGE and SDNN in the deviance model	87
5.4	Basis function of AC in the deviance model	87
5.5	Basis functions of the stepwise procedures for the martingale and deviance scenario	90
5.6	Basis functions of the nonlinear covariates of the multivariate models after the stepwise procedures (the two scenarios differ in the bottom panels: the martingale model included SDNN, the deviance model included LVEF)	93
5.7	Kaplan-Meier curves of all five final models	97
5.8	Comparison of the snapshot and stepwise models with the classical Cox model via martingale residuals	98

List of Tables

4.1	Results of the simulations: fraction of simulation cycles with spuriously chosen basis functions, correct identified interaction terms and correct identified nonlinear influence of x_{11} in scenario C	42
4.2	Results of the simulations with an interaction between a continuous and a binary covariate	51
4.3	Formulae for generating survival times and the hazard function	55
4.4	Results of the simulation studies on exponential distributed survival time data in percent	58
4.5	Results of the simulation studies on Weibull distributed survival time data ($v = 2$) in percent	74
4.6	Results of the simulation studies on Weibull distributed survival time data ($v = 5$) in percent	75
5.1	Main characteristics of the patients after myocardial infarction	80

5.2	Cox PH model of the infarction data	81
5.3	The final Cox PH model of the infarction data	81
5.4	Multivariate MARS model with martingale residuals of the Cox null model as response	83
5.5	Multivariate MARS model with deviance residuals of the Cox null model as response	84
5.6	Stepwise analysis with martingale residuals as response	89
5.7	Stepwise analysis with deviance residuals as response	89
5.8	Multivariate MARS model with the covariates found in the stepwise procedure with martingale residuals of the Cox null model as response	91
5.9	Multivariate MARS model with the covariates found in the stepwise procedure with deviance residuals of the Cox null model as response	92
5.10	Coefficients and standard error of the 5 Cox models (conventional Cox and Cox after MARS models) . . .	96
5.11	Log-likelihoods of the four final models and the con- ventional Cox approach	97
5.12	Variances of the martingale residuals of the five final models	99
5.13	Coefficients and standard error of the 3 MARS mod- els on the restricted data	101

Introduction

The methodology of regression analysis presents a wide fund of possibilities to model the influence of some prognostic factors on a certain dependent covariate. One is usually interested in the strength of the influence as well as on the form, whether linear, nonlinear, threshold type or as interaction between two of the covariates.

Linear models are widespread but a very bounded approach of describing the association between different variables and the response. Enhancements of this model type are generalized linear models [19] and generalized additive models [13].

The application of Fractional Polynomials [23], which offers a selection of polynomials as possible type of functional form, shows an improvement compared to linear models as well. However the analysis of interactions is restricted to interactions between a binary and a continuous covariate [24].

Especially the locating of interactions between certain covariates always raises problems, for the user has to find existing interactions himself by trying all possibilities manually.

In this thesis I would like to present another kind of modelling approach - the multivariate adaptive regression splines (MARS). Piecewise linear splines with alternating slope enable to model linear as well as nonlinear influences of the covariates and particularly interactions, i.e. MARS is a type of segmented regression approach. The combination of several so called basis functions, which are determined by data-driven knots, yields a flexible and practicable setting. The model building process runs successively, starting with a null model and including the best basis function per iteration, concerning a certain fitting criterion. All possible model expansions, i.e. further basis functions, are calculated per iteration step and the basis function, which minimizes the fitting criterion is included in the enhanced model. These basis functions can imply single covariates, so-called hockeystick functions, i.e. functions including a knot, and interactions between two different variables.

This approach, which has been implemented in the context of linear and logistic regressions, was extended for the use of survival type data in this thesis, using martingale [3] and deviance residuals [26] of a Cox model [8] as response to MARS. These two types of residuals were compared regarding their power within this new approach.

In the first chapter the theory of survival time analysis and especially Cox regression is introduced in the framework of counting

process theory and martingale basics are brought in.

In chapter 2 multivariate adaptive regression splines (MARS) are presented in detail.

In chapter 3 the new theory of MARS used in the context of survival time data is presented. Martingale and deviance residuals of a Cox model without any covariates are used as response variables in a MARS model, respectively, to detect influential covariates together with their functional form.

Chapter 4 introduces simulation studies about MARS and its abilities. The aim of the studies is to check whether MARS actually is able to identify the appropriate model, particularly in the situation of an interaction. Furthermore simulation studies on the new approach with martingale and deviance residuals of a Cox model as response in a MARS setting are described. For that purpose exponential and Weibull distributed survival times were analysed.

In chapter 5 data on patients who suffered a myocardial infarction are presented and evaluated by using the newly introduced approach of survival MARS. A snapshot analysis (including all possible covariates simultaneously) and also a stepwise procedure are executed for both martingale and deviance residual scenarios.

The last chapter summarizes the conclusions of this thesis and the outlook on further work on the topic.

Chapter 1

Survival time data and the counting process theory

Many medical investigations are engaged with the incidence of a certain event, e.g. relapse after healing or death after a chemotherapy. Besides one is often interested in the length of time until this event occurs. Consequently the investigator has to deal with two dependent covariates, and the most common modelling approach for that situation is survival time analysis, more precisely the Cox proportional hazards model [8]. This approach is introduced in the following chapter in the context of counting process theory, which was introduced in the 1970's by Odd O. Aalen [1].

1.1 Basics

The data

Let n individuals be observed in a study and let T_i^* be the time from beginning of observation until the occurrence of the event for the i th subject. Typically this value cannot be observed for all patients, because of restrictions to the observation time. Some subjects may still be under risk at the end of the study, so one has just observed the censoring time C_i , which is defined as the time from the beginning to the end of the study. The basic assumption in this context is the independence of T_i^* and C_i . Now let $T_i = \min(T_i^*, C_i)$ be the observed followup time and $\delta_i = I(T_i^* \leq C_i)$ the status, which indicates, whether a subject had the event or not.

Suppose that the T_i 's are iid with distribution function $F(t) = Pr(T \leq t)$ and density function $f(t) = \frac{dF(t)}{dt}$. The survivor function $S(t)$, i.e. the probability to survive at least until time t , is

$$S(t) = Pr(T > t).$$

The counting process notation of survival time data substitutes the

pair (T_i, δ_i) by the pair of stochastic processes $(N_i(t), Y_i(t))$ with

$N_i(t)$ = the number of observed events in $[0, t]$ for subject i and

$$Y_i(t) = \begin{cases} 1 & , \text{ if subject } i \text{ is under risk at time } t. \\ 0 & , \text{ else} \end{cases}$$

This approach can handle multiple events and multiple time intervals under observation. However, in this thesis only the case with $N_i(t) \in \{0, 1\}$ ought to be discussed.

For counting processes are a special type of stochastic processes, it is reasonable to briefly introduce the theory and terminology of this kind of modelling technique here.

Stochastic processes

Data and the occurrence of events over a specific period of time $T = [0, t]$ or $T = [0, t)$, $0 < t \leq \infty$, are often modelled with stochastic processes.

A stochastic process X is a collection of random variables

$$(X(t) : t \in T)$$

i.e. every value is observed at a defined time-point.

Thus, at time t all data are available which were observed until t .

This information, which was generated by the process X on $[0, t]$, is

represented by the filtration or history $(\mathcal{F}_t : t \in T)$ and grows over time, i.e. $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$.

A counting process is a stochastic process $(N(t) : t \in T)$ with $N(0) = 0$ and $N(t) < \infty$, whose paths are right-continuous with probability 1, piecewise constant, and have jumps of height +1 [11]. The most frequent application of counting processes is the determination of the number of events of a certain type within a specific time-interval: $N(t) - N(s)$ would denote the number of observed events in $[s, t]$.

1.2 The hazard function and its estimation

The hazard function is defined as

$$\begin{aligned} \lambda(t) &= \lim_{h \downarrow 0} \frac{\Pr(t < T \leq t + h | T > t)}{h} \\ &= \frac{f(t)}{S(t)}. \end{aligned} \quad (1.1)$$

The estimator of the cumulative hazard function $\Lambda(t) = \int_0^t \lambda(s) ds$ is based on the summed processes $\bar{Y}(t) = \sum_i Y_i(t)$ and $\bar{N}(t) = \sum_i N_i(t)$. $\bar{Y}(t)$ depicts the number of subjects who are at risk at time t and $\bar{N}(t)$ depicts the total number of events up to and including t .

For a short period of time it holds that

$$\begin{aligned}\Lambda(s+h) - \Lambda(s) &\approx \lambda(s)h & (1.2) \\ &= Pr(\text{event in}(s, s+h)|\text{at risk at } s)\end{aligned}$$

This expression can be estimated by $\frac{\bar{N}(s+h) - \bar{N}(s)}{\bar{Y}(s)}$. We obtain the Nelson-Aalen estimator by summing these quantities over subintervals of $(0, t]$, which are small enough to contain at most one event time:

$$\hat{\Lambda}(t) = \int_0^t \frac{d\bar{N}(s)}{\bar{Y}(s)} \quad (1.3)$$

where $d\bar{N}(t)$ is the number of events occurring exactly at t .

The variance of the Nelson-Aalen estimator is assessed consistently by

$$var [\hat{\Lambda}(t)] = \int_0^t \frac{d\bar{N}(s)}{[\bar{Y}(s)]^2} \quad (1.4)$$

which is just the sum of the scaled increments $\frac{\Delta_h \bar{N}(s)}{\bar{Y}(s)}$ with $\Delta_h \bar{N}(t) = \bar{N}(t+h) - \bar{N}(t)$, the number of events in a small interval from t to $t+h$.

The Nelson-Aalen estimator has an interesting property:

$$\sum_{i=1}^n \hat{\Lambda}(T_i) = \sum_{i=1}^n N_i(T_i) \quad (1.5)$$

i.e. the total estimated hazard, summed over all subjects, equals the total number of observed events. Since $\hat{\Lambda}(T_i) = \int_0^\infty Y_i(s) d\hat{\Lambda}(s)$ and $N_i(T_i) = N_i(\infty)$ the latter property can be written as

$$\sum_i \left[N_i(\infty) - \int_0^\infty Y_i(s) \hat{\Lambda}(s) \right] = 0, \quad (1.6)$$

which is of high relevance and will be described in greater detail in section 4.

To establish a model in terms of the history or filtration, let $dN_i(t)$ be the increment in N_i in an arbitrary small time interval $[t, t + dt]$ and consider the fact, that \mathcal{F}_{t-} contains all information of $[0, t)$. Then it holds that

$$\mathcal{E}(dN_i(t) | \mathcal{F}_{t-}) = Y_i(t) \lambda(t) dt. \quad (1.7)$$

1.3 The Cox PH model

Beside the analysis of time until event, one is often interested in the influence of additional variables on the survival time and the outcome of the study, i.e. does the subject experience the interesting event or not under certain conditions determined by some covariates, e.g. prognostic factors.

Since its introduction the classical Cox proportional hazards model [8] is most often used to estimate the effect of one or more covariates

on the survival of a certain subject. Let $X = (X_1, \dots, X_p)$ denote p measured covariates on a given subject. Then the corresponding hazard function is defined as

$$\lambda(t) = \lambda_0(t)e^{f(X)} \quad (1.8)$$

with an unspecified nonnegative baseline hazard $\lambda_0(t)$ as a function of time. The baseline hazard represents the hazard of an "average" subject of the data, thus the hazard of a particular subject is just a multiple of the baseline hazard. In the Cox model the most important assumption is the proportionality of the hazard rates:

$$\frac{\lambda(t|X)}{\lambda(t|X=0)} = \frac{\lambda_0(t)e^{f(X)}}{\lambda_0(t)} = e^{f(X)}$$

This means, that the hazard rate of a particular subject remains the same over the whole observation time, i.e. the influence of all covariates is independent of time.

Inference

As the baseline hazard is not estimated by parametric methods, likelihood inference cannot be performed for the Cox model. The coefficients β can though be estimated by maximizing a "partial likelihood" function, even when λ_0 is left completely unspecified.

The partial likelihood for β [8] is defined as

$$L(\beta) = \prod_{i=1}^n \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\beta' X_i)}{\sum_{j \in R_k} Y_j(t) \exp(\beta' X_j)} \right\}^{dN_i(t)}, \quad (1.9)$$

with R_k being the set of subjects at risk at time t .

The log partial likelihood is accordingly given by

$$\log L(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[Y_i(t) \beta' X_i - \log \left(\sum_{j \in R_k} Y_j(t) \exp(\beta' X_j) \right) \right] dN_i(t). \quad (1.10)$$

Maximum partial likelihood estimates $\hat{\beta}$ are found by solving the equation $U(\beta) = \frac{\partial}{\partial \beta} \log L(\beta) = 0$ with

$$U(\beta) = \sum_{i=1}^n \int_0^{\infty} \left[X_i - \frac{\sum_{j \in R_k} Y_j X_j \exp(\beta' X_j)}{\sum_{j \in R_k} Y_j \exp(\beta' X_j)} \right] dN_i(t). \quad (1.11)$$

Breslow [6, 7] proposed an estimate for the cumulative baseline hazard in the case with covariates. Let $\hat{\beta}$ be partial likelihood estimations of the coefficients. Then the i th subject in the sample has an estimated event rate of $Y_i(s) \exp\{\hat{\beta}' X_i(s)\} \lambda_0(s)$. A subject in the risk set with rate $\exp\{\hat{\beta}' X_i(s)\} \lambda(s)$ will have an event with same probability as $\exp\{\hat{\beta}' X_i(s)\}$ cases with rate $\lambda_0(s)$. A heterogenous sample could be visualized like that:

$\sum_{i=1}^n Y_i(s) \exp\{\hat{\beta}' X_i(s)\}$ cases under risk at time s all have an event probability of $\lambda_0(s)$. Thus $\Lambda_0(t)$ can be estimated as follows:

$$\hat{\Lambda}_0(t) = \int_0^t \frac{\sum_i dN_i(t)}{\sum_i Y_i(s) \exp\{\hat{\beta}' X_i(s)\}} \quad (1.12)$$

1.4 Martingale basics

The expression introduced in (1.6),

$$M_i(t) = N_i(t) - \int_0^t Y_i(s) \lambda(s) ds, \quad (1.13)$$

is a martingale process with respect to the history defined above. A martingale concerning the filtration \mathcal{F}_t is a stochastic process with right-continuous paths and left-sided limits and has the key property of any martingale: for every $t > 0$ it is

$$\mathcal{E}(dM(t)|\mathcal{F}_{t-}) = 0.$$

Thus $\mathcal{E}(M(t)|\mathcal{F}_s) = M(s)$ for every $0 \leq s < t$, i.e. $\mathcal{E}(M(t)|M(u); 0 \leq u \leq s) = M(s)$. That is, a martingale is a process without drift, which means that the best prediction of any future value is its actual value when it is conditioned on the

past.

Martingale increments are uncorrelated, but not necessarily independent. For $t, u, s > 0$ holds:

$$\text{Cov}[M(t), M(t+u) - M(t)] = 0$$

$$\text{Cov}[M(t) - M(t-s), M(t+u) - M(t)] = 0$$

The most simple martingale is a symmetric random walk in discrete time.

The Doob-Meyer decomposition theorem [10, 20] indicates, that any counting process is uniquely decomposable into a sum of a martingale and a so-called compensator, which is a predictable right-continuous process (0 at time 0). Since $N(t)$ is a counting process of the occurred events in $[0, t]$ and $\int_0^t Y_i(s)\lambda(s)ds$ in (1.13) is a right-continuous and predictable process,

$$\begin{aligned} M_i(t) &= N_i(t) - \int_0^t Y_i(s)\lambda(s)ds \\ &\equiv N_i(t) - \Lambda_i(t) \end{aligned}$$

is in fact a martingale with the above properties.

Chapter 2

Multivariate adaptive Regression Splines

The theory of multivariate adaptive regression splines (MARS) was developed by Jerome Friedman [12] in 1991. Charles Kooperberg [16] designed a comparable approach especially for categorical covariates with a short enhancement for continuous responses.

Both proposals are based on tree structures, i.e. display an advancement of CART (classification and regression trees).

As open source software on MARS is so far available only from Kooperberg (function *polymars* in the library *polspline* on R 2.3.0), his approach was used in this thesis.

The intention of this dissertation was to establish a MARS approach to deal with survival time data, thus censored observations.

The general approach of multivariate adaptive regression splines will

now be presented in this chapter. The extension on survival time data is then introduced in chapter 3.

2.1 The approach

Let y be the dependent response, which can be continuous or binary, and let $\mathbf{X} = (X_1, \dots, X_p) \in \mathcal{D} \subset \mathbb{R}^p$ be the set of potential predictive covariates. Then we assume that the data are generated from an unknown "true" model. In case of a continuous response this would be

$$\begin{aligned} y &= f(X_1, \dots, X_p) + \epsilon \\ &= f(\mathbf{X}) + \epsilon, \end{aligned} \quad (2.1)$$

in case of a binary response, i.e. using a logistic regression, this would be

$$\text{logitPr}(Y = 1) = f(X_1, \dots, X_p) + \epsilon. \quad (2.2)$$

The distribution of the error ϵ is member of the exponential family [19]. MARS approximates f by applying functions, which include interactions of at most second order. That means we use the model

$$f(\mathbf{X}) = g_0 + \sum_{j_1} g_{j_1}(X_{j_1}) + \sum_{j_1 < j_2} g_{j_1, j_2}(X_{j_1}, X_{j_2}) + \epsilon. \quad (2.3)$$

Linear splines and their tensor products are used to model the function $g(\cdot)$. A one-dimensional spline can be written as

$$g(x) = b_{-1} + b_0x + \sum_{k=1}^K b_k(x - t_k)_+, \quad (2.4)$$

with the so-called hockeystick-function

$$(x - t_k)_+ = \begin{cases} x - t_k & , \text{ if } x \geq t_k, \\ 0 & , \text{ else,} \end{cases} \quad (2.5)$$

and the knot t_k in the range of the observed values of \mathbf{X} . For this reason the function g is situated in a linear space with the $K + 2$ basis functions

$$1, x \text{ and } (x - t_k)_+, \quad k = 1, \dots, K.$$

The interaction g_{j_1, j_2} is modeled by means of tensor product splines

$$g_{12}(x_1, x_2) = g_1(x_1) \times g_2(x_2). \quad (2.6)$$

Thus the following model results:

$$\begin{aligned} g_0 &= \beta_0 \\ g_{j_1}(X_{j_1}) &= \sum_{i=1}^M \beta_i^{j_1} B_i^{j_1}(X_{j_1}), \\ g_{j_1, j_2}(X_{j_1}, X_{j_2}) &= \sum_{i=1}^M \beta_i^{j_1, j_2} B_i^{j_1, j_2}(X_{j_1}, X_{j_2}), \end{aligned}$$

with M being the number of basis functions in the model. The B s represent spline basis functions as described above and the β s are coefficients. Now the MARS model can be written as

$$f(\mathbf{X}) = \sum_{i=1}^M \beta_i B_i(\mathbf{X}), \quad (2.7)$$

with the following possible modes of basis functions $B(\mathbf{X})$:

- 1
- x_i
- $(x_i - t_k)_+$
- $x_i x_j$
- $(x_i - t_k)_+ x_j$
- $(x_i - t_k)_+ (x_j - t_l)_+$

In the MARS approach the coefficients are estimated using the Least Squares method with the coefficient matrix written as

$$\hat{\beta} = (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} Y. \quad (2.8)$$

\mathbf{X}^* is the design matrix, i.e. the matrix of the selected basis functions, and Y is the response vector.

2.2 Goodness-of-Fit criterion

As measure for the degree of fit or lack of accuracy of the model the generalized cross-validation criterion is used:

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N [y_i - \hat{f}_M(\mathbf{x}_i)]^2}{\left[1 - \frac{d \cdot M}{N}\right]^2} \quad (2.9)$$

with M being the number of basis functions in the model, whereas \hat{f} denotes the fitted values of the current MARS model.

The numerator is the common residual sum of squares, which is penalized by the denominator, which accounts for the increasing variance in the case of increasing model complexity. The penalizing parameter d can be chosen by the user himself, a conventional value is $d = 4$. A smaller d generates a larger model with more basis functions, a larger d creates a smaller model with less basis functions.

2.3 Candidates and Knots

MARS is a stepwise process. In the stepwise addition process basis functions are added until the maximal allowed model size (which can be assessed by the user) is reached. The largest model generally overfits the data. Then in a second stage – the stepwise deletion

process – all "unnecessary" basis functions are removed again until a final model is obtained which is best considering the GCV, i.e. the one with the minimum GCV.

In the first step of the addition process a constant model is fitted. Subsequently the number of candidate basis functions depends on the number of possible knots per predictor variable. To keep the procedure fast and the results robust the number of possible knots per predictor and also the possible candidates per step are limited. To determine the number of potential knots of a specific covariate an order statistic is computed and a subset of it is then chosen as potential knots. Commonly these are about 20 knots per predictor, at most every third value is chosen yet. This prevents the design matrix of the basis functions to become singular. However, the user can adjust the number of candidate knots himself.

The following basis functions are possible candidates:

- x_i , $i = 1, \dots, p$
- $(x_i - t_{ik})_+$, if x_i is already in the model
- $x_i x_j$, if x_i **and** x_j are already in the model
- $x_i(x_j - t_{jk})_+$, if $x_i x_j$ **and** $(x_j - t_{jk})_+$ are already basis functions
- $(x_i - t_{ik})_+(x_j - t_{jk})_+$, if $x_i(x_j - t_{jk})_+$ **and** $(x_i - t_{ik})_+ x_j$ are in the model yet

These conditions force linear terms to be involved and result in a better interpretability of the final model.

In the first iteration – after the fit of the constant model – a linear basis function on one of the predictor variables is fitted. The second iteration accounts for both linear basis functions on another covariate and basis functions with knots of the covariate already in the model.

The model to choose in every step during the forward process is the one out of all possible models which minimises the GCV.

In the backward process one basis function is deleted per step and the GCV is computed for the reduced model. The model which yields the smallest increase of GCV becomes the new one.

The maximum model size is arbitrary. The default proposed by Kooperberg is $\min(6n^{1/3}, \frac{n}{4}, 100)$ with n being the number of cases in the data set. Unfortunately, Kooperberg did not describe, why he decided for these values.

Forward and backward process yield a sequence of models with different numbers of basis functions. The one out of this sequence of varying sized models which offers the minimum GCV is chosen as "best" and defined as final fit.

Example:

Let $y = 3x^3 + \epsilon$ with $x \sim U[-1, 1]$ and $\epsilon \sim N(0, \sigma^2)$ with $\sigma^2 \in \{0.25, 4\}$. The application of MARS on these data resulted in the following two models with 6 and 2 knots and GCVs of 0.0633 and 4.046405, respectively:

$$\begin{aligned} f_{\sigma^2=0.25}(x) &= 6.66 + 9.69x - 4.67(x + 0.90)_+ - 2.45(x + 0.64)_+ \\ &\quad - 2.38(x + 0.37)_+ + 1.88(x - 0.33)_+ + 2.19(x - 0.60)_+ \\ &\quad + 3.57(x - 0.81)_+ \end{aligned}$$

$$f_{\sigma^2=4}(x) = 1.94 + 4.55x - 4.58(x + 0.40)_+ + 4.33(x - 0.44)_+$$

The figures 2.1 and 2.2 show the scatterplots with the corresponding MARS regression curves (red line) of this example. The green lines display the true relationship $y = 3x^3$. The figures show, that the MARS approximation almost completely overlays the true function, thus yields a very good fit to the data, even in the case of a rather large error variance.

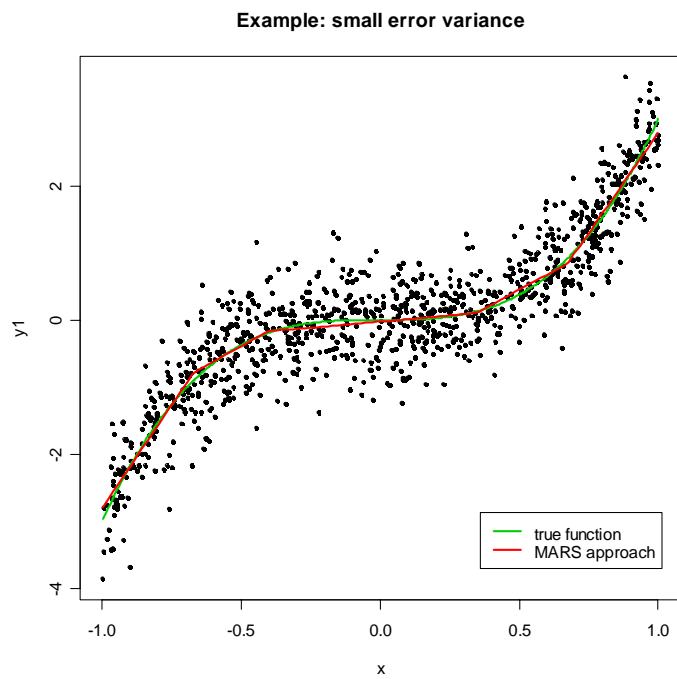


Figure 2.1: MARS approximation of a cubic relationship, small error variance of $\sigma^2 = 0.25$.

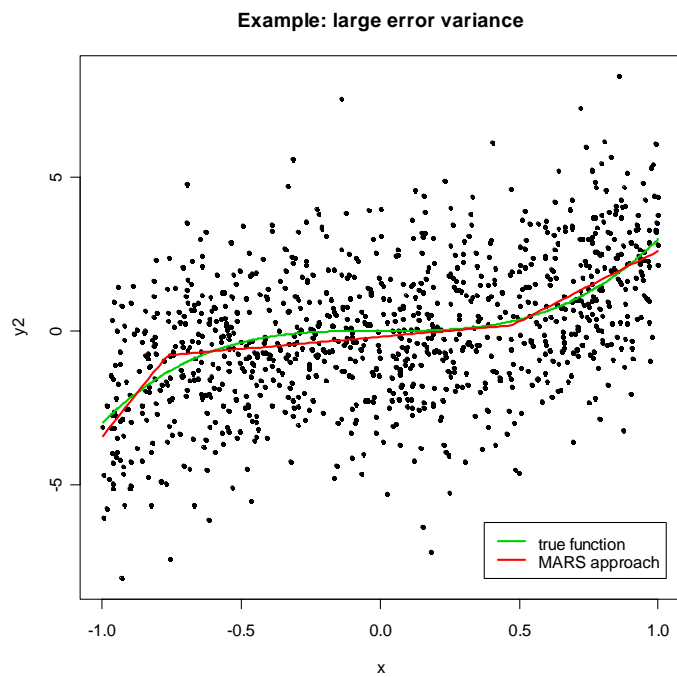


Figure 2.2: MARS approximation of a cubic relationship, large error variance of $\sigma^2 = 4$.

Chapter 3

MARS in Survival time data

3.1 Basic assumptions

Multivariate adaptive regression splines are so far implemented and used for continuous and binary responses. The intention of this thesis was to expand this technique to survival time data. For that purpose a new approach was developed, which uses martingale or deviance residuals of a Cox model as response in a common MARS approach [11, 14].

Remind that martingale residuals are based on the fact, that the process

$$\begin{aligned}
M_i(t) &= N_i(t) - \int_0^t Y_i(s) \exp\{\beta' X_i(s)\} d\hat{\Lambda}_0(s) & (3.1) \\
&\equiv N_i(t) - \int_0^t Y_i(s) \lambda(s) ds \\
&\equiv \delta_i - \hat{\Lambda}_i(t)
\end{aligned}$$

is a martingale, when the proportional hazards assumption holds. The process $N_i(t)$ counts the number of occurred events of individual i at time t , thus is identical to δ_i , because only $N_i(t) \in \{0, 1\}$ are regarded in this thesis.

The used quantities in (3.1) are defined as follows:

$\mathbf{Y}_i(\mathbf{t})$ is an indicator, if subject i is under risk immediately before t

$\mathbf{N}_i(\mathbf{t})$ is a counting process of the events

$\mathbf{X}_i(\mathbf{t})$ is the vector of covariates

$\hat{\Lambda}_0(\mathbf{t})$ is the Breslow estimator of the cumulative baseline hazard function

Assume that

$$\begin{aligned}
\lambda(t|X, Z)dt &= h(X) \exp\{\beta' Z\} d\Lambda_0(t) & (3.2) \\
&\equiv \exp\{f(X)\} \exp\{\beta' Z\} d\Lambda_0(t)
\end{aligned}$$

i.e. the correct functional form is only known for the covariate vector Z but not for the single non time-dependent covariate X .

According to Fleming & Harrington [11] it holds that

$$\begin{aligned} \mathcal{E}(\hat{M}(t)|X) &\approx \frac{1 - \bar{h}(t_0)}{h(X)} \mathcal{E}(N(t)|X) + \\ &+ \mathcal{E} \left\{ \int_0^t Y(s) \exp\{\beta' Z\} [\bar{h}(t_0) - \bar{h}(s, Z)] d\Lambda_0(s)|X \right\} \end{aligned} \quad (3.3)$$

with

$$\bar{h}(s, Z) = \frac{\mathcal{E}\{e^{f(X)} Y(s)|Z\}}{\mathcal{E}\{Y(s)|Z\}} = \frac{\mathcal{E}\{h(X) Y(s)|Z\}}{\mathcal{E}\{Y(s)|Z\}}$$

$$\bar{h}(s) = \mathcal{E}\bar{h}(s, Z)$$

and t_0 a fixed time point. The function $\bar{h}(s, Z)$ depicts the mean value of $h(X)$ over time as well as the expected distribution of the risk set at a given time point.

Inverting equation (3.3) yields

$$-\log \left\{ 1 - \frac{\mathcal{E}(\hat{M}(t)|X)}{\mathcal{E}(N(t)|X)} \right\} \approx \quad (3.4)$$

$$\approx [f(X) - \log\{\bar{h}(t_0)\}] -$$

$$-\log \left[1 - \frac{h(X)\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}[\bar{h}(t_0) - \bar{h}(s, Z)]d\Lambda_0(s)|X\right\}}{\bar{h}(t_0)\mathcal{E}\{N(t)|X\}} \right]$$

$$= f(X) - \log \left\{ \frac{\bar{h}(t_0)\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}h(X)d\Lambda_0(s)|X\right\}}{\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}h(X)d\Lambda_0(s)|X\right\}} - \frac{h(X)\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}[\bar{h}(t_0) - \bar{h}(s, Z)]d\Lambda_0(s)|X\right\}}{\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}h(X)d\Lambda_0(s)|X\right\}} \right\}$$

$$\Rightarrow -\log \left\{ 1 - \frac{\mathcal{E}(\hat{M}(t)|X)}{\mathcal{E}(N(t)|X)} \right\} \approx \quad (3.5)$$

$$\approx f(X) - \log \left[\frac{\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}\bar{h}(s, Z)h(X)d\Lambda_0(s)|X\right\}}{\mathcal{E}\left\{\int_0^t Y(s)e^{\beta'Z}h(X)d\Lambda_0(s)|X\right\}} \right]$$

$$= f(X) - \log \tilde{h}$$

$$\equiv f(X) - \tilde{f}$$

whereas the function \tilde{h} ...

”... depends in a complex way upon both the censoring pattern and the true functional form and will in general not be constant in X . When X is independent of Z , however, $\bar{h}(s, Z)e^{\beta'Z}$ will be independent of X , and smoothing provided by the integral will cause \tilde{f} to have small variation compared to f .”

(Fleming & Harrington)

Taylor series approximation results in the following relationship:

$$\mathcal{E}(\hat{M}|X) \approx [f(X) - \tilde{f}] \cdot c \quad (3.6)$$

with c being the number of events divided by the sample size.

Thus, a smoothed plot of the \hat{M}_i against the particular covariate X should reveal the correct functional form $f(X)$.

As an example assume we have a model with p covariates already included. Now we plot the martingale residuals of this model against the next covariate X and perform a MARS fit. The form of the fit is an indicator of the form of $f(X)$ (linear, nonlinear, threshold etc.).

Klein & Moeschberger gave an example of bone marrow trans-

plant data. They calculated the martingale residuals of a Cox model without the interesting covariate *wtime* (waiting time until transplantation in months). Figure 3.1 shows the plot of these residuals versus the covariate *wtime*, the Lowess smooth, which is the proposal of Klein & Moeschberger, and the MARS approach. Apparently the functional form which was found by MARS exceedingly coincides with the proposed Lowess smooth. The found MARS model has the following form:

$$\begin{aligned} \hat{f}_{mart}(wtime) = & - 0.31 + 0.02 \cdot wtime - 0.03 \cdot (wtime - 34)_+ \\ & - 0.04 \cdot (wtime - 71)_+ + 0.06 \cdot (wtime - 98)_+ \end{aligned}$$

Martingale residuals are defined on $(-\infty, 1]$ and thus highly skewed. To determine whether this skewness does in any way affect the outcome, the new method was also applied to a transformation of the martingale residuals, namely deviance residuals, which are symmetrically defined on $(-\infty, \infty)$. If all covariates are time-fixed, the deviance residual for subject i is defined as

$$d_i = \text{sign}(\hat{M}_i) \cdot \sqrt{-\hat{M}_i - N_i \log \left(\frac{N_i - \hat{M}_i}{N_i} \right)}. \quad (3.7)$$

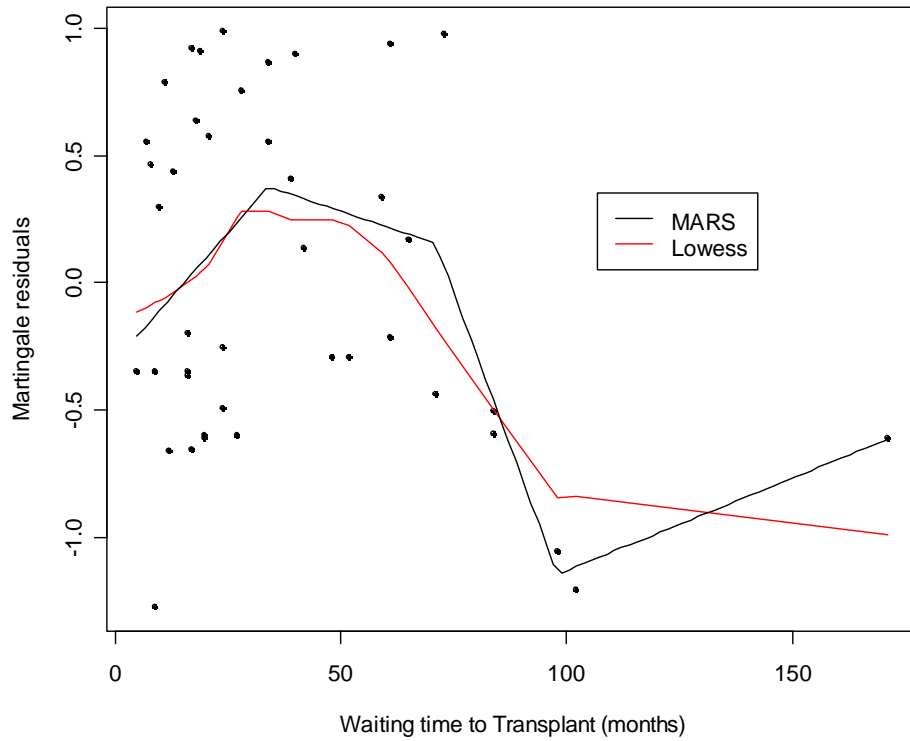


Figure 3.1: Example from Klein & Moeschberger with martingale residuals

The example from Klein & Moeschberger with deviance residuals is shown in Figure 3.2. Note that the MARS approach found the same knots for both martingale and deviance residuals as response.

The model with deviance residuals looks as follows:

$$\begin{aligned} \hat{f}_{devi}(wtime) = & - 0.39 + 0.03 \cdot wtime - 0.03 \cdot (wtime - 34)_+ \\ & - 0.06 \cdot (wtime - 71)_+ + 0.07 \cdot (wtime - 98)_+ \end{aligned}$$

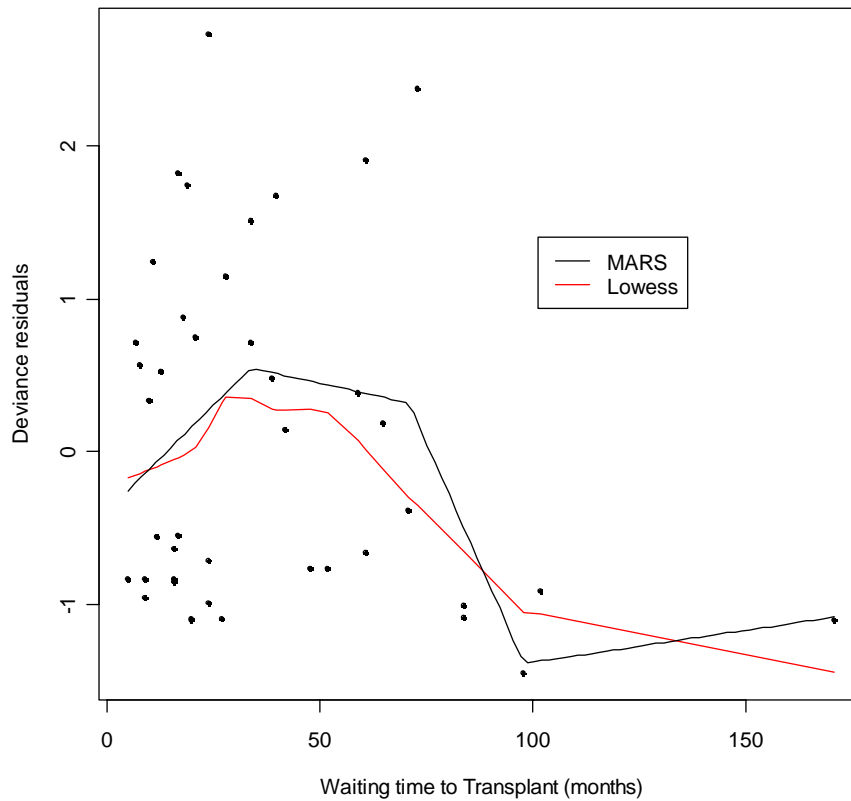


Figure 3.2: Example from Klein & Moeschberger with Deviance residuals

3.2 Enhancements

One intention of this thesis was to extend the proof of Fleming & Harrington [11] to the situation of interactions, i.e. it shall be shown that a smoothed plot of martingale residuals of a model without the two covariates A and B against A and B (3D-plot) is also able to reveal the correct relationship and influence of A and B on survival time.

This would mean, that the method can detect influential covariates and their correct functional form simultaneously at one step, thus

the usage of residuals of a Cox model without any covariates as dependent variable in a multivariate MARS setting is an adequate approach.

Let $f(A, B)$ be an arbitrary function of the two non time-dependent covariates A and B , whereas $f(A, B)$ can receive the following forms:

- $f(A, B) = f(A) + f(B)$
- $f(A, B) = f(A) \cdot f(B)$

Assume that

$$\begin{aligned} \lambda(t|A, B, Z) &= e^{f(A,B)} e^{\beta'Z} d\Lambda_0(t) & (3.8) \\ &\equiv h(A, B) e^{\beta'Z} d\Lambda_0(t) \end{aligned}$$

describes a model where the functional form of the covariate vector Z is known, but the function h of the variables A and B is unknown. $\hat{M}(t)$ again denotes the martingale residuals of a model which does not include A and B .

Thus it holds that the expected martingale residuals can be written as follows:

$$\begin{aligned}\mathcal{E}\{\hat{M}|A, B\} &= \mathcal{E}\{N(t)|A, B\} + \mathcal{E}\{\Lambda(t)|A, B\} \\ &\equiv \mathcal{E}\{N(t)|A, B\} + \mathcal{E}\left\{-\int_0^t Y(s)e^{\hat{\beta}'Z}d\hat{\Lambda}_0(s)|A, B\right\}\end{aligned}$$

This expression is then dilated by the term $e^{\beta'Z}\bar{h}(s, Z)d\Lambda_0(s)$, whereas \bar{h} represents an average of the function h over the expected event time of a subject with covariate matrix A and B as well as over the expected risk set at time t :

$$\begin{aligned}\mathcal{E}\{\hat{M}|A, B\} &= \mathcal{E}\{N(t)|A, B\} + \mathcal{E}\left\{-\int_0^t Y(s)e^{\beta'Z}\bar{h}(s, Z)d\Lambda_0(s)|A, B\right\} + \\ &\quad + \mathcal{E}\left\{\int_0^t Y(s)\left[e^{\beta'Z}\bar{h}(s, Z)d\Lambda_0(s) - e^{\hat{\beta}'Z}d\hat{\Lambda}_0(s)\right]|A, B\right\} \\ &= \text{term1} + \text{term2} + \text{term3}\end{aligned}$$

with $\bar{h}(s, Z) = \frac{\mathcal{E}\{h(X)Y(s)|Z\}}{\mathcal{E}\{Y(s)|Z\}}$ and $\bar{h}(s) = \mathcal{E}\bar{h}(s, Z)$.

Term 2 and term 3 are now regarded separately.

Start with term 3:

$$\begin{aligned}
term3 &= \mathcal{E} \left\{ \int_0^t Y(s) \left[e^{\beta'Z} \bar{h}(s, Z) d\Lambda_0(s) - e^{\hat{\beta}'Z} d\hat{\Lambda}_0(s) \right] | A, B \right\} \\
&= \mathcal{E} \left\{ \int_0^t Y(s) \left[e^{\beta'Z} \bar{h}(s, Z) d\Lambda_0(s) - e^{\hat{\beta}'Z} \frac{d\bar{N}(s)}{\sum_{j=1}^n Y_j(s) e^{\hat{\beta}'Z_j}} \right] | A, B \right\} \\
&= - \mathcal{E} \left\{ \sum_{i=1}^n \int_0^t \frac{Y(s) e^{\hat{\beta}'Z}}{\sum_{j=1}^n Y_j(s) e^{\hat{\beta}'Z_j}} [dN_j(s) - \right. \\
&\quad \left. - Y_j(s) \frac{e^{\beta'Z}}{e^{\hat{\beta}'Z}} e^{\hat{\beta}'Z_i} \bar{h}(s, Z) d\Lambda_0(s)] | A, B \right\} \\
&= - \mathcal{E} \{ martingale \} \rightarrow 0
\end{aligned}$$

if $e^{\beta'Z} \approx e^{\hat{\beta}'Z}$

The latter assumption holds if the estimator $\hat{\beta}$ is unbiased.

Thus, just term 2 is remaining. Choose t_0 a fixed time point:

$$\begin{aligned}
term2 &= - \mathcal{E} \left\{ \int_0^t Y(s) e^{\beta' Z} \frac{\bar{h}(t_0)}{h(A, B)} h(A, B) d\Lambda_0(s) | A, B \right\} + \\
&\quad + \mathcal{E} \left\{ \int_0^t Y(s) e^{\beta' Z} [\bar{h}(t_0) - \bar{h}(s, Z)] d\Lambda_0(s) | A, B \right\} \\
&= - \mathcal{E} \left\{ \int_0^t Y(s) e^{\beta' Z} \frac{\bar{h}(t_0)}{h(A, B)} h(A, B) d\Lambda_0(s) | A, B \right\} + R(t, A, B) \\
&= - \frac{\bar{h}(t_0)}{h(A, B)} \mathcal{E} \left\{ \int_0^t Y(s) e^{\beta' Z} h(A, B) d\Lambda_0(s) | A, B \right\} + R(t, A, B)
\end{aligned}$$

$R(t, A, B)$ just acts as a placeholder for the second term which is needed later on in the proof.

Because

- $e^{\beta' Z} h(A, B) d\Lambda_0(s) \equiv \lambda(s | A, B, Z)$
- $\int_0^t Y(s) \lambda(s | A, B, Z) \equiv \Lambda(t)$
- $\mathcal{E}(M) = \mathcal{E}(N) - \mathcal{E}(\Lambda) = 0 \quad \Rightarrow \quad \mathcal{E}(\Lambda) = \mathcal{E}(N)$

it holds that

$$term2 = -\frac{\bar{h}(t_0)}{h(A, B)} \mathcal{E}\{N(t)|A, B\} + R(t, A, B) \quad (3.9)$$

Combining now these findings with the remaining term 2 we get

$$\begin{aligned} \mathcal{E}\{\hat{M}(t)|A, B\} &\approx \mathcal{E}\{N(t)|A, B\} - \frac{\bar{h}(t_0)}{h(A, B)} \mathcal{E}\{N(t)|A, B\} + R(t, A, B) \\ &= \left[1 - \frac{\bar{h}(t_0)}{h(A, B)}\right] \mathcal{E}\{N(t)|A, B\} + R(t, A, B) \end{aligned}$$

This equation is now extended by the factor $1/(\mathcal{E}\{N(t)|A, B\})$:

$$\begin{aligned} \frac{\mathcal{E}\{\hat{M}(t)|A, B\}}{\mathcal{E}\{N(t)|A, B\}} &\approx 1 - \frac{\bar{h}(t_0)}{h(A, B)} + \frac{R(t, A, B)}{\mathcal{E}\{N(t)|A, B\}} \\ \Rightarrow 1 - \frac{\mathcal{E}\{\hat{M}(t)|A, B\}}{\mathcal{E}\{N(t)|A, B\}} &\approx \frac{\bar{h}(t_0)}{h(A, B)} + \frac{R(t, A, B)}{\mathcal{E}\{N(t)|A, B\}} \\ &= \frac{\bar{h}(t_0)}{h(A, B)} \left[1 - \frac{h(A, B)R(t, A, B)}{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\}}\right] \end{aligned}$$

$$\begin{aligned}
&\Rightarrow \log \left[1 - \frac{\mathcal{E}\{\hat{M}(t)|A, B\}}{\mathcal{E}\{N(t)|A, B\}} \right] \\
&\approx \log \bar{h}(t_0) - f(A, B) + \log \left[1 - \frac{h(A, B)R(t|A, B)}{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\}} \right] \\
&\approx f(A, B) - \log \bar{h}(t_0) - \log \left[1 - \frac{h(A, B)R(t, A, B)}{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\}} \right] \\
&= f(A, B) - \log \left[\bar{h}(t_0) \left\{ 1 - \frac{h(A, B)R(t, A, B)}{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\}} \right\} \right] \\
&= f(A, B) - \log \left[\bar{h}(t_0) \left\{ \frac{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\} - h(A, B)R(t, A, B)}{\bar{h}(t_0)\mathcal{E}\{N(t)|A, B\}} \right\} \right] \\
&= f(A, B) - \log \left[\frac{\bar{h}(t_0)\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}h(A, B)d\Lambda_0(s)|A, B\}}{\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}h(A, B)d\Lambda_0(s)|A, B\}} - \right. \\
&\quad \left. - \frac{h(A, B)\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}[\bar{h}(t_0) - \bar{h}(s, Z)]d\Lambda_0(s)|A, B\}}{\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}h(A, B)d\Lambda_0(s)|A, B\}} \right] \\
&= f(A, B) - \log \left[\frac{\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}\bar{h}(s, Z)h(A, B)d\Lambda_0(s)|A, B\}}{\mathcal{E}\{\int_0^t Y(s)e^{\beta'Z}h(A, B)d\Lambda_0(s)|A, B\}} \right]
\end{aligned}$$

$$\begin{aligned} \Rightarrow -\log \left[1 - \frac{\mathcal{E}\{\hat{M}(t)|A, B\}}{\mathcal{E}\{N(t)|A, B\}} \right] &= f(A, B) - \log \tilde{h} \\ &= f(A, B) - \tilde{f} \end{aligned}$$

Thus we have

$$-\log \left[1 - \frac{\mathcal{E}\{\hat{M}(t)|A, B\}}{\mathcal{E}\{N(t)|A, B\}} \right] \approx f(A, B) - \tilde{f} \quad (3.10)$$

$\mathcal{E}(\hat{M}(t)|A, B)$ and $\mathcal{E}(N(t)|A, B)$ can be approximated by a smoothed estimation like LOWESS (Cleveland, 1979) or MARS, perhaps obtained by a smoothed scatterplot of \hat{M}_i against A and B . Hence,

$$f(A, B) - \tilde{f} \approx -\log \left\{ 1 - \frac{\text{smooth}(\hat{M}, A, B)}{\text{smooth}(N, A, B)} \right\} \quad (3.11)$$

A first order Taylor series approximation of the form

$$f(a + h) \approx f(a) + \frac{h}{1!} f'(a)$$

with $a = 1$ and $h = -\frac{\text{smooth}(\hat{M}, A, B)}{\text{smooth}(N, A, B)}$

yields the following:

$$\begin{aligned}
 f(A, B) - \tilde{f} &\approx -\log(1) + \frac{\mathit{smooth}(\hat{M}, A, B)}{\mathit{smooth}(N, A, B)} \cdot \frac{1}{1} \\
 &= \frac{\mathit{smooth}(\hat{M}, A, B)}{\mathit{smooth}(N, A, B)}
 \end{aligned}$$

$$\Rightarrow \mathcal{E}\{\hat{M}(t)|A, B\} = \mathit{smooth}(\hat{M}, A, B) = [f(A, B) - \tilde{f}] \cdot c$$

with c being the expected number of events per subject, i.e. the total number of events divided by the sample size.

The proof shows that the method [11] even works in the context of interactions and the simultaneous modelling of more than one covariate is possible in the survival MARS framework.

Chapter 4

Simulation studies

Simulation studies are an important statistical tool to analyse performance, power and property of a new statistical method or approach. The user creates a set of data, whereas the response variable y depends in a predefined way on the covariates. Thus, one actually knows which results has to be obtained, because the functional coherence between dependent variable and predictor variables is self-defined.

Jerome Friedman [12] introduced some simulation studies for his MARS approach. He showed that MARS is able to detect pure noise data and recognize mere additive relationships, i.e. the approach avoids to find "structure when it is not present" [12]. As a matter of course, it was then also shown, that the method reliably detects structure "when it does exist", albeit disturbed by pure noise variables.

Jerome Friedmans approach differs from the one used in this thesis, namely the approach of Kooperberg. Firstly, two basis functions (not only one as suggested by Kooperberg) are entered into the model per step in the forward selection procedure, namely the sibling pair $\pm(x - t)_+$, secondly interactions of higher order than $p = 2$ are allowed and thirdly the goodness-of-fit criterion GCV is defined somewhat different:

$$GCV_{Friedman} = \frac{1}{N} \frac{\sum_{i=1}^N [y_i - \hat{f}_M(x_i)]^2}{[1 - \frac{C(M)+d \cdot M}{N}]^2}$$

with

$$C(M) = \text{trace}[B(B^T B)^{-1} B^T] + 1$$

and B the $M \times N$ matrix of the M basis functions.

On this account the results of Friedman may differ substantially from the MARS approach of Kooperberg.

Charles Kooperberg [16] showed MARS based simulations solely for classification problems, i.e. the dependent variable was categorical. He compared several classification methods and arrived at the conclusion that MARS performs at least as well as other established methods for binary response, but yields the enormous advantage of being much faster in fitting huge datasets.

Due to the differing approach of Friedman and the mere binary response situation in Kooperbergs investigations further simulation studies on the ability of MARS will be performed in this thesis, especially for structures which are of high interest in the medical context. Among these are nonlinear effects of the covariates and/or interactions between them. Furthermore the power of MARS in the setting of survival time data is of high interest.

The thesis only presents studies on continuous dependent variables, for just these are interesting in the context of survival MARS. Indeed, simulation studies on binary responses were done, but are not shown in this paper.

Several simulation studies were done to investigate the power of multivariate adaptive regression splines in these special settings. First of all the general performance of MARS was checked: functionality, stability and reliability of this approach were analysed in several scenarios. These studies are presented in the first section.

The second part deals with the ability of MARS to assess interactions between a continuous and a binary covariate, e.g. age and gender.

In the last section the simulations for survival data are introduced with different distributions for the survival time, more precisely exponential and Weibull distribution.

4.1 Monograph of the approach

Three scenarios with diverse influences of the covariates and changing error variance were established to analyse the power of MARS. Let $x_1, \dots, x_{10} \sim U[-1, 1]$ be some covariates and $\epsilon \sim N(0, \sigma^2)$ with $\sigma^2 \in \{0.1, 1, 2\}$ the error. Now in the two scenarios A and B the dependent variable y is defined as

$$\mathbf{A} \quad : \quad y = 10x_1x_2 + 5x_3 + 10 \sin(4x_4) + \log |x_5| + \epsilon$$

$$\mathbf{B} \quad : \quad y = x_1x_2 + x_3 + \sin(4x_4) + \log |x_5| + \epsilon$$

Thus, nonlinear effects of the covariates, interactions and the mere nuisance parameters x_6, \dots, x_{10} are to be identified by the method. Scenario B was performed to check on the ability of MARS to find even small effects of the covariates. The differing variation ought to show whether MARS can handle increasing noise in the data.

A third scenario was established with regard to the subsequent analysis of survival time data, i.e. here the same functional relationship should be assumed as later on.

For that purpose additional variables x_{11} to x_{15} with different distributions were generated:

$$x_{11} \sim U[30, 75]$$

$$x_{12} \sim Bi(n, 0.5)$$

$$x_{13} \sim N(0, 1)$$

$$x_{14} \sim U[0, 10]$$

$$x_{15} \sim Bi(n, 0.7)$$

with n being the sample size. The following relationship was then performed:

$$\mathbf{C} \quad : \quad y = 0.04 \left(\frac{x_{11}}{10} \right)^2 - 6x_{12} + 0.1x_{11}x_{12}$$

Thus, x_{13} to x_{15} are again mere nuisance parameters and it ought to be found a nonlinear structure and an interaction term between a continuous and a binary variable.

1000 simulations with 500 and 1000 observations, respectively, were computed for every scenario and differing values of the penalizing parameter d . Remind that d is the tuning parameter in the denominator of the generalized crossvalidation criterion. A common value of d is 4, smaller values yield a larger model with more basis functions, larger values of d result in smaller models with less basis functions.

Results are documented as fraction of models with false basis functions (`%false`) and fraction of models with interaction between x_1 and x_2 (`% x_1x_2`) and x_{11} and x_{12} (`% $x_{11}x_{12}$`), respectively.

For scenario C additionally the fraction of models with nonlinear x_{11} is documented (`%knots(x_{11})`).

A model with false basis functions is defined as follows:

1. a model with not all of the simulated coherences found, e.g. interaction not found or linear influence of x_3, x_4, x_5 (scenario A and B) or x_{11} (scenario C)
2. a model with too many basis functions, e.g. basis functions of the nuisance parameters or false interactions

Table 4.1 shows the results of this first simulation study for penalizing parameters $d \in \{3, 4, 5\}$.

Table 4.1: Results of the simulations: fraction of simulation cycles with spuriously chosen basis functions, correct identified interaction terms and correct identified nonlinear influence of x_{11} in scenario C

		d = 3						
σ^2	sample size	A		B		C		
		%false	% x_1x_2	%false	% x_1x_2	%false	% $x_{11}x_{12}$	%knots(x_{11})
0.1	500	37.3	100	41.7	100	6.1	100	100
	1000	40.6	100	37.3	100	7.0	100	100
1	500	17.9	100	18.1	99.9	4.9	100	100
	1000	20.2	100	20.4	100	5.1	100	100
2	500	24.7	100	14.5	46.8	5.1	100	100
	1000	19.9	100	19.7	90.1	6.2	100	100
		d = 4						
σ^2	sample size	A		B		C		
		%false	% x_1x_2	%false	% x_1x_2	%false	% $x_{11}x_{12}$	%knots(x_{11})
0.1	500	21.7	100	22.1	100	1.2	100	100
	1000	20.5	100	16.5	100	2.1	100	100
1	500	4.1	100	5.7	100	1.2	100	100
	1000	6.4	100	5.3	100	1.5	100	100
2	500	26.3	100	5.8	22.7	0.6	100	100
	1000	6.0	100	4.4	70.7	2.0	100	100
		d = 5						
σ^2	sample size	A		B		C		
		%false	% x_1x_2	%false	% x_1x_2	%false	% $x_{11}x_{12}$	%knots(x_{11})
0.1	500	9.8	100	11.8	100	0.7	100	100
	1000	11.9	100	11.6	100	0.8	100	100
1	500	2.2	100	2.0	92.9	0.5	100	100
	1000	2.2	100	3.1	100	0.7	100	100
2	500	41.8	100	6.7	12.6	0.2	100	100
	1000	3.16	100	2.8	49.8	0.5	100	100

The studies show the following results:

Scenarios A and B

1. amount of models with false basis functions:

In both scenarios the fraction of models with spuriously chosen basis functions reduces with increasing penalizing parameter d . An exception forms scenario A with $\sigma^2 = 2$ and a sample size of 500, here the fraction of models with spuriously chosen basis functions is at least 25% ($d = 3$), which may be in fact caused by the combination of small sample size and large error variance.

2. fraction of models with interaction:

In scenario A 100% of the simulations result in models with an interaction term between x_1 and x_2 .

In scenario B the size of the error variance affects the fraction of models with interaction term. An error variance of 0.1 produces 100% of models with interaction term and a $\sigma^2 = 1$ yielded at least 92.9% models with interaction term. In the case of $\sigma^2 = 2$ the sample size strongly influences the findings. A large sample size results in a higher proportion of models with interaction term.

Besides, the size of d has an effect on this topic: the smaller

d , i.e. the more basis functions in the largest model of the stepwise addition process, the larger the interaction fraction.

3. In the case of a small error variance ($\sigma^2 \leq 1$) the sample size has no influence on the quality of the findings in scenario A. The fraction of models with false basis functions remains approximately equal. However in the case of a larger error variance ($\sigma^2 = 2$) the sample size do affect the outcome. A larger sample size results in a smaller amount of spurious models and vice versa.

In scenario B the connection between sample size and error variance cannot be observed. One third of the cases shows a higher fraction of false models for 1000 observations than for 500 observations, however in two thirds of the cases a larger sample size results in a smaller amount of wrong models.

Figures 4.1 to 4.3 show the mean functional form of x_3 to x_{10} for a selection of simulated scenarios A and B. The interaction of x_1 and x_2 was not plotted due to the lack of interpretability of the figure. The plots for the variables x_6 to x_{10} demonstrate a mean influence of these variables of zero, i.e. even if some models have found the mere nuisance parameters as influential, the overall result must be that they have no influence on the response at all. The plots of the results for the other parameter combinations look quite identical, thus are not printed.

The high amount of models with false basis functions in the case of $\sigma^2 = 2$ results from the fact, that x_5 is seldomly detected as influential on the response, which is due to the small simulated coefficient of 1. The larger the simulated effect is, the more likely the according variable is entered into the model. Considering a random sample of the models with false basis functions more precisely showed, that the false chosen basis functions had very small coefficients, thus the effect of them remains small and hence can be neglected.

Scenario C

Scenario C yields in 100% of the cases models with found interaction and nonlinear x_{11} and at most 7% spuriously chosen basis functions, which means the entering of other covariables than x_{11} and x_{12} into the model. Figure 4.4 shows the mean results for a sample size of 1000 and the parameter combination $d = 4$, $\sigma^2 = 1$. All other combinations yielded comparable plots and thus are not presented. The results are convincing: in all simulation cycles the MARS approach finds the correct functional relationship between dependent variable and covariates and only a few cycles detected influences of the nuisance parameters, which is due to the value of the penalizing parameter, as the fraction of false basis functions diminishes with increasing d and decreasing allowed model size in the stepwise addition process.

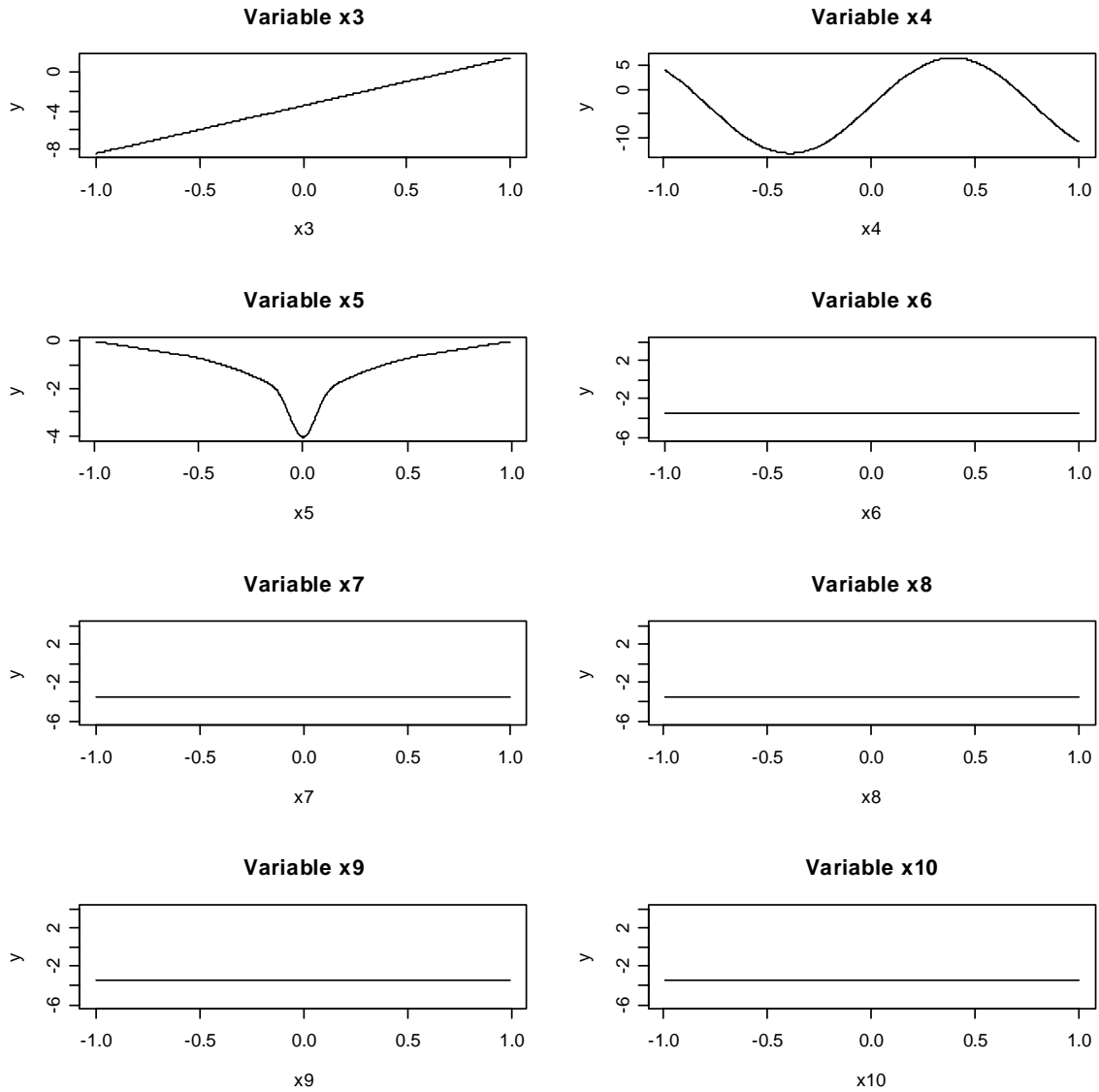


Figure 4.1: scenario A, 500 observations, $\sigma^2 = 0.1$, $d = 3$

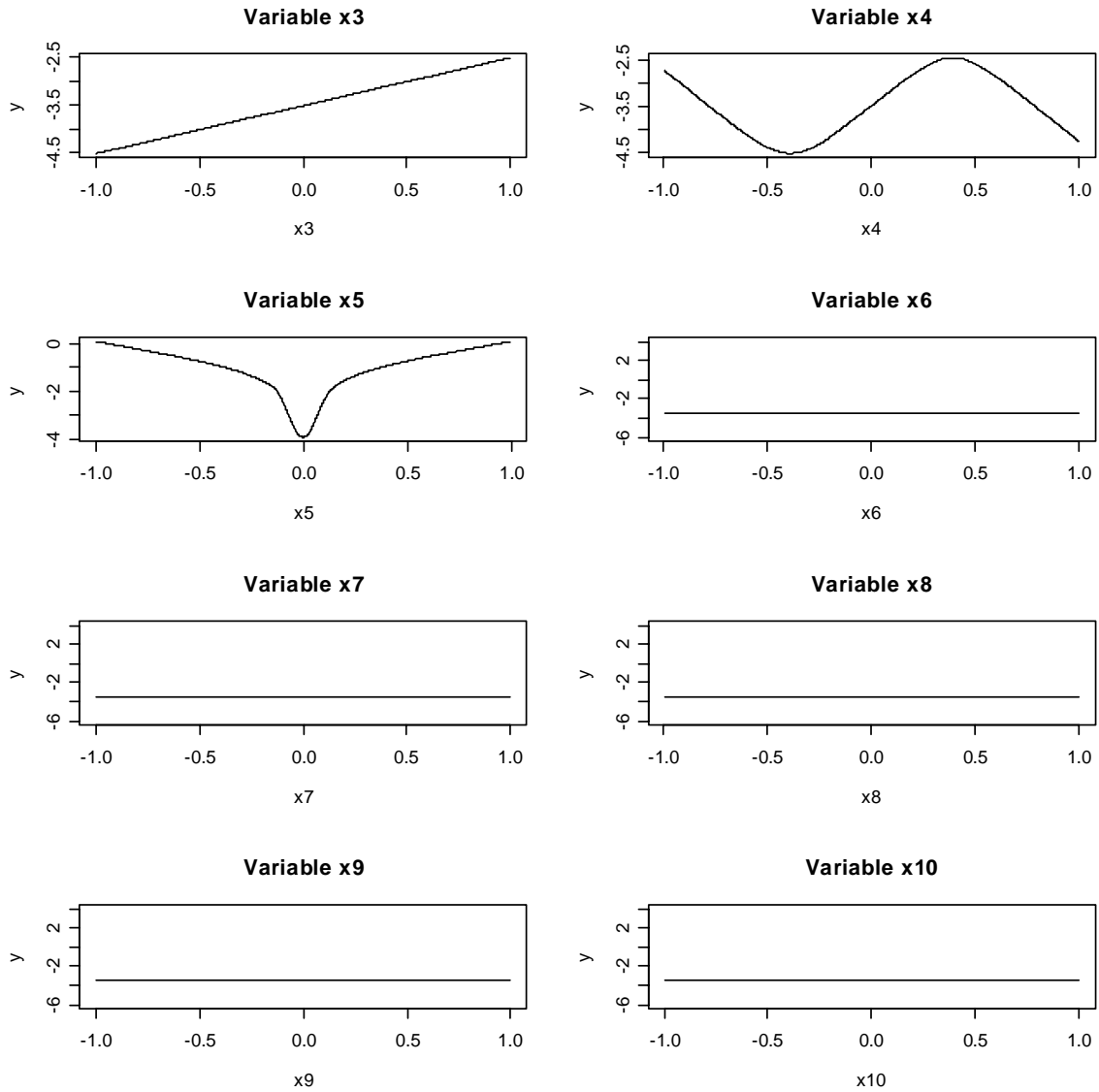


Figure 4.2: scenario B, 1000 observations, $\sigma^2 = 1$, $d = 4$

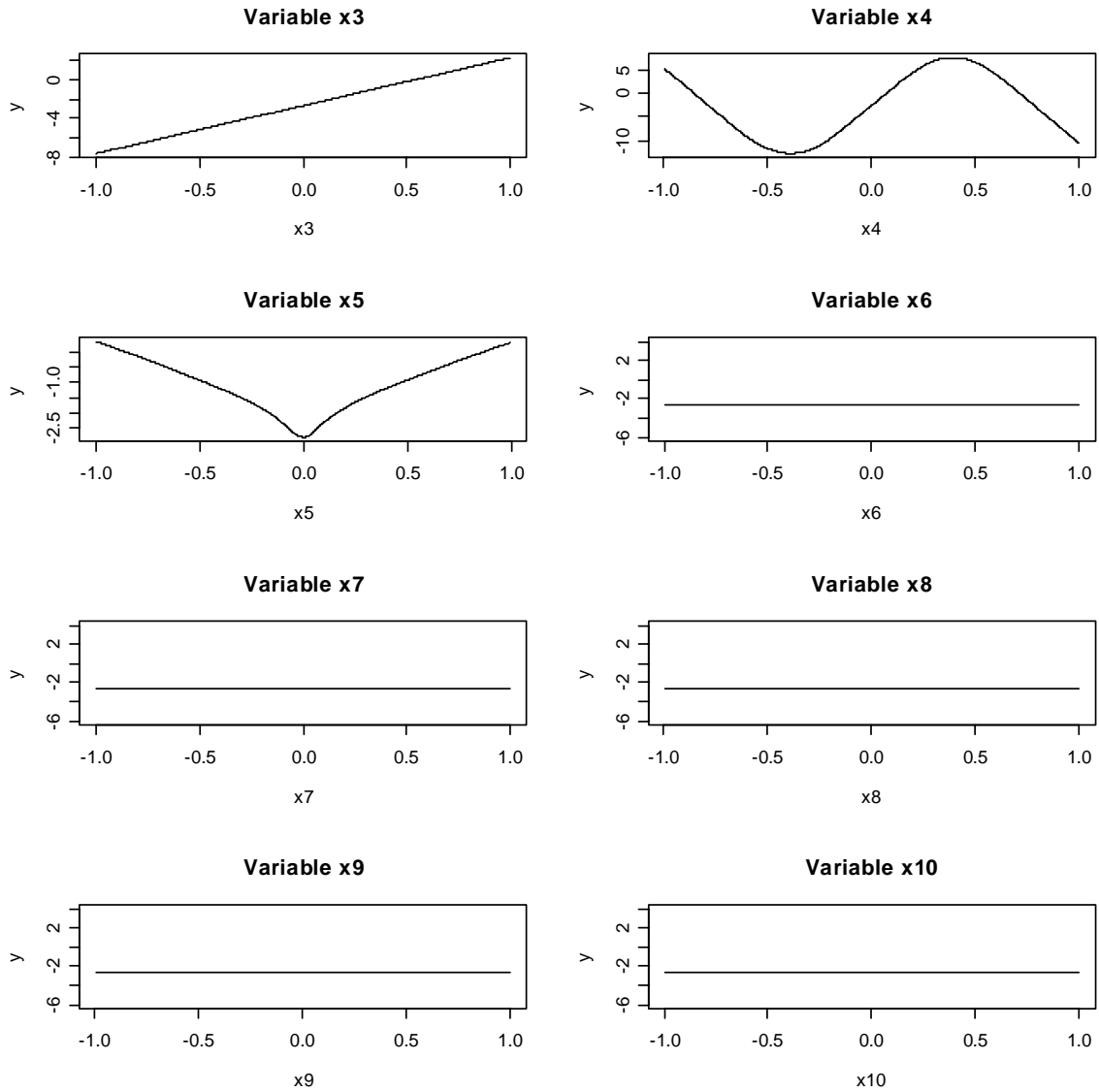


Figure 4.3: scenario A, 1000 observations, $\sigma^2 = 2$, $d = 5$

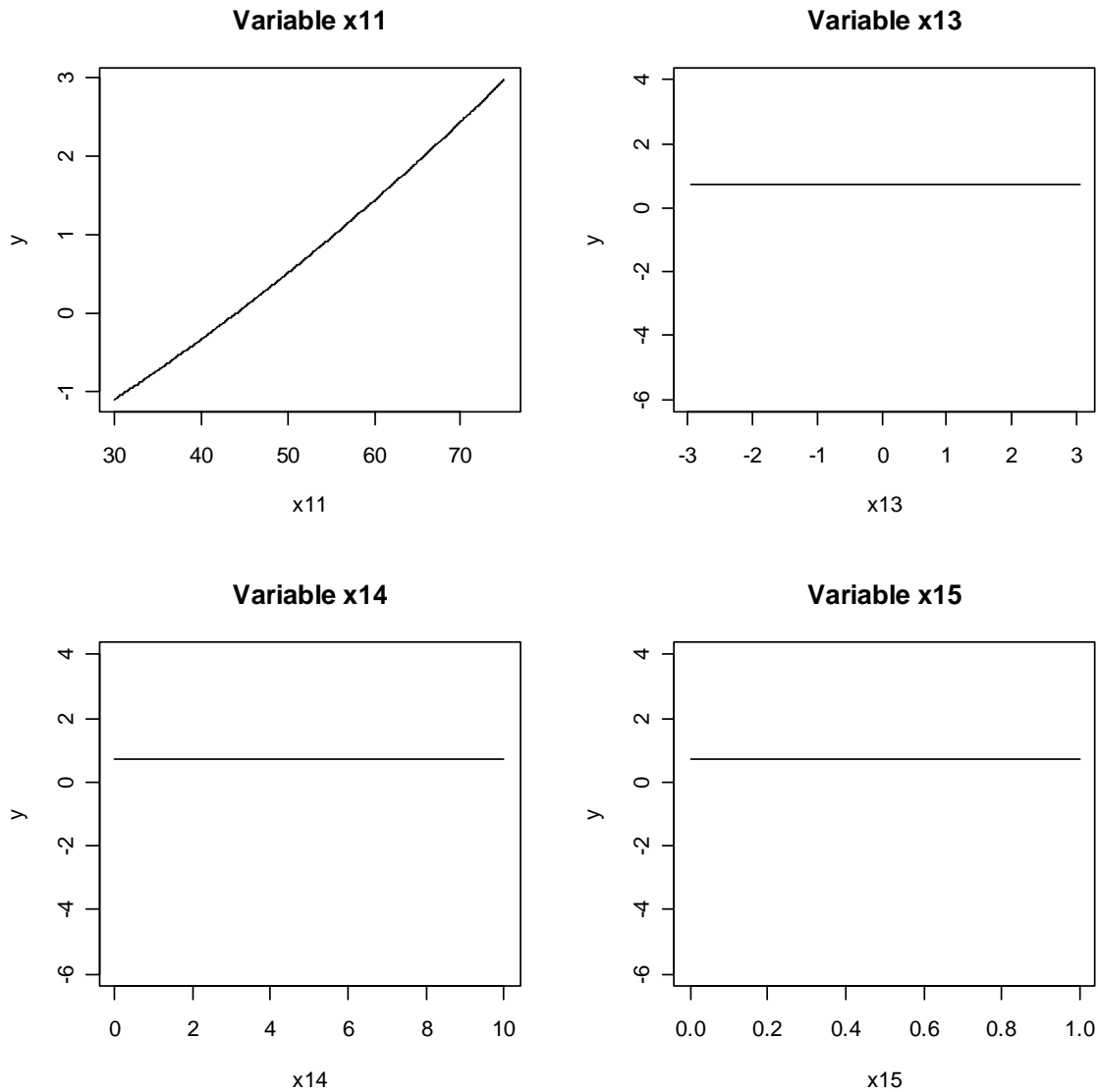


Figure 4.4: scenario C, 1000 observations, $\sigma^2 = 1$, $d = 4$

The results show, that MARS is able to find linear and nonlinear effects of covariates on the response. Even when the fraction of models with false basis functions is large the mean functional form remains more or less correct, only the log-transformation of x_5 in scenario B becomes more v-shaped in the mean result because of a higher

fraction of models, which do not include this relationship. Even the handling of binary covariates is no challenge for the method.

The conclusion of these simulation studies must be, that multivariate adaptive regression splines present a reasonable way to model complicated types of connections between response and covariates, including interactions and nonlinear effects.

4.2 Interaction studies

The analysis of a special type of influence of covariates on the dependent variable shall be deepened in this section, namely the interaction between a continuous and a binary covariate, e.g. age and gender, which was already outlined in the previous section. Other methods often have problems to detect these coherences, so it is of high interest, whether MARS is able to reveal such kind of connection between covariates. 1000 simulations with 500 cases, respectively, were performed with the following relationship:

$$y = \frac{(z - 40)^2 - 10z - 2x - 7x(z - 50)_+ + \epsilon}{100} \quad (4.1)$$

with $x \sim Bi(500, 0.5)$, $z \sim U[40, 75]$ and $\epsilon \sim N(0, \sigma^2)$, $\sigma^2 \in \{1, 10, 25, 50\}$. The variable x can be seen as "gender", z can be

seen as "age". As measure of goodness of fit $R^2 = \frac{SSE}{SST}$ was calculated and for all simulations reported as mean R^2 and the range of the mean R^2 . Table 4.2 shows the results of these simulation studies.

Table 4.2: Results of the simulations with an interaction between a continuous and a binary covariate

σ^2	mean(R^2)	range(R^2)
1	0.9991	0.9989 - 0.9993
10	0.9974	0.9962 - 0.9981
25	0.9880	0.9838 - 0.9905
50	0.9561	0.9397 - 0.9661

Figure 4.5 shows the mean results of the interaction studies for the four different error variances.

Table and graphics yield a good overview of the performance of MARS in this particular scenario. All grades of error rate result in convincing outcomes with an R^2 of at least 0.94. The graphs clearly show the ability of the approach to distinct between the two groups $x = 0$ and $x = 1$, even in the case of strong noise in the data.

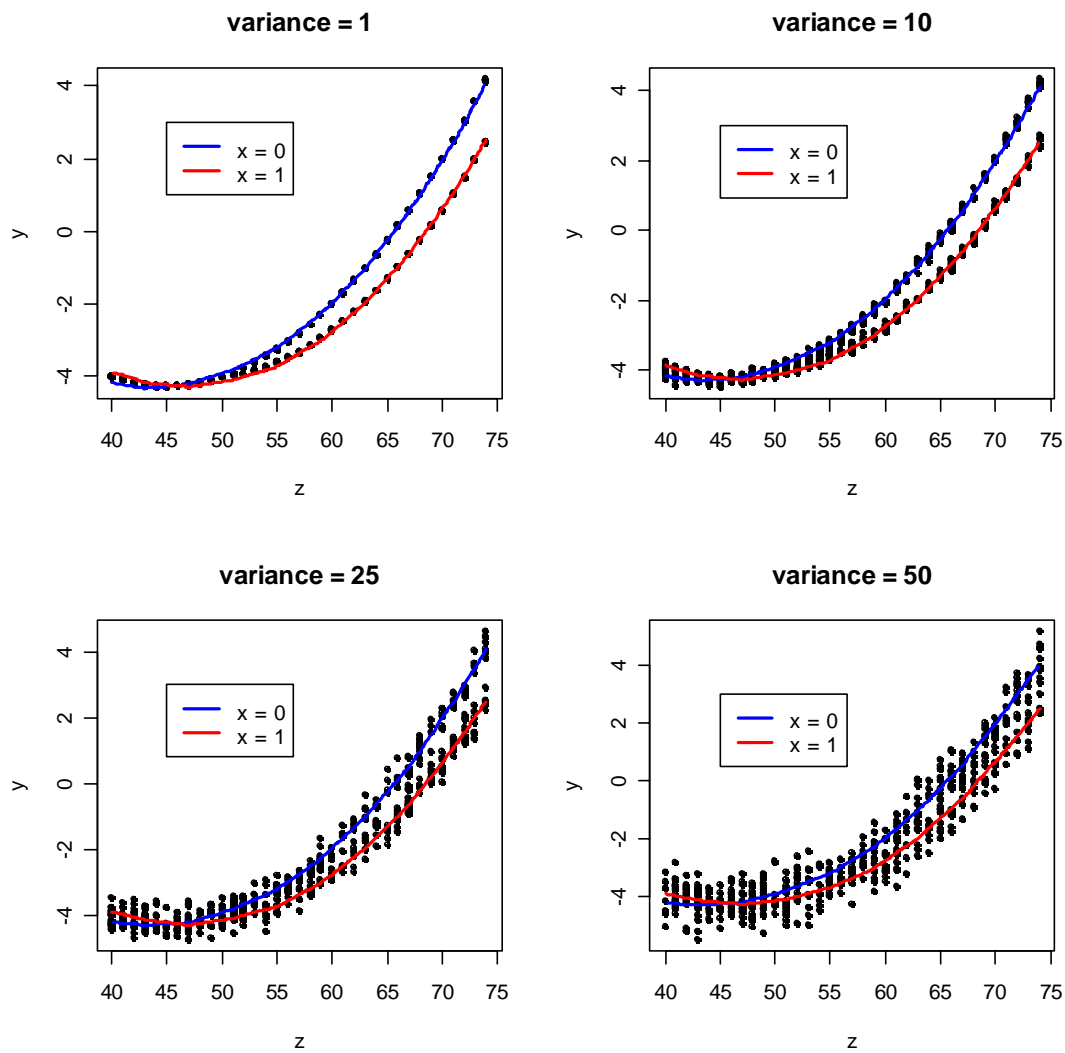


Figure 4.5: MARS results for an interaction between a binary and a continuous covariate

Hence, the simulation study demonstrated that MARS is up to find even this difficult kind of interaction between a continuous and a binary variable.

4.3 Simulations on survival time data

The actually interesting type of data to test the approach of multivariate adaptive regression splines is in fact survival time data. In this situation, apart from a status variable, which indicates whether the event of interest has occurred or not, an observational time was measured for every subject in the study.

The new approach in this thesis proposes the use of residuals of the Cox null model, namely martingale or deviance residuals, as new response variable in a common MARS setting.

In this section simulation studies on this topic shall be introduced. But first of all, the method to generate survival time data used in this thesis is presented.

The simulation of survival time data

A method to generate survival times to simulate Cox proportional hazards models [5] is the following:

The survival function of the Cox PH model is defined as

$$S(t|x) = \exp[-\Lambda_0(t) \exp(f(x))] \quad (4.2)$$

with

$$\Lambda_0(t) = \int_0^t \lambda_0(u) du \quad (4.3)$$

being the cumulative baseline hazard function. Consequential the distribution function of the Cox model is given by

$$F(t|x) = 1 - \exp[-\Lambda_0(t) \exp(f(x))]. \quad (4.4)$$

Set Y a random variable with distribution function F , then $U = F(Y)$ is uniformly distributed on $[0, 1]$: $U \sim U[0, 1]$. Furthermore, if $U \sim U[0, 1]$, then it holds that $(1 - U) \sim U[0, 1]$.

Now let T be a survival time of the Cox model, then it follows that

$$U = 1 - F(t|x) = \exp[-\Lambda_0(t) \exp(f(x))] \approx U[0, 1] \quad (4.5)$$

As $\lambda_0(t) > 0$ for all t , $\Lambda_0(t)$ can be inverted and the survival time T of the Cox model can be expressed as

$$T = \Lambda_0^{-1}(t)[-log(U) \exp(f(x))] \quad (4.6)$$

with U being a random variable with $U \sim U[0, 1]$.

Table 4.3 shows the formulae for generating survival times and the hazard function for exponential and Weibull distributed time variables.

Table 4.3: Formulae for generating survival times and the hazard function

	distribution	
	exponential	Weibull
survival time	$T = -\frac{\log(U)}{\lambda \exp(f(x))}$	$T = \left(-\frac{\log(U)}{\lambda \exp(f(x))}\right)^{\frac{1}{v}}$
hazard function	$\lambda(t x) = \lambda \exp(f(x))$	$\lambda(t x) = \lambda \exp(f(x))vt^{v-1}$

4.3.1 Simulation studies with exponential distribution

On the basis of the approach of generating survival times from above the following data was generated for $n = 1000$ observations:

$$\begin{aligned}
 x_1 &\sim U[30, 75] & x_4 &\sim U[0, 10] \\
 x_2 &\sim Bi(n, 0.5) & x_5 &\sim Bi(n, 0.7) \\
 x_3 &\sim N(0, 1) & u &\sim U[0, 1] \\
 \lambda &= 0.5
 \end{aligned}$$

and the exponential distributed survival time t is defined as

$$t = \frac{-\log(u)}{0.5 \cdot \exp\left(0.04 \left(\frac{x_1}{10}\right)^2 - 6x_2 + 0.1x_1x_2\right)} \cdot 100. \quad (4.7)$$

Then a new variable 'censoring time' is defined as $ct \sim U[0, t_{m\%}]$ with $t_{m\%}$ being the according quantile of t and $m \in \{0.89, 0.7, 0.26\}$. These values of m yield censoring rates of 25%, 50% and 85%, respectively.

Now the censoring indicator $\delta_i, i = 1, \dots, n$ is generated as follows:

$$\delta_i = \begin{cases} 1 & , \text{ if } t_i \leq ct_i \\ 0 & , \text{ else} \end{cases}$$

At last for every censored subject, i.e. for every subject with $\delta = 0$, the value of its time variable t was substituted by its corresponding censoring time ct . This resulting time variable was now used as observational time for further analysis.

1000 simulation cycles were calculated for 4 different values of the penalizing parameter d : $d \in \{2, 3, 4, 5\}$, with martingale and deviance residuals from the Cox null model as response, respectively. Table 4.4 shows the results of these simulation studies.

The findings are reported as fraction of models with

- a nonlinear x_1 , i.e. MARS found at least one knot for this covariate (`%knots(x1)`)
- found interaction between x_1 and x_2 (`%x1x2`) and

- false basis functions (`%false`) all in all, whereas models with false basis functions are meant to contain
 - the covariate x_1 in only linear form
 - no interaction term between x_1 and x_2
 - covariates x_3 , x_4 or x_5 or
 - other interactions than between x_1 and x_2

Figures 4.6 to 4.9 again display the mean basis functions of the executed simulation cycles, exemplarily one type of residual for every simulated penalizing parameter d and for varying censoring rates. The interaction between x_1 and x_2 is clearly observable as well as the nonlinear effect of x_1 . For the other mere nuisance parameters little influence is identifiable, as the mean slope is sort of zero in all scenarios. The fact that the slope marginally increases for the nuisance parameters in the case of smaller d is easy to explain by the higher number of allowed basis functions in these settings. Thus it is more likely for basis functions of x_3 to x_5 to wrongly enter the model.

Table 4.4: Results of the simulation studies on exponential distributed survival time data in percent

d	censoring rate	martingale residuals			deviance residuals		
		%knots(x_1)	% x_1x_2	%false	%knots(x_1)	% x_1x_2	false
2	25%	90.5	100	31.2	87.0	100	35.9
	50%	99.6	100	24.0	100	99.9	25.8
	85%	100	85	40.8	100	87.5	36.9
3	25%	60.7	100	43.3	71.1	100	32.7
	50%	98.0	100	7.6	99.9	99.8	6.8
	85%	99.9	84.4	23.1	100	85.4	20.7
4	25%	38.5	100	62.4	57.9	100	43.8
	50%	96.4	100	5.2	99.7	99.9	20.5
	85%	99.9	81.1	21.1	99.9	82.4	20.5
5	25%	25.7	100	74.7	44.2	100	56.4
	50%	92.1	100	8.8	99.5	99.5	2.1
	85%	99.3	74.8	26.8	99.7	76.6	25.4

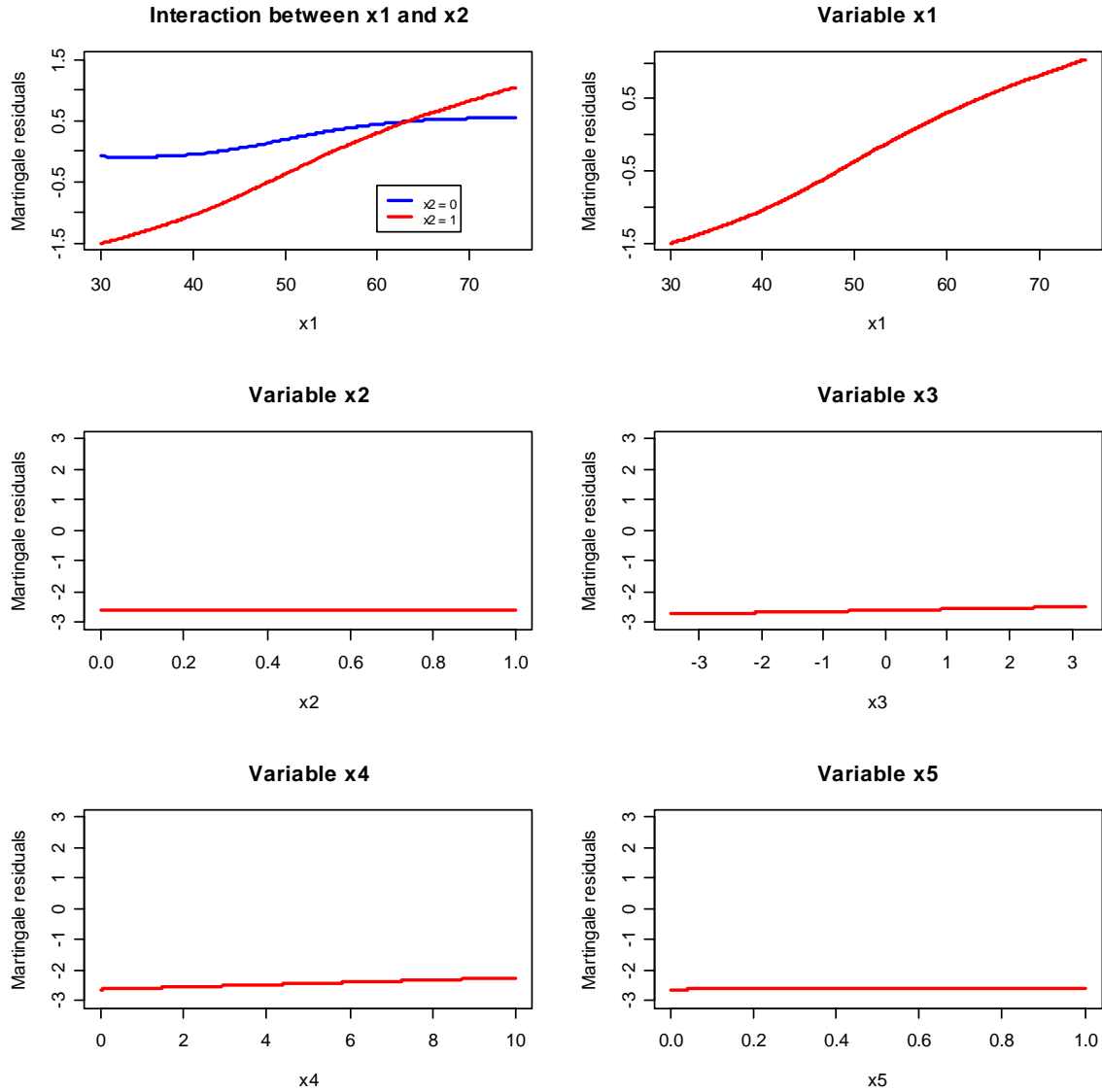


Figure 4.6: Martingale residuals, $d = 2$, 25% censoring rate

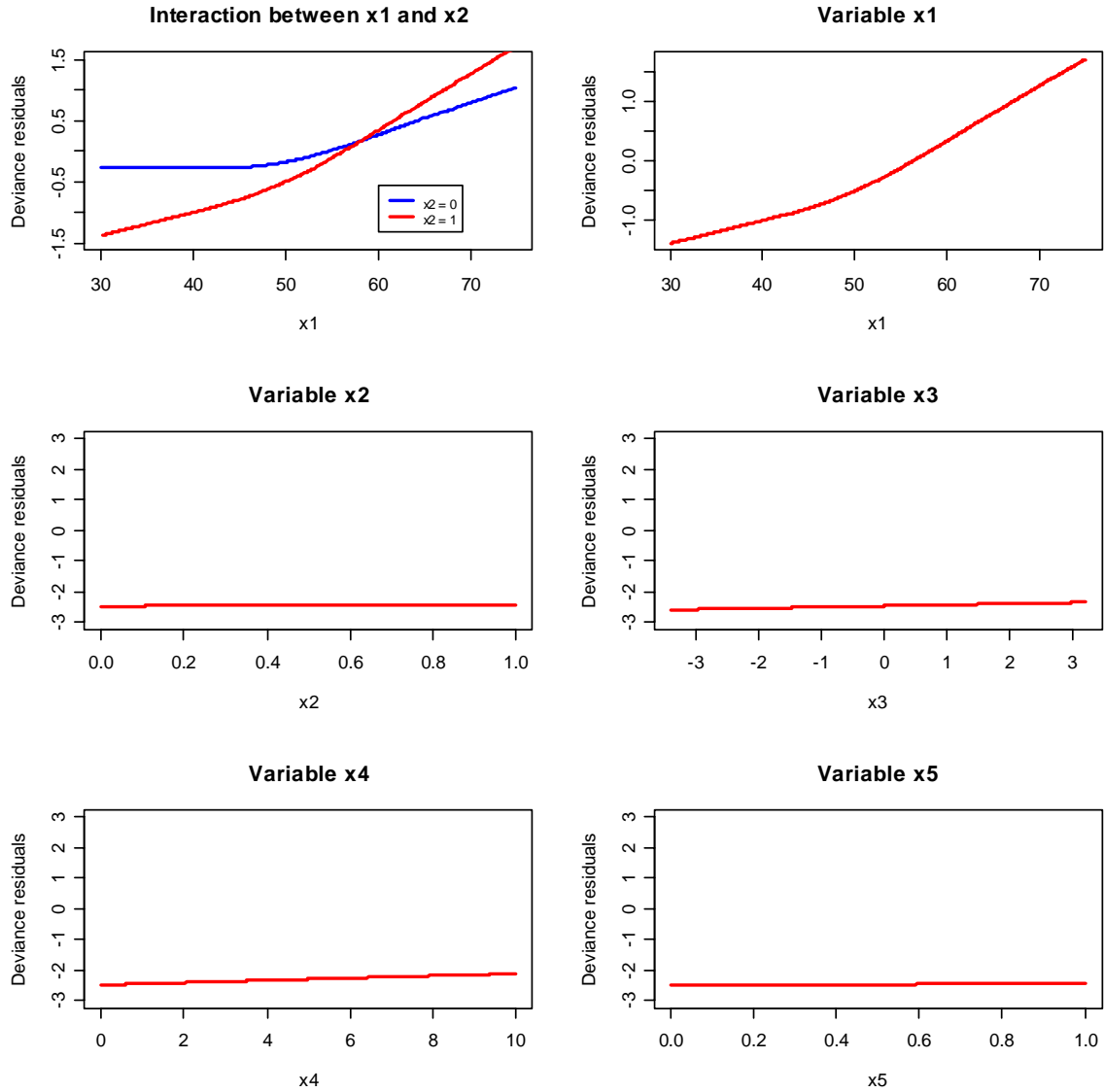


Figure 4.7: Deviance residuals, $d = 3$, 50% censoring rate

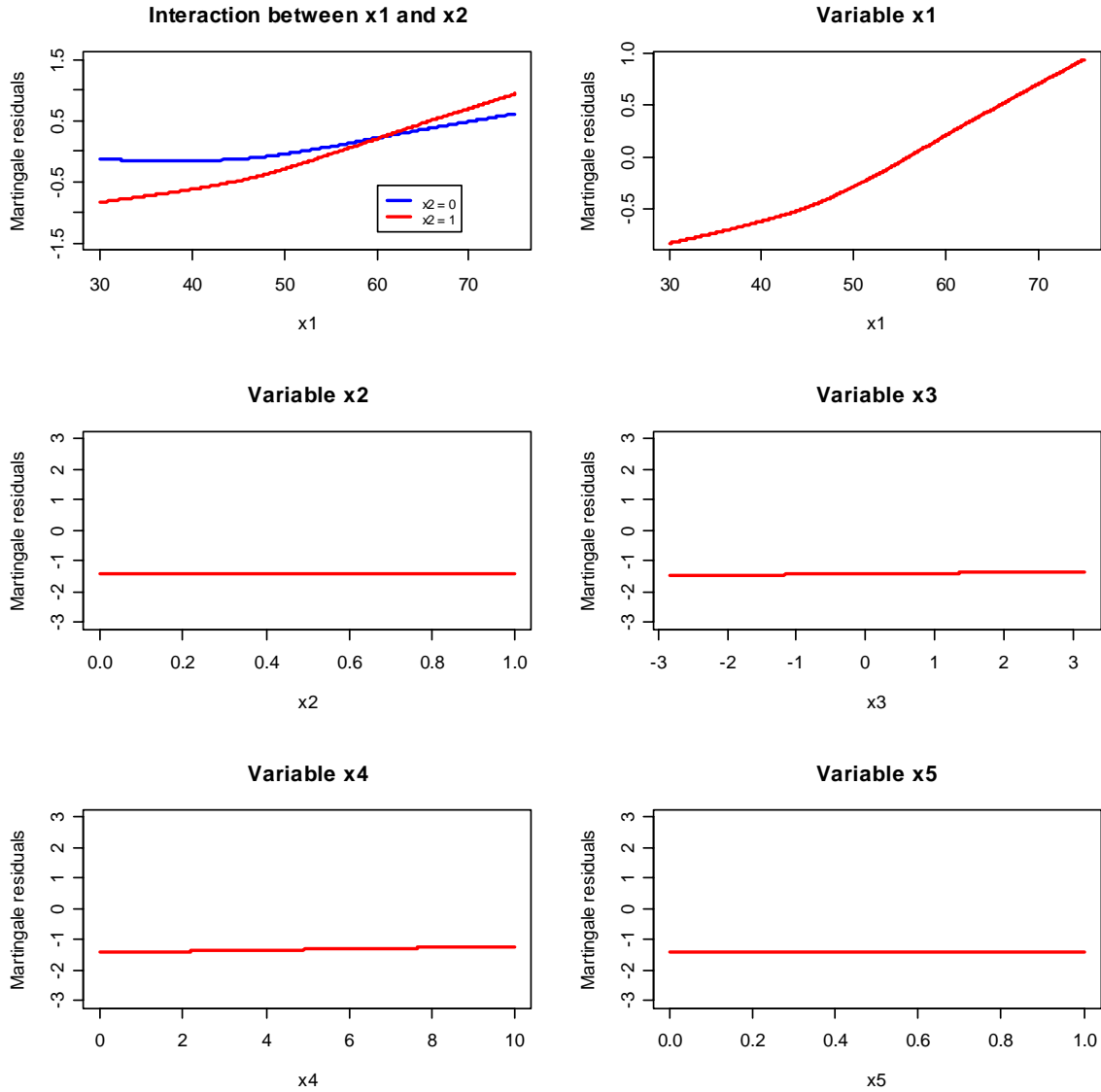


Figure 4.8: Martingale residuals, $d = 4$, 50% censoring rate

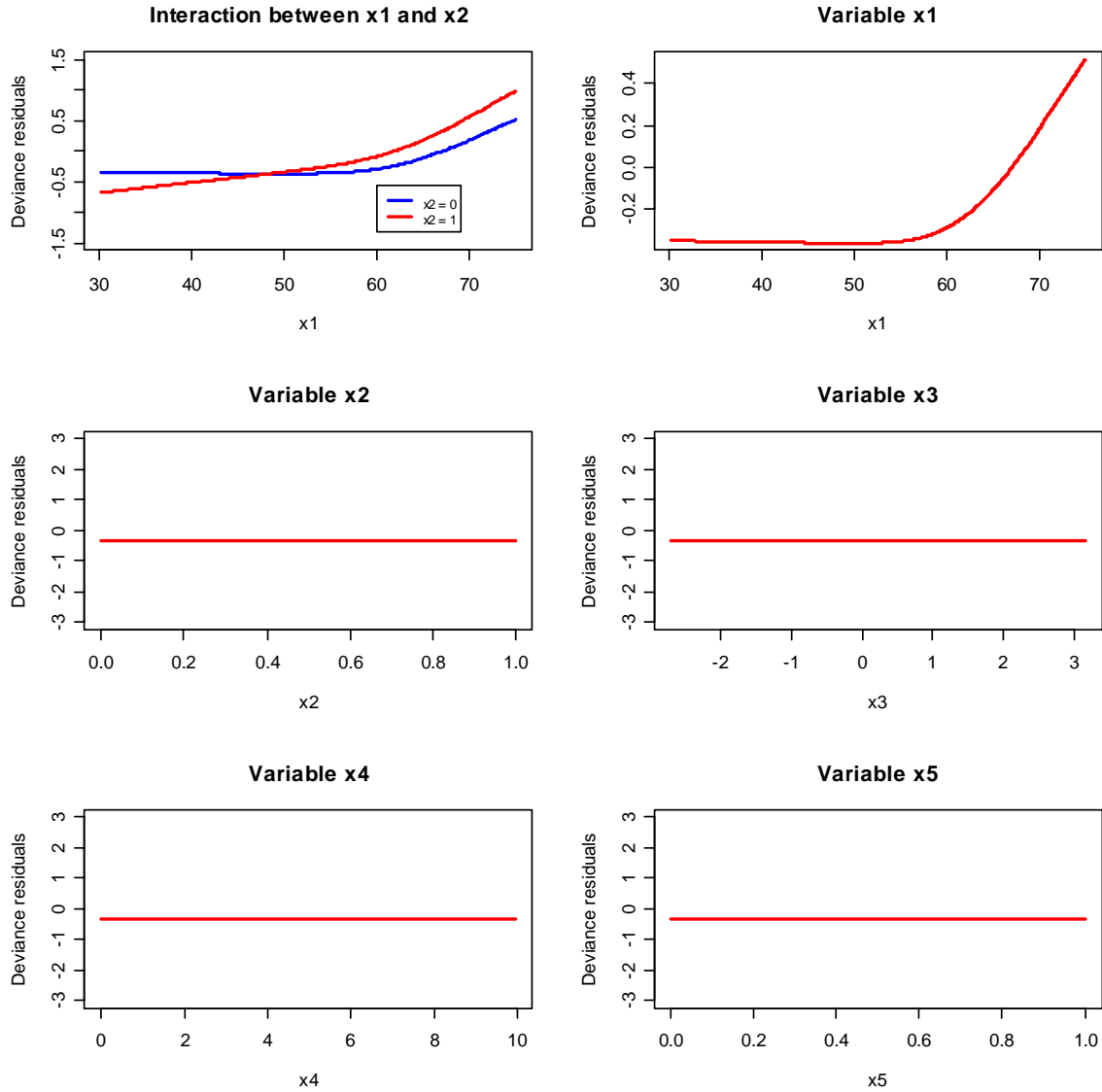


Figure 4.9: Deviance residuals, $d = 5$, 85% censoring rate

The following results could be observed:

- up to 100% of the models detected a nonlinear influence of x_1 in the case of high censoring rates.
- in at least 99.5% of the simulation cycles with moderate to low censoring rates the interaction term was found by the approach
- a censoring rate of 50% yields the lowest fraction of spuriously chosen basis functions in all settings
- in the case of penalizing parameter $d = 2$ too many basis functions were found, thus the fraction of false basis functions is relatively high
- in the case of penalizing parameter $d = 5$ too less basis functions could be found due to the size restriction, thus the fraction of false basis functions is high here, too
- the approach with martingale residuals of the Cox null model as response finds knots in variable x_1 a little more frequently as the approach with deviance residuals
- the two approaches are approximately equal considering the fraction of models with found interaction term between x_1 and x_2
- also the amount of spuriously chosen basis functions is nearly equal in both settings

As measure of the performance of the new method, martingale residuals were used. On this account, in every simulation cycle the basis functions found by the MARS approach were included as covariates in a Cox model and the resulting martingale residuals and their variance were calculated. Additionally, classical Cox models were performed on the simulated data in every cycle and the corresponding martingale residuals and their variance were calculated, too.

Exemplarily, two of the different scenarios are displayed here:

Figure 4.10 shows the scatterplot of the resulting variances together with the bisecting line of the scenario with martingale residuals as response in a MARS approach, a penalizing parameter $d = 4$ and a censoring rate of 50%. In 77.1% of the simulations the MARS approach yields a smaller variance than the classical Cox approach ($p(\text{sign test}) < 0.0001$). The classical Cox model yields a mean variance of 0.505, whereas the MARS approach shows a slightly less mean variance of 0.490.

Figure 4.11 shows the scatterplot of variances of the scenario with deviance residuals as response, a penalizing parameter $d = 4$ and a censoring rate of 50%. Again, most of the items (84.9%) lie under the bisecting line, thus MARS shows smaller variances of the martingale residuals as Cox, i.e. MARS yields a better fit to the data ($p(\text{sign test}) < 0.0001$). In this scenario the classical Cox approach yields a mean variance of 0.511, the MARS approach with deviance

residuals as response yields a mean variance of 0.494.

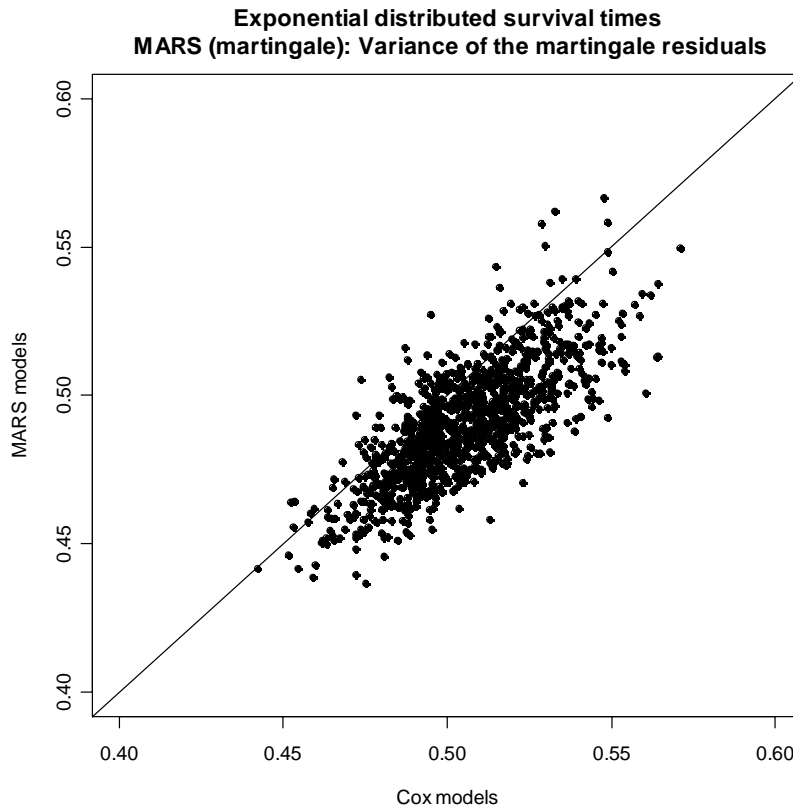


Figure 4.10: Comparison of the variances of MARS with martingale residuals as response and the classical Cox approach with exponential distributed survival times and a censoring rate of 50% ($d = 4$)

4.3.2 Simulation studies with Weibull distribution

The density function of the Weibull distribution is defined as

$$f(t) = \lambda v t^{v-1} \exp(-\lambda t^v) \quad (4.8)$$

with shape parameter v and scale parameter λ . A shape parameter

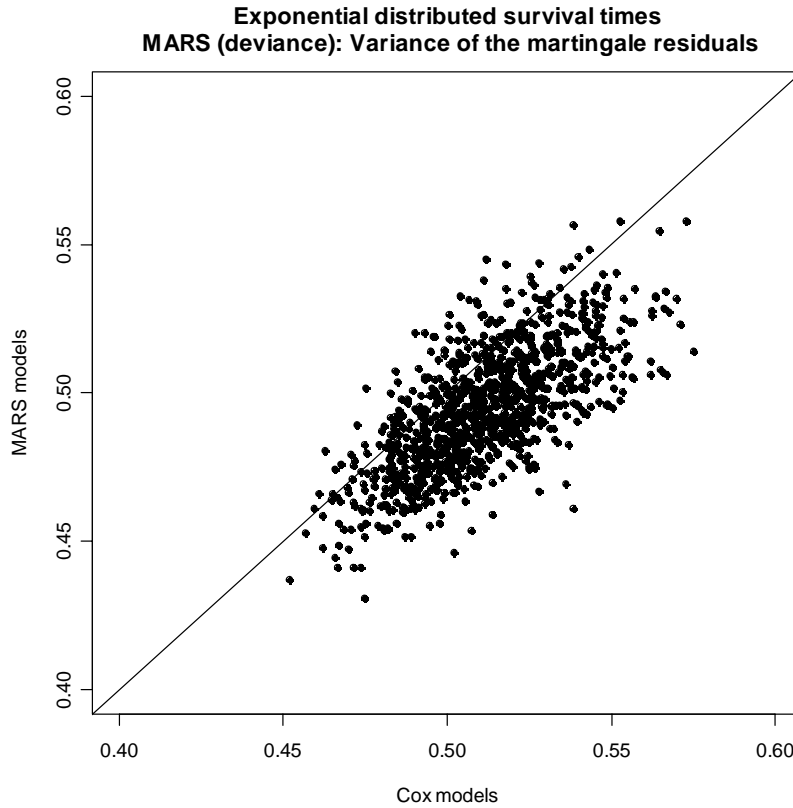


Figure 4.11: Comparison of the variances of MARS with deviance residuals as response and the classical Cox approach with exponential distributed survival times and a censoring rate of 50% ($d = 4$)

v of 1 results in the special case of an exponential distribution. Figure 4.12 shows the density function of the Weibull distribution for varying scale and shape parameters and figure 4.13 displays martingale and deviance residuals of the Cox null models for the data defined below for shape parameter $v \in \{2, 5\}$ and scale parameter $\lambda \in \{2, 5\}$.

The graphics show the following: the larger the shape parameter v , the shorter the resulting observation times for each subject, i.e.

the main part of the observed time values lays in a very narrow area. Correspondingly the interquartile ranges of the martingale and deviance residuals gets smaller for increasing values of v . This means that for most of the subjects very similar survival times are observed, and thus a statistical analysis gets more difficult.

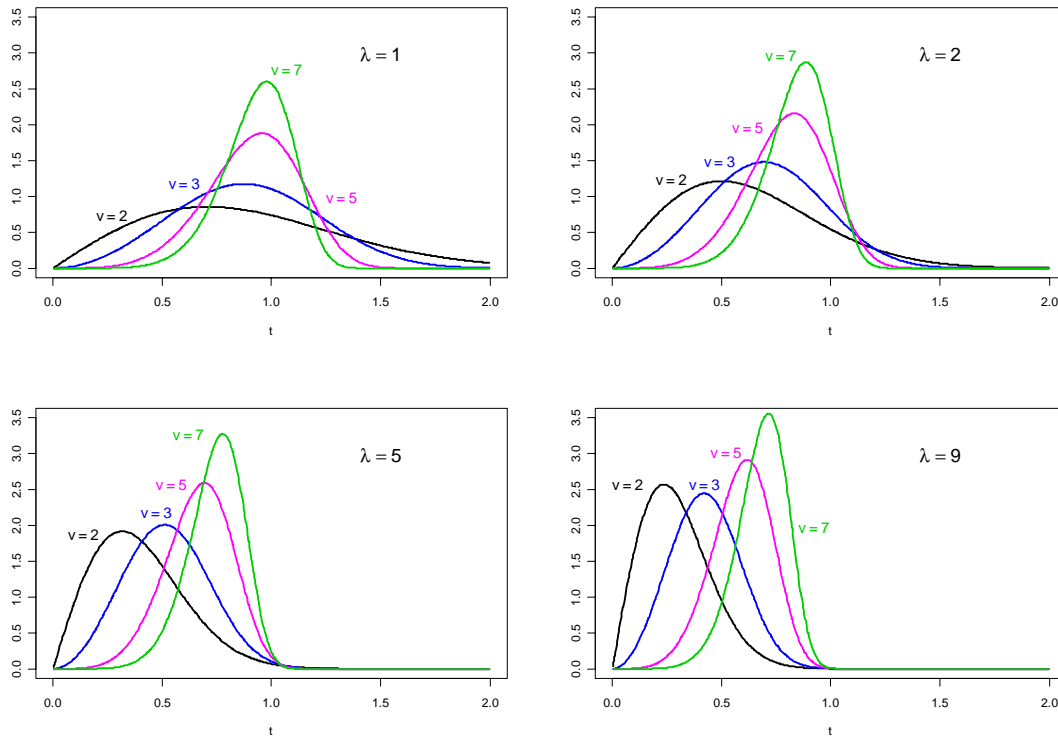


Figure 4.12: density functions of the Weibull distribution

For the simulation studies the same influence of the covariates on the survival time is assumed as in the case of the exponential distributed survival time. Thus the observational time t is defined according to Augustin [5] as

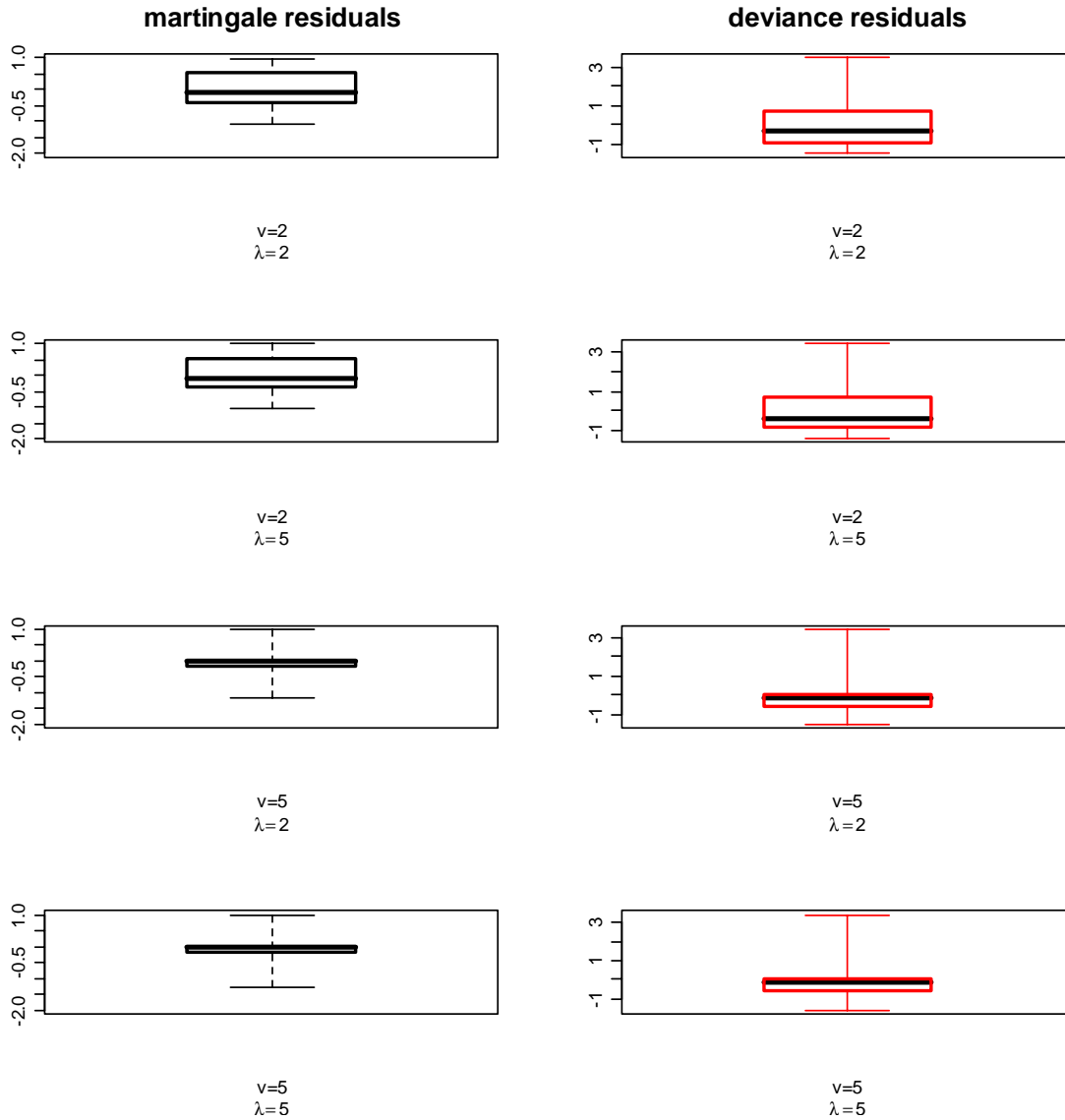


Figure 4.13: Martingale and deviance residuals for several parameters

$$t = \left(-\frac{\log(u)}{\lambda \exp(0.04(\frac{x_1}{10})^2 - 6x_2 + 0.1x_1x_2)} \right)^{\frac{1}{v}} \quad (4.9)$$

with covariates

$$\begin{aligned} x_1 &\sim U[30, 75] & x_4 &\sim U[0, 10] \\ x_2 &\sim Bi(n, 0.5) & x_5 &\sim Bi(n, 0.7) \\ x_3 &\sim N(0, 1) & u &\sim U[0, 1] \end{aligned}$$

and shape parameter $v \in \{2, 5\}$ and scale parameter $\lambda \in \{2, 5\}$, considering the shrinking interquartile ranges for increasing parameter values.

Again the variable 'censoring time' is generated as $ct \sim U[0, t_{m\%}]$ with $t_{m\%}$ being the according quantile of t and $m \in \{0.26, 0.7, 0.89\}$, which yields censoring rates of 25%, 50% and 85%, respectively. The censoring indicator δ_i is defined as

$$\delta_i = \begin{cases} 1 & , \text{ if } t_i \leq ct_i \\ 0 & , \text{ else} \end{cases}$$

Lastly the survival times of all censored objects ($\delta_i = 0$) are substituted by the censoring time ct_i . This final observational time variable was then used for further analysis.

Again 1000 simulation cycles were executed for the different settings, i.e. for varying shape and scale parameters and changing censoring rate and penalizing parameter d . Tables 4.5 and 4.6 show the results of the simulation studies with a Weibull distributed survival time with shape parameter $v = 2$ and $v = 5$, respectively.

The results can be resumed as follows:

- to find a nonlinear relationship moderate to high censoring rates are beneficial, in that case in most of the scenarios more than 90% of the cycles identified at least one knot for variable x_1
- for the identification of interactions between two covariates low to moderate censoring rates are favorable, here the method yielded up to 100% success rate for all scenarios
- the moderate penalizing parameter $d = 3$ gained best results regarding the number of false basis functions, here the fraction reduced to a minimum
- $d = 2$ yields a larger fraction of false basis functions, because too much potential basis functions are allowed to be entered into the model during the stepwise addition process, thus more false basis functions may remain in the final model
- $d = 5$ yields a larger fraction of false basis functions, because

the largest allowed model during the stepwise addition process is forced to remain too small to find all correct functional coherences

- as expected, a higher shape parameter v provided worse results due to the smaller interquartile range of the residuals
- a censoring rate of approximately 50% again yielded best results, as it already has in the case of exponential distributed survival times

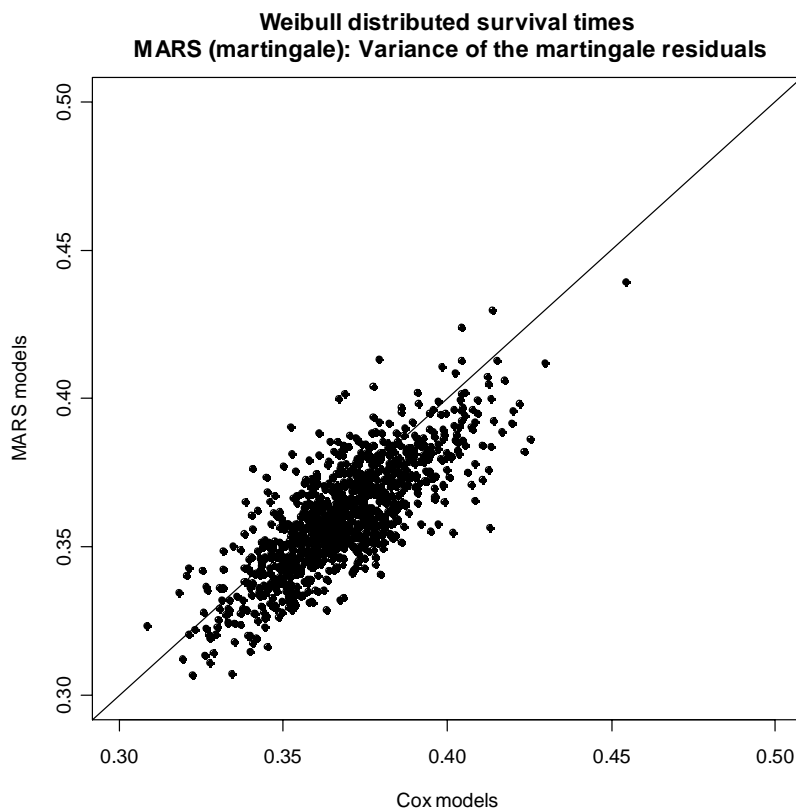


Figure 4.14: Comparison of the variances of MARS with martingale residuals as response and the classical Cox approach with Weibull distributed survival times ($v = 2, \lambda = 5$) and a censoring rate of 50% ($d = 4$)

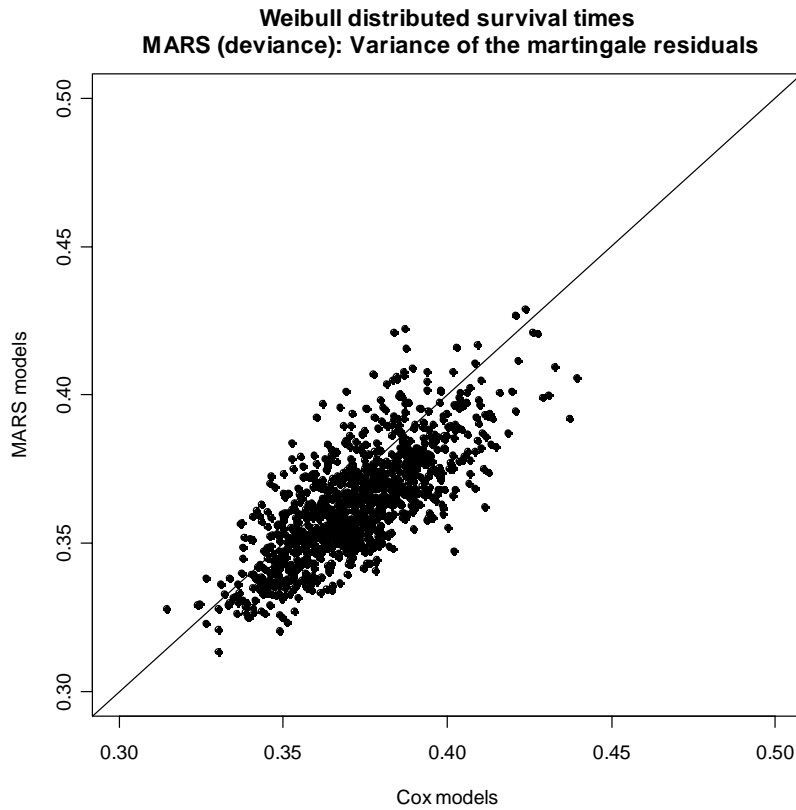


Figure 4.15: Comparison of the variances of MARS with deviance residuals as response and the classical Cox approach with Weibull distributed survival times ($v = 2, \lambda = 5$) and a censoring rate of 50% ($d = 4$)

Again, the variances of the corresponding martingale residuals were compared and two of the numerous scenarios are displayed here:

Figure 4.14 shows the scatterplot of the variances for the scenario with martingale residuals as response, Weibull distributed survival times with shape parameter $v = 2$ and scale parameter $\lambda = 5$, a penalizing parameter $d = 4$ and a censoring rate of 50%. 77.1% of the MARS models yield smaller variances than the classical Cox approach ($p(\text{sign test}) < 0.0001$). The simulations produced a mean

variance for the classical Cox model of 0.368, whereas the MARS approach yields a slightly smaller mean variance of 0.360.

Figure 4.15 shows the corresponding variances for the scenario with deviance residuals as response, Weibull distributed survival times with shape parameter $v = 2$ and scale parameter $\lambda = 5$, a penalizing parameter $d = 4$ and a censoring rate of 50%. Again, most of the items (77.1%) lie under the bisecting line, thus the MARS approach yields smaller variances than Cox, i.e. MARS shows a better performance than Cox ($p(\text{sign test}) < 0.0001$). The classical Cox approach adds up to a mean variance of 0.372, the MARS approach results in a smaller mean variance of 0.363.

Table 4.5: Results of the simulation studies on Weibull distributed survival time data ($v = 2$) in percent

d	λ	censoring rate	martingale residuals			deviance residuals		
			%knots(x_1)	% x_1x_2	%false	%knots(x_1)	% x_1x_2	%false
2	2	25%	69.9	100.0	49.1	96.3	100.0	31.7
		50%	99.6	99.6	27.2	100.0	92.8	30.0
		85%	99.7	80.0	45.6	99.9	84.4	43.7
	5	25%	72.6	100.0	49.5	96.6	100.0	27.9
		50%	99.4	99.4	24.6	99.8	94.8	28.7
		85%	99.8	80.9	45.4	99.5	82.1	43.4
3	2	25%	48.3	100.0	56.5	88.7	100.0	18.3
		50%	97.5	99.4	9.3	99.8	94.9	11.7
		85%	99.6	74.1	32.9	99.0	76.5	30.8
	5	25%	48.2	100.0	54.7	88.2	100.0	17.8
		50%	97.6	99.7	9.8	100.0	95.7	9.1
		85%	98.6	73.9	34.5	99.2	75.9	31.5
4	2	25%	31.8	100.0	69.1	79.6	100.0	23.2
		50%	93.8	99.3	9.1	99.5	94.9	9.5
		85%	97.6	62.6	41.4	97.6	69.4	34.1
	5	25%	29.3	100.0	71.6	78.1	100.0	23.8
		50%	95.1	99.3	7.7	98.4	95.0	9.3
		85%	97.9	65.0	38.1	98.5	66.4	36.7
5	2	25%	17.2	100.0	82.8	63.9	100.0	36.4
		50%	88.3	99.7	12.8	97.5	95.7	8.1
		85%	94.6	52.7	50.6	97.6	55.3	46.5
	5	25%	19.3	100.0	81.2	63.4	100.0	37.2
		50%	89.8	99.7	11.0	97.7	94.7	8.2
		85%	92.5	55.5	47.4	97.7	57.4	44.7

Table 4.6: Results of the simulation studies on Weibull distributed survival time data ($v = 5$) in percent

d	λ	censoring rate	martingale residuals			deviance residuals		
			%knots(x_1)	% x_1x_2	%false	%knots(x_1)	% x_1x_2	%false
2	2	25%	66.7	100.0	53.0	94.2	99.7	33.4
		50%	95.0	92.8	34.4	97.7	79.2	45.3
		85%	95.7	61.4	62.3	95.3	62.4	60.7
	5	25%	66.2	100.0	52.3	93.5	99.3	33.1
		50%	95.3	91.3	38.0	99.0	80.2	43.0
		85%	97.0	60.9	59.4	94.2	61.4	61.9
3	2	25%	47.7	100.0	56.0	84.9	100.0	21.3
		50%	88.6	85.9	29.7	91.6	77.4	34.2
		85%	87.9	42.7	65.6	89.5	46.1	62.0
	5	25%	46.1	100.0	57.7	86.2	99.4	20.3
		50%	89.4	87.1	28.4	95.0	75.5	33.4
		85%	90.7	41.3	66.9	89.3	43.6	64.0
4	2	25%	31.7	100.0	69.5	75.4	99.2	26.8
		50%	77.4	81.0	37.6	89.8	69.1	40.6
		85%	77.9	30.2	74.9	84.1	27.2	77.1
	5	25%	29.9	100.0	71.0	69.4	99.1	32.6
		50%	77.2	82.9	38.9	90.8	70.9	37.8
		85%	83.4	29.7	74.3	84.5	29.4	75.4
5	2	25%	19.6	100.0	80.8	58.0	99.4	43.0
		50%	64.3	73.6	53.8	84.8	59.1	51.5
		85%	74.6	16.4	86.2	73.0	19.8	84.0
	5	25%	19.2	100.0	81.2	58.5	99.5	42.3
		50%	63.8	71.3	56.4	77.7	58.2	56.4
		85%	69.8	17.3	87.0	76.9	18.1	84.6

Chapter 5

Application on real data: prediction of mortality after myocardial infarction

The new method of survival MARS was applied to a cohort of patients who survived an acute myocardial infarction [4]. The data of 2376 patients were collected at the German Heart Center, Munich, and the Klinikum rechts der Isar, Munich.

Results from clinical studies indicate that mortality of survivors of a myocardial infarction can be reduced by implanting a defibrillator [22, 18]. The left ventricular ejection fraction is a common predictor for risk of mortality, more precisely patients with a LVEF $\leq 30\%$ are regarded as high-risk patients and get implanted a defibrillator. However, particularly patients with higher ejection fraction may

have the same risk, i.e. there exist groups of patients with higher risk than expected. Therefore, Bauer et al. [4] established a more precise method to predict mortality in survivors of a myocardial infarction.

This approach determines heart-rate variability, a marker for autonomous responsiveness of the heart. Heart-rate variability is influenced by vagal as well as sympathetic modulation of the sinus node. There is evidence that a decrease of vagal activity increases mortality. Though, conventional measures of heart-rate variability like standard deviation of all normal-to-normal intervals (SDNN) do not distinguish between vagal and sympathetic effects.

Bauer et al. [4] approximately determined vagal and sympathetic effects by separate assessment of deceleration-related and acceleration-related heart-rate variability using 24-h Holter recordings. These recordings were then modified by a signal processing technique of phase-rectified signal averaging, which yields the separate presentation of deceleration and acceleration capacity.

Apart from deceleration (DC) and acceleration capacity (AC), further predictors were investigated, from which the following were used for the analysis with MARS: left ventricular ejection fraction (LVEF), presence of pathological heart-rate turbulence (HRT), presence of diabetes mellitus, SDNN, age and presence of arrhythmias.

Heart-rate turbulence describes the response of the sinus node to a premature ventricular contraction. This quantity is characterised by two numerical parameters, the turbulence onset (TO) and the turbulence slope (TS) [25]. Turbulence onset compares the RR intervals (time duration between two consecutive R waves of the ECG) immediately after with immediately before a premature ventricular contraction, i.e. describes the immediate initial acceleration, measured in percent. The turbulence slope is the maximum slope of a regression line assessed over any sequence of five subsequent sinus-rhythm RR intervals within the first 20 sinus-rhythm RR intervals after a premature ventricular contraction, i.e. the speed of subsequent deceleration, measured in ms per RR interval. A TS of less than 2.5 is reckoned pathological, as well as a TO greater than 0. Are both parameters pathological, the heart-rate turbulence is also defined as abnormal.

Arrhythmia is present, if a patient suffers at least 10 premature ventricular contractions per hour or at least one non-sustained ventricular tachycardia in 24 hours [25].

Bauer et al. [4] figured out deceleration capacity to be the covariate with the highest predictive power among all examined variables. Thus, the analysis in this thesis shall pay special attention to that quantity.

5.1 Data preparation

Only patients with no missing data were included into the analysis. Furthermore, just patients with a left ventricular ejection fraction of more than 30% were to be investigated.

Another cutback of the data resulted from omitting 2.5% of the data with the smallest and largest values of the deceleration capacity, respectively. This was done to factor out extreme values, which are easy to recognize in the left panel of figure 5.1. The right panel shows the histogram of deceleration capacity after trimming the data.

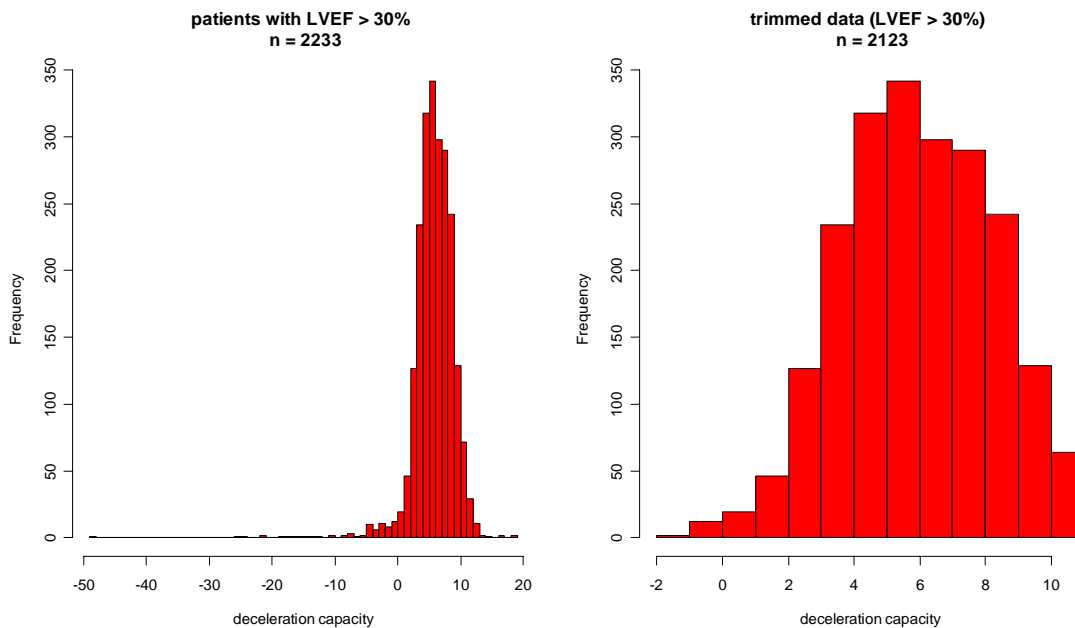


Figure 5.1: Histogram of deceleration capacity for all data and trimmed by extreme values of DC

The basic characteristics of the remaining 2123 patients are summarized in table 5.1.

Table 5.1: Main characteristics of the patients after myocardial infarction

Variable	Mean (SD)	Number (%)
GENDER (male)		1699 (80.0)
DC	5.92 (2.27)	
AC	-7.15 (2.62)	
LVEF	54.78 (10.71)	
HRT (yes)		144 (6.8)
DIABETES (yes)		371 (17.5)
SDNN	98.75 (34.14)	
AGE	58.14 (10.28)	
ARRHYTHMIA (yes)		338 (15.9)

The simulation studies did not exactly clarify whether the usage of martingale or deviance residuals of a Cox model yields better results. Hence, the whole analysis of the infarction data was done with both types of residuals as response in a common MARS approach.

5.2 The classical Cox PH method

Before starting with the new approach a conventional analysis with a Cox regression shall be performed. For these purposes the covariates DC, AC, LVEF, HRT, DIABETES, SDNN, AGE and ARRHYTHMIA were included in a Cox PH model. Table 5.2 shows

the resulting model for these eight covariates.

Table 5.2: Cox PH model of the infarction data

Covariate	Coefficient	SE(coeff)	p-value
DC	-0.104	0.044	0.017
AC	-0.017	0.030	0.570
LVEF	-0.025	0.008	< 0.001
HRT	0.787	0.190	< 0.001
DIABETES	0.508	0.163	0.002
SDNN	-0.004	0.003	0.160
AGE	0.048	0.010	< 0.001
ARRHYTHMIA	0.123	0.181	0.49
<i>log-likelihood = -1244.183</i>			

Three of the covariates (AC, SDNN and ARRHYTHMIA) do not show significant effects concerning the Wald statistic, thus a reduced model with only five prognostic factors was performed. Table 5.3 summarizes this final Cox model of the infarction data.

Table 5.3: The final Cox PH model of the infarction data

Covariate	Coefficient	SE(coeff)	p-value
DC	-0.131	0.040	< 0.001
LVEF	-0.027	0.017	< 0.001
HRT	0.800	0.188	< 0.001
DIABETES	0.524	0.162	< 0.001
AGE	0.047	0.009	< 0.001
<i>log-likelihood = -1245.525</i>			

So, the five covariates deceleration capacity, left ventricular ejection fraction, heart rate turbulence, diabetes and age were identified as influential on survival time and the risk of death by the conventional Cox PH approach.

Higher values of deceleration capacity and LVEF result in a reduced mortality rate, whereas risk of death increases with the presence of a heart rate turbulence and diabetes. Older patients also show a higher risk.

Now survival analysis using MARS is to be checked with the present data on myocardial infarction.

5.3 A "snapshot"

A first usage of the new approach with martingale and deviance residuals of the Cox null model as response in a conventional MARS model was done by simultaneously including all eight above mentioned covariates DC, AC, AGE, LVEF, HRT, SDNN, DIABETES and ARRHYTHMIA in a multivariate model. The resulting MARS models with 10 and 11 included basis functions, respectively, are shown in tables 5.4 and 5.5. As a measure for goodness-of-fit, the generalized cross validation criterion is specified at the bottom of the tables. Yet, the GCVs of martingale and deviance model are

not directly comparable, due to the differing ranges of the residuals. Deviance residuals range from $-\infty$ to ∞ , martingale residuals from $-\infty$ to 1, thus the deviance model's residual sum of squares yields larger values than the one from martingale residuals.

Table 5.4: Multivariate MARS model with martingale residuals of the Cox null model as response

Predictor	Coefficient	SE
AGE	0.009	0.002
LVEF	-0.002	0.001
DIABETES	0.033	0.017
SDNN	-0.007	0.001
HRT	0.010	0.030
DC	0.054	0.017
$(\text{SDNN} - 63)_+$	0.007	0.001
$(\text{AGE} - 69.8)_+$	0.024	0.007
AGE·DC	-0.001	0.0003
DIABETES·HRT	0.161	0.054

$$GCV = 0.08166$$

The two models show broad accordance in the choice of predictors:

- Both models detect AGE, LVEF, DIABETES, SDNN and HRT as influential on survival. The coefficients show identical signs in the two models, i.e. the direction of influence is clear-cut.
- Both models found AGE being nonlinear with the same knot $t_{AGE} = 69.8$ and identical signs of the coefficients.

Table 5.5: Multivariate MARS model with deviance residuals of the Cox null model as response

Predictor	Coefficient	SE
AGE	0.005	0.002
LVEF	-0.006	0.001
DIABETES	0.084	0.043
SDNN	-0.025	0.006
HRT	0.308	0.072
AC	-0.024	0.008
$(\mathbf{AGE} - 69.8)_+$	0.076	0.017
DIABETES·HRT	0.455	0.133
$(\mathbf{AC} + 3.41)_+$	-0.529	0.145
$(\mathbf{AC} + 5.43)_+$	0.165	0.038
$(\mathbf{SDNN} - 54)_+$	0.025	0.006

$$GCV = 0.49242$$

- SDNN is also detected nonlinear in both models, whereas the knots differ ($t_{SDNN} = 63$ in the martingale model, $t_{SDNN} = 54$ in the deviance model). Though, the direction of influence is equal.
- The interaction between DIABETES and HRT was found in both models, too, again with identical signs of the coefficients.

However, some discrepancies do exist:

- In the martingale model DC is detected as influential, whereas in the deviance model this covariate does not appear, while here AC was included into the model as a nonlinear effect on survival.
- The martingale model included a further interaction, namely between AGE and DC.

Recapulatory the martingale model yields an increase in risk of death with increasing age and DC and presence of diabetes as well as heart rate turbulence. The risk decreases with increasing LVEF. Figure 5.2 shows the basis functions of the two nonlinear covariates AGE and SDNN in the martingale model. The risk of death moderately increases until the age of 70, for older patients the risk increases even more. On the contrary, an increase of SDNN until 63 heavily decreases risk of death, with higher values of SDNN the risk slowly increases again.

The deviance model also yields an increasing risk of death with increasing age and the presence of diabetes and heart rate turbulence, whereas the risk again decreases with increasing LVEF.

Figure 5.3 and 5.4 show the nonlinear basis functions of AGE, SDNN and AC. Risk increases with increasing age and increases even more

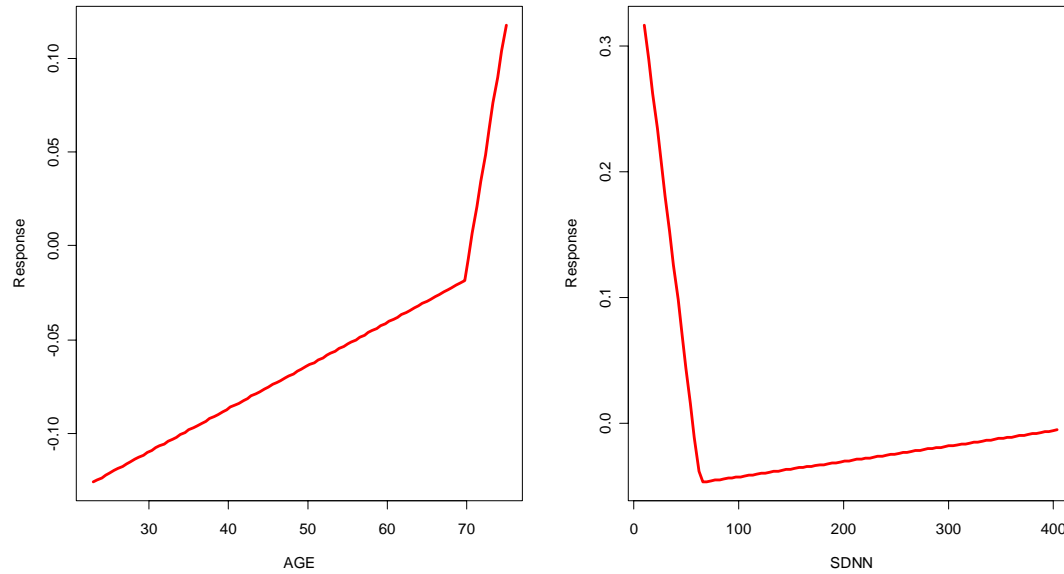


Figure 5.2: Basis functions of SDNN and AGE in the martingale model

for patients at the age of 69.8 or older. An increase of SDNN decreases risk rapidly until a value of 54, for higher values of SDNN the risk still moderately decreases. The acceleration capacity shows a rather strange functional form. Firstly the risk decreases slightly until a value of -5.4, for values between -5.4 and -3.4 the risk heavily spurts and afterwards it decreases again.

By reason of the differing results between deviance and martingale model, the fact that the deceleration capacity was not included into the deviance model and the peculiar functional form of the acceleration capacity in the deviance model a stepwise analysis was performed.

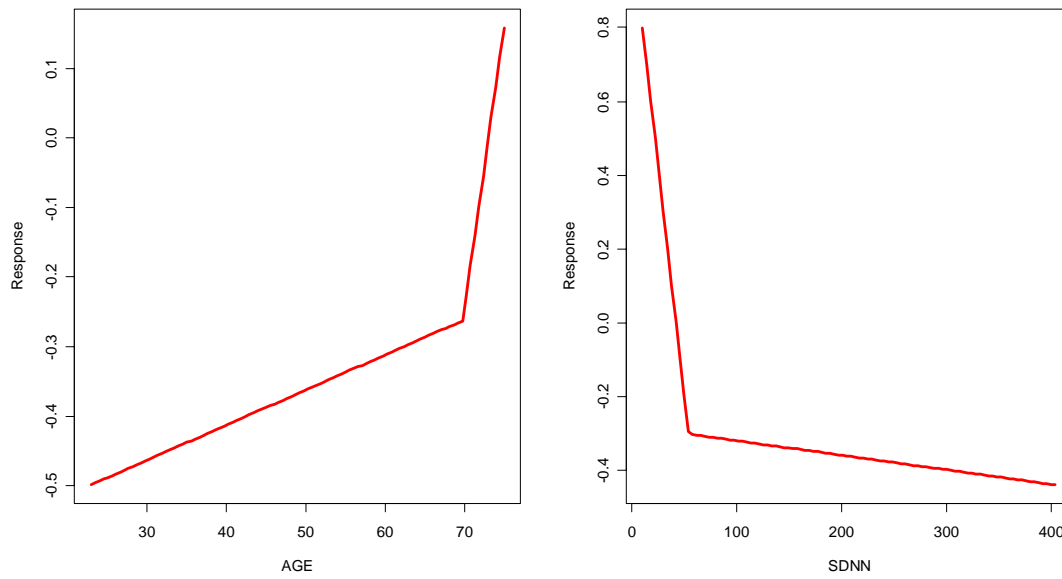


Figure 5.3: Basis functions of SDNN and AGE in the deviance model

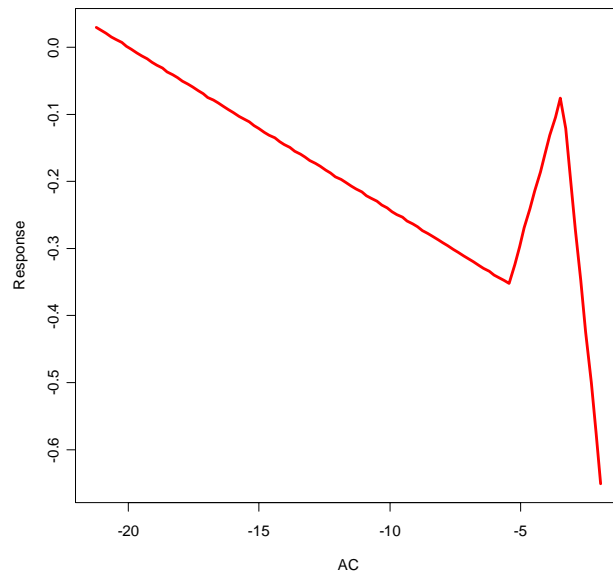


Figure 5.4: Basis function of AC in the deviance model

5.4 Stepwise analysis

The stepwise computation began with all possible univariate MARS models, i.e. eight MARS models with martingale and deviance residuals of the Cox model without any covariates as response, respectively, were calculated, using AC, DC, AGE, LVEF, HRT, DIABETES, SDNN or ARRHYTHMIA as single covariate.

The model with the smallest generalized cross validation criterion, the goodness-of-fit criterion within the MARS approach, was selected. Subsequently, a Cox model with just this covariate was calculated to assess the according residuals. These residuals were then used in seven new univariate MARS models to search for the next most influential covariate.

This procedure was done until no further MARS model was found or all covariates were incorporated.

Tables 5.6 and 5.7 illustrate the sequence of found basis functions in the order of inclusion.

These two sequences show almost absolute accordance, except for the variable SDNN, which has one more knot in the martingale model (besides the knots differ in value), and DIABETES, which is only included in the deviance model.

Figure 5.5 shows the basis functions found by the stepwise procedures for both scenarios. The left panels illustrate the basis func-

Table 5.6: Stepwise analysis with martingale residuals as response

step	basis function	cefficient	GCV
1	DC	-0.062	0.084
	$(\mathbf{DC} - 5.43)_+$	0.059	
2	HRT	0.164	0.089
3	AGE	0.001	0.091
	$(\mathbf{AGE} - 69.8)_+$	0.031	
4	SDNN	0.016	0.090
	$(\mathbf{SDNN} - 36)_+$	-0.034	
	$(\mathbf{SDNN} - 54)_+$	0.018	
5	LVEF	-0.002	0.088
<i>no further models found</i>			

Table 5.7: Stepwise analysis with deviance residuals as response

step	basis function	cefficient	GCV
1	DC	-0.515	0.510
	$(\mathbf{DC} - 5.43)_+$	0.145	
2	HRT	0.411	0.495
3	AGE	0.003	0.489
	$(\mathbf{AGE} - 69.8)_+$	0.074	
4	SDNN	-0.018	0.479
	$(\mathbf{SDNN} - 63)_+$	0.018	
5	LVEF	-0.006	0.474
6	DIABETES	0.113	0.472
<i>no further models found</i>			

tions of the martingale approach, the right panels show the basis functions of the deviance approach.

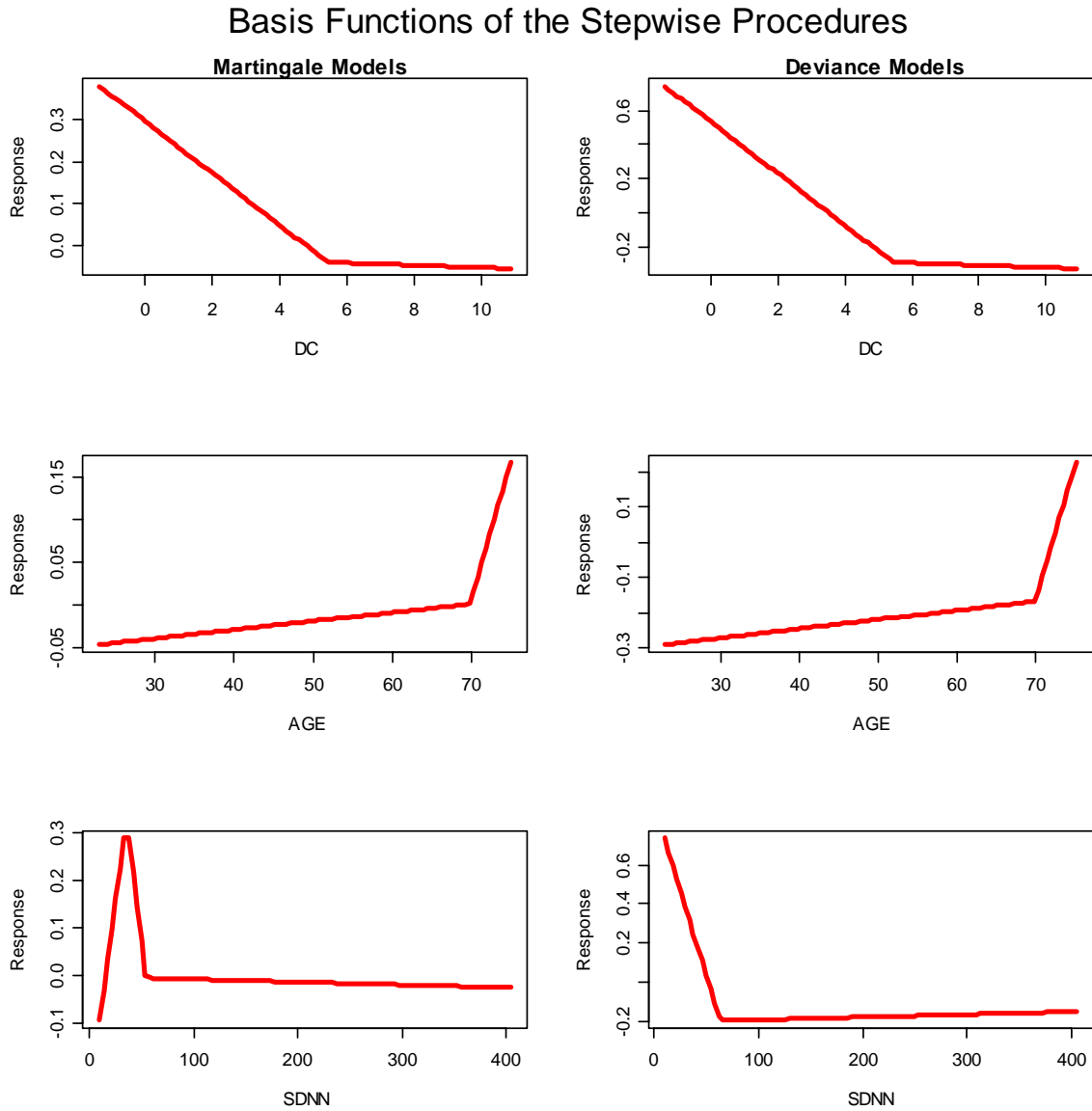


Figure 5.5: Basis functions of the stepwise procedures for the martingale and deviance scenario

The stepwise proceeding does of course refuse the consideration of interactions between the covariates, hence the above detected influential covariates are finally all at once built in a multivariate MARS model. This proceeding represents the counterpart to a classical regression procedure with an initial stepwise modelling, followed by the multivariate analysis, in which only the significant covariates of the stepwise model are included. This multivariate procedure was again done with martingale and deviance residuals of the Cox null model as response, respectively. Tables 5.8 and 5.9 show the findings.

Table 5.8: Multivariate MARS model with the covariates found in the stepwise procedure with martingale residuals of the Cox null model as response

Predictor	Coefficient	SE
HRT	0.147	0.026
DC	-0.033	0.007
$(\mathbf{DC} - 5.43)_+$	0.041	0.010
AGE	0.002	0.001
$(\mathbf{AGE} - 69.8)_+$	0.032	0.007
LVEF	-0.002	0.001
SDNN	-0.010	0.002
$(\mathbf{SDNN} - 54)_+$	0.010	0.002

$$GCV = 0.08172$$

The multivariate deviance model has included an interaction between HRT and DIABETES which was already found in the snap-

Table 5.9: Multivariate MARS model with the covariates found in the stepwise procedure with deviance residuals of the Cox null model as response

Predictor	Coefficient	SE
HRT	0.267	0.074
DC	-0.091	0.016
$(\mathbf{DC} - 5.43)_+$	0.111	0.024
AGE	0.004	0.002
$(\mathbf{AGE} - 69.8)_+$	0.070	0.017
LVEF	-0.006	0.002
DIABETES	0.086	0.043
HRT·DIABETES	0.461	0.133
$(\mathbf{LVEF} - 78)_+$	0.171	0.059

$$GCV = 0.49137$$

shot analysis, and detected the influence of LVEF as nonlinear. Instead SDNN does not appear in the deviance model. Knots for DC and AGE are again identical, which suggests that these functional forms are quite certain. Also the variables HRT and LVEF can be supposed to be influential on mortality.

Figure 5.6 shows the different basis functions of the nonlinear covariates of the two multivariate models arising from the stepwise procedure. The left panels again illustrate the basis functions of the martingale approach, the left panels show the basis functions of the deviance approach.

Basis Functions of the Multivariate Models after Stepwise Procedure

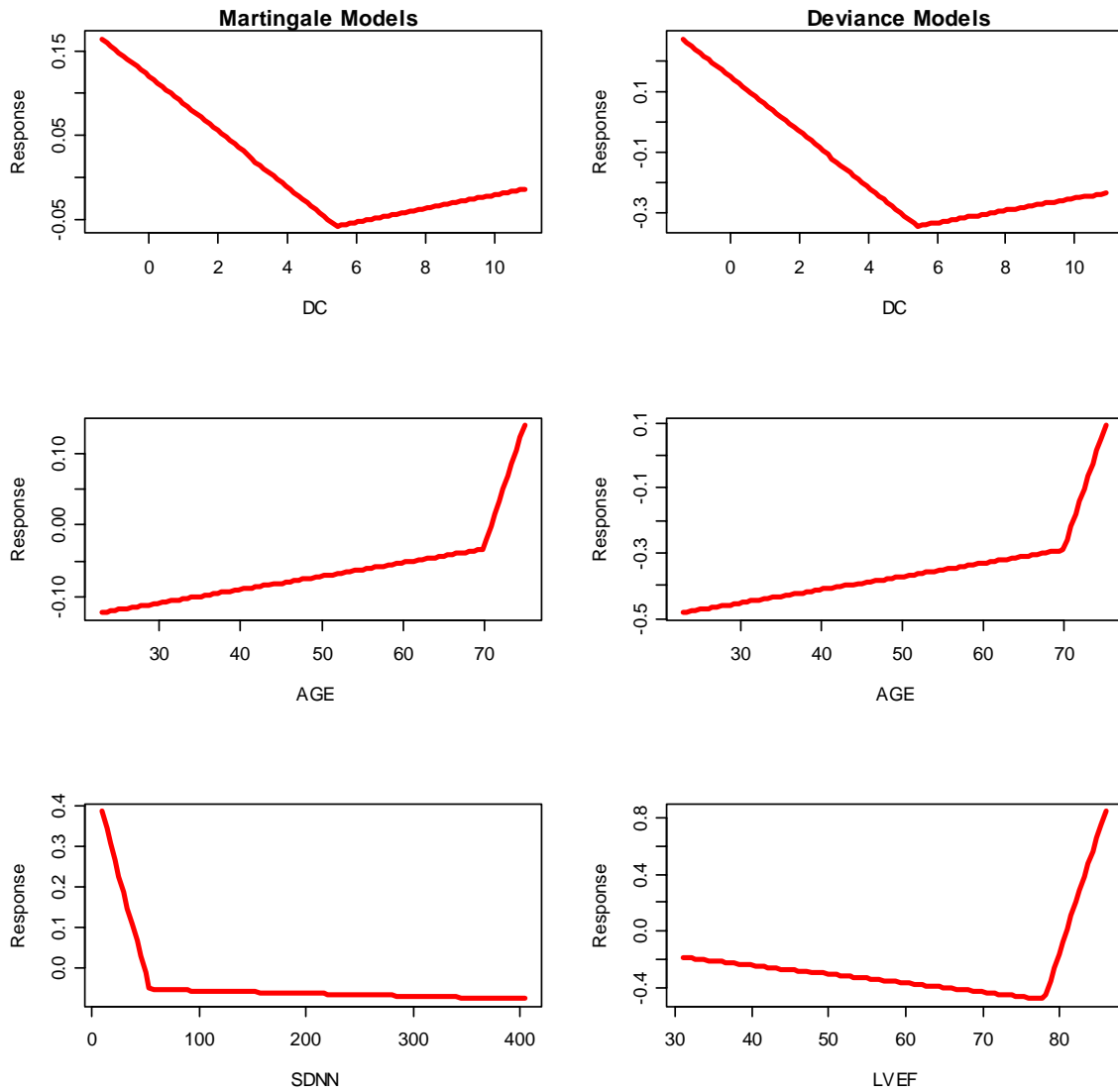


Figure 5.6: Basis functions of the nonlinear covariates of the multivariate models after the stepwise procedures (the two scenarios differ in the bottom panels: the martingale model included SDNN, the deviance model included LVEF)

5.5 Final Results

Comparing now the findings of the snapshot and stepwise analyses, the following conclusions can be made about the nature of coherence between mortality and the observed covariates:

- Age presents a nonlinear functional form, whereas increasing age results in a higher risk of death. The risk increases enormously for $AGE > 69.8$.
- The presence of a pathological heart rate turbulence increases risk of mortality.
- The left ventricular ejection fraction has an effect on mortality, too. The higher the value of LVEF the lower mortality.
- The standard deviation of all normal-to-normal-intervals (SDNN) can also be regarded as influential on risk of death. Risk decreases until a value between 54 and 63, for higher values risk seems to settle down at a constant level.
- The presence of diabetes mellitus increases the mortality rate.
- Deceleration capacity seems to be a predictor for mortality after myocardial infarction. The higher DC the lower the risk of death. For values greater than 5.43 the risk approximately levels off.

- There is also evidence for an interaction between diabetes mellitus and heart rate turbulence. A positive sign indicates increasing risk for patients who feature at least one of the symptoms.

The four final models - snapshot and multivariate after stepwise with martingale and deviance residuals, respectively - differ marginally in the included covariates and the found knots.

However, the goodness-of-fit criterion GCV indicates no tremendous differences between snapshot and multivariate after stepwise models for both scenarios.

Finally, all the identified basis functions (linear, with knot and interaction) were included as covariates into Cox models. Some covariates were not significant and therefore excluded in a second cycle. Interestingly no interaction found during the MARS modeling was significant in the following Cox analysis. Table 5.10 summarizes the significant coefficients along with the corresponding standard errors of the five final Cox models: conventional Cox from section 5.2, snapshot with martingale and deviance residuals and multivariate after stepwise with martingale and deviance residuals.

The log-likelihoods of these Cox models are presented in table 5.11. One degree of freedom was assessed for every coefficient as well as for every knot. The likelihood-ratio tests yield no significant

differences between the five models, i.e. the MARS models do not offer a higher goodness-of-fit as the classical Cox approach.

Table 5.10: Coefficients and standard error of the 5 Cox models (conventional Cox and Cox after MARS models)

Variable	Cox	snapshot		multivariate after stepwise	
		martingale	deviance	martingale	deviance
DC	-0.131 (0.04)	-	-	-0.219 (0.06)	-0.249 (0.05)
AC	-	-	-0.091 (0.03)	-	-
AGE	0.047 (0.01)	0.041 (0.01)	0.040 (0.01)	0.039 (0.01)	0.049 (0.01)
LVEF	-0.027 (0.02)	-0.030 (0.01)	-0.028 (0.01)	-0.026 (0.01)	-0.028 (0.01)
HRT	0.800 (0.19)	0.863 (0.18)	0.760 (0.19)	0.756 (0.19)	0.756 (0.19)
DIABETES	0.524 (0.16)	0.479 (0.16)	0.554 (0.16)	-	0.538 (0.16)
SDNN	-	-0.040 (0.01)	-0.033 (0.01)	-0.040 (0.01)	-
(DC - 5.43)₊	-	-	-	0.287 (0.11)	0.306 (0.10)
(AC + 5.43)₊	-	-	0.529 (0.15)	-	-
(AC + 3.41)₊	-	-	-1.274 (0.58)	-	-
(AGE - 69.8)₊	-	0.168 (0.06)	0.155 (0.06)	0.148 (0.06)	-
(LVEF - 78)₊	-	-	-	-	0.406 (0.13)
(SDNN - 54)₊	-	-	0.030 (0.01)	0.037 (0.01)	-
(SDNN - 63)₊	-	0.039 (0.01)	-	-	-

Figure 5.7 shows the Kaplan-Meier curves of all five final models with 95% confidence bands. The curves are estimated by centering the covariates by their means. Especially at the beginning of the observational time the five models proceed concordantly and feature high concordance even later on.

Table 5.11: Log-likelihoods of the four final models and the conventional Cox approach

	martingale (df)	deviance (df)
snapshot	-1238.868 (9)	-1236.921 (14)
multivariate after stepwise	-1238.102 (11)	-1238.790 (9)
Cox	-1245.525 (5)	

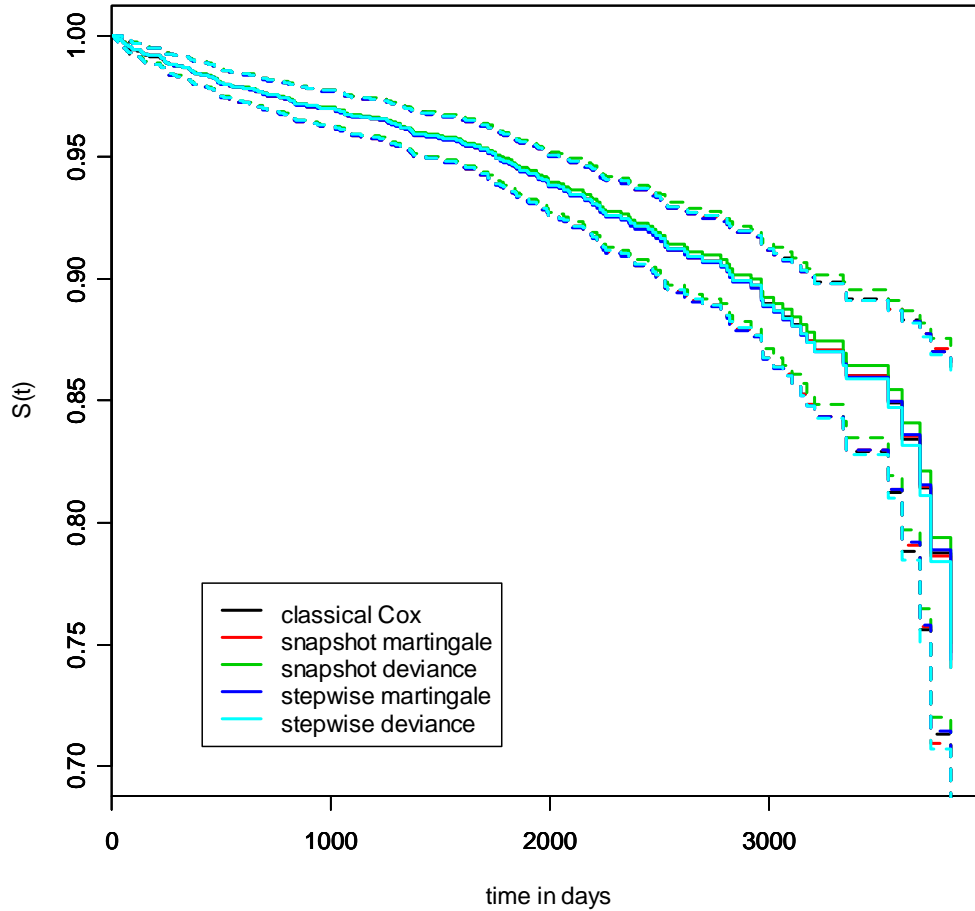


Figure 5.7: Kaplan-Meier curves of all five final models with 95% confidence bands

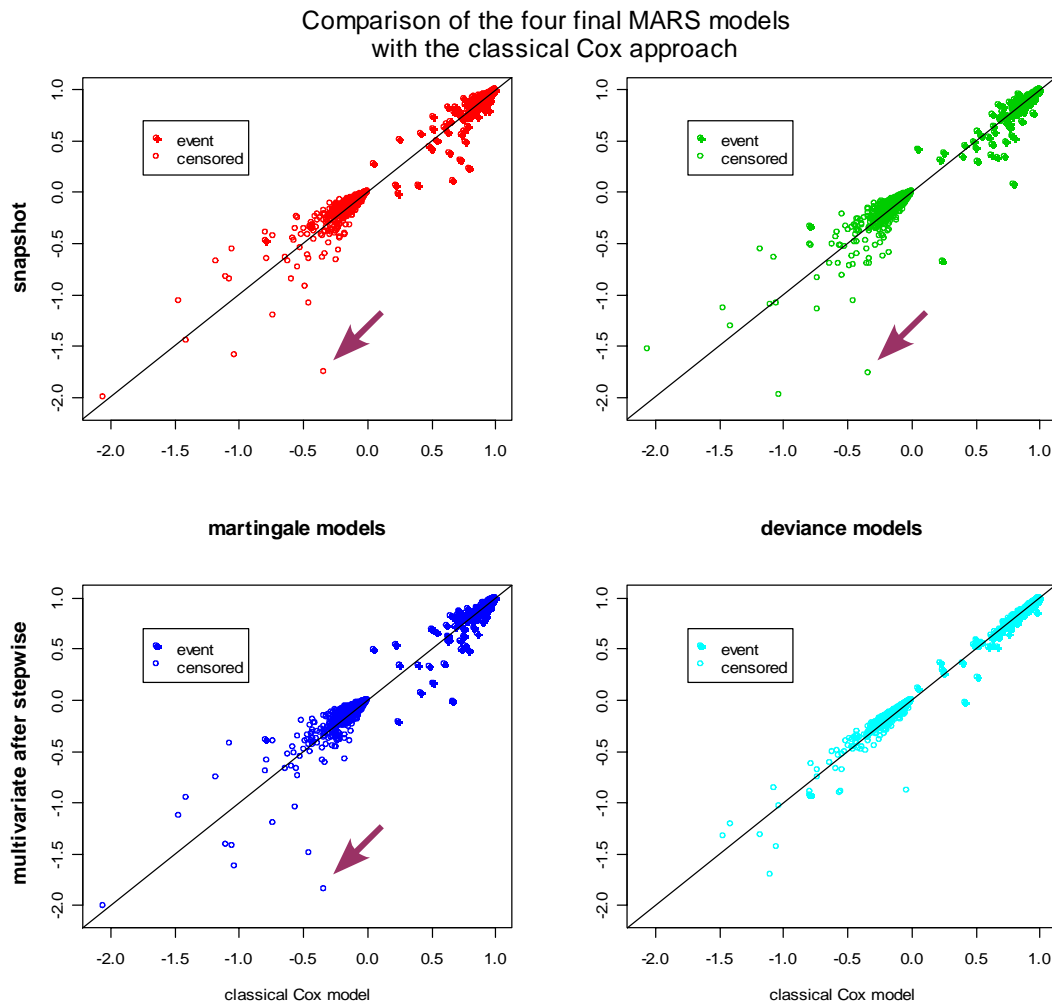


Figure 5.8: Comparison of the snapshot and stepwise models with the classical Cox model via martingale residuals

Figure 5.8 displays a comparison of the four final MARS models with the classical Cox model. For that purpose the martingale residuals of the corresponding Cox models were plotted against each other, labeled by the survival status. The top panels show the snapshot models, the bottom panels show the multivariate after stepwise models. The left panels present the martingale models, the right

panels present the deviance models. All MARS models show high concordance with the classical Cox model, except for some extreme deviations.

Table 5.12 summarizes the variances of the martingale residuals of the five models. A decrease in variance of the MARS models compared to the classical Cox model would suggest a better performance of the new method. However, all approaches exhibit comparable variances. Thus, a difference in performance and fit cannot be detected between the several approaches.

Table 5.12: Variances of the martingale residuals of the five final models

Approach	Variance
classical Cox	0.0898
snapshot martingale	0.0902
snapshot deviance	0.0902
stepwise martingale	0.0913
stepwise deviance	0.0921

The arrows in figure 5.8 mark one special patient (censored), who features enormous departure from the bisecting line in three of the four models (snapshot martingale, snapshot deviance and stepwise martingale). The classical Cox model estimates the low risk of that person quite well (martingale residual of -0.34), but the three MARS models predicted him a quite large risk of death. The data of this person was now investigated in detail. In doing so it turned out that this man exhibits an SDNN value of 10, the smallest SDNN value

of the whole sample. In fact this value lies more than 1.5 times the interquartile range apart from the median of the sample, and thus can be regarded as outlier.

SDNN is not a predictor in the classical Cox model, but is included in the three MARS models snapshot martingale, snapshot deviance and multivariate after stepwise martingale with a negative coefficient, respectively. Thus, the MARS models predict him very high risk, although he did not die. For SDNN was not included as covariate in the multivariate after stepwise model with deviance residuals, its prediction for that special person is much better.

The multivariate after stepwise model with deviance residuals as response seems to coincide best with the classical Cox approach. The reason for that lies in the high concordance of their included covariates. Both models included DC, AGE, LVEF, HRT and DIABETES as predictors. The MARS approach just additionally includes a knot for DC and LVEF, respectively.

As a result of the observances about SDNN the three corresponding MARS models were again calculated on the data set without this particular patient with the smallest value of SDNN, to investigate whether the MARS approach is affected by outliers. However, the analyses on the restricted data generated the same models as on the whole data set, only the knot of the variable SDNN shifted to the value of 55, which lies between the former knots. In the re-

Table 5.13: Coefficients and standard error of the 3 MARS models on the restricted data

	snapshot		stepwise
Variable	martingale	deviance	martingale
DC	-	-	-0.102 (0.04)
AC	-	-0.088 (0.03)	-
AGE	0.043 (0.01)	0.042 (0.01)	0.037 (0.01)
LVEF	-0.031 (0.01)	-0.028 (0.01)	-0.028 (0.01)
HRT	0.864 (0.18)	0.857 (0.18)	0.771 (0.19)
DIABETES	0.441 (0.17)	0.508 (0.17)	-
SDNN	-0.059 (0.01)	-0.049 (0.01)	-0.058 (0.01)
$(\mathbf{AC} + 5.43)_+$	-	0.506 (0.15)	-
$(\mathbf{AC} + 3.41)_+$	-	-1.452 (0.60)	-
$(\mathbf{AGE} - 69.8)_+$	0.161 (0.06)	0.156 (0.06)	0.168 (0.06)
$(\mathbf{SDNN} - 55)_+$	0.058 (0.01)	0.047 (0.02)	0.058 (0.01)

stricted snapshot deviance model the deceleration capacity is merely included as linear predictor.

Again Cox models with the found basis functions as covariates were calculated and the corresponding coefficients and their standard errors are summarized in table 5.13.

All coefficients keep the same size as in the former models and the signs remain equal.

Thus, the MARS approach do not seem to be notably susceptible to outliers, the method appears rather stable.

Recapulatory, there is no evidence that the new developed MARS approach yield better results than the classical Cox proceeding. The additional inclusion of nonlinear effects yields no protruding improvement of the outcome. The reason for that can be, that there are in fact no nonlinear relationships and interactions between different covariates in these data.

The four different MARS approaches - snapshot and multivariate after stepwise each with martingale and deviance residuals, respectively - can be assumed equal, the differences are marginal and obviously just express in negligible effects.

Three main predictors were identified by all five approaches: heart rate turbulence, age and left ventricular ejection fraction. These covariates seem to have an influence on the mortality rate for granted. Deceleration capacity, diabetes and SDNN were identified as predictors by at least three models, thus some influence can be assumed for them, too.

As snapshot and stepwise procedure do not distinctively differ for these data, it can be noticed that the multivariate approach is rather robust and hardly susceptible for noise in the data.

The whole analyses in this thesis could not reveal essential differences between the two considered approaches with martingale and deviance residuals. The application on the infarction data shows that the proposal with deviance residuals is as applicable as the one with martingale residuals.

Chapter 6

Conclusions

6.1 Summary

The main focus of medical research often lies on the identification of new prognostic factors. To find the correct form of influence, i.e. the right functional relationship between response and predictor, is a challenge for statisticians. Especially in the environment of survival time data the bandwidth of available procedures is quite small. Typically the modelling of nonlinear effects and interactions is rather difficult with the established methods.

The new approach using residuals of a Cox null model as response in a MARS setting provides a useful tool to deal with the aforementioned problems. Linear as well as nonlinear effects of the covariates and even interactions are dependably identified adapted to the available data.

The thesis gives an overview of the underlying techniques and works out the details of the new approach. Simulation studies yield circumstantial evidence of the power of this method. High percentages of the simulated relationships were identified by MARS.

Simulation studies on the general performance of the method yield convincing results. Nonlinear effects of covariates were found at a high percentage, the stronger, i.e. the larger, the simulated effect the more dependably the approach works. Interactions between two continuous covariates were found almost ever, except for a small effect paired with a large error variance. The penalizing parameter d in the generalized cross validation criterion also effects the performance of the approach. The larger the value, i.e. the smaller the model is allowed to grow, the more effects are not identified, thus the poorer the final model.

By means of interaction studies it was checked whether MARS is able to detect an interaction between a continuous and a binary covariate. As measure of the goodness-of-fit the statistics $R^2 = \frac{SSE}{SST}$ was calculated and summarized as mean R^2 . All simulation studies in case of interactions yielded at least a mean R^2 of 0.94 (range(R^2): 0.940 - 0.999), i.e. simulated interactions between a continuous and a binary covariate are steadily detected by the MARS approach, even for a very large error variance of 50.

At last, the simulation studies on survival type data yield compara-

ble results for exponential and Weibull distributed survival times, respectively. Nonlinear effects on the response were particularly found for rather large censoring rates, whereas the recognition of interactions benefits from rather low censoring rates. A moderate censoring rate of 50% results in the smallest percentage of false basis functions. Besides, a larger penalizing parameter d again yields worse models in terms of the number of spuriously chosen basis functions. This is obvious because a larger parameter causes a stronger penalization which in turn leads to models with less basis functions, i.e. the probability of not detected but existing functional coherences increases with increasing values of d .

The comparison of martingale and deviance residuals of a Cox model without any covariates as response in a common MARS approach does not yield protruding differences in power and performance, so the conclusion of this thesis must be, that the user himself has to choose among these two possibilities.

The calculation of a Cox PH model with the basis functions of MARS as covariates was done to investigate the goodness-of-fit of the new approach. The resulting martingale residuals were compared with the residuals from the common Cox model of the original covariates. This comparison reveals a significantly smaller variation of the residuals in the case of the survival MARS approach. Survival MARS can therefore be regarded as superior to the established

method of Cox, when nonlinear effects and interactions do exist in the data.

Considering the results of the whole simulation analysis, one can conclude that the new developed MARS approach with residuals of a Cox model as response is a reasonable method to discover effects in a survival time setting, for linear as well as nonlinear effects and also interactions are certainly detected in the data.

The analysis of the data on survivors of a myocardial infarction does not reveal any advantages of the survival MARS approach. All possible proceedings - classical Cox, multivariate and stepwise MARS - result in comparable models concerning the log-likelihoods of the corresponding Cox PH models, thus can be treated as equal. The main predictors for mortality are clearly detected, these basis functions (i.e. variables) wherein the models do disagree can be regarded as marginal. The variances of the corresponding Cox model residuals are comparable as well, the proceeding with the new approach does not increase the goodness-of-fit. Possibly there do in fact not exist any nonlinear effects and interactions in the data and if so, MARS can not perform better than the classical approach.

Recapitulatory, survival MARS presents a feasible and powerful alternative to the common method Cox PH and is even superior to Cox in the case of nonlinear effects and interactions, as MARS pro-

vides simple data-driven and hands-off modelling of these types of functional forms.

6.2 Outlook

The new MARS approach is a powerful way to detect prognostic factors. Yet, the simulation studies show uncertainties in the choice of basis functions. Thus, a mode to stabilize the procedure may be the involvement of bootstrapping, i.e. selection of the final model among a range of models calculated on bootstrap samples of the original data. By this means the final model may be freed from nuisance effects and unimportant covariates.

The whole thesis dealt with situations with predictors which do not depend on observational time. Hence, another extension of the approach ought to be the application on time-dependent covariates, i.e. on situations where the assumption of proportional hazards is not fulfilled. The exploration of the behaviour of MARS in this scenario could be a further step into the direction of more flexible modelling.

So far, the available software used in this thesis only allows for interactions of at most second order. A further possibility of continuation of the work with MARS may also be to expand the syntax

so that interactions of higher order are possible.

Appendix A

Notation

A partial listing of symbols and other notation used frequently in this thesis:

T, t	time variable
C	censoring time
$I(x \leq k)$	indicator function, = 1 if $x \leq k$
δ_i	status variable, = 1 if individual i had an event
$Pr(X)$	probability of X
$N_i(t)$	counting process: number of events in the time interval $[0, t]$ for subject i ($N_i(t) \in \{0, 1\}$ in this thesis)
$Y_i(t)$	= 1 if subject i is under risk at time t
\mathcal{F}_t	filtration or history

$S(t)$	survivor function
$F(t)$	distribution function
$\lambda(t)$	hazard function
$\Lambda(t)$	cumulative hazard function
$\hat{\Lambda}_0(t)$	Breslow estimator of the cumulative baseline hazard
$\hat{\Lambda}(t)$	Nelson-Aalen estimator for the cumulative hazard function
$d\bar{N}(t)$	number of events occurring exactly at time t
$dN_i(t)$	increment in N_i in an arbitrary small time interval $[t, t + dt]$
$\mathcal{E}(X Y)$	expectation of X conditional on Y
$L(\beta)$	partial likelihood of the Cox PH model
$(x - t_k)_+$	hockeystick function in MARS, $= (x - t_k)$ if $x \geq t_k$
t_k	knot, in the range of the observed values of X
B	spline basis function in MARS
d	penalizing parameter in the generalized crossvalidation criterion
$M_i(t)$	martingale residuals of subject i at time t
v	shape parameter of the Weibull distribution
λ	scale parameter of the Weibull distribution

Appendix B

R functions

B.1 Simulations on a continuous response

This function was used to perform the simulation studies described in section 4.1.

```
"marssim" <- function(simsize=10,noobs=500,sigma=0.1,scenario=1,gcv=4)
{
  predicted1 <- 0
  predicted2 <- 0
  predicted3 <- 0
  predicted4 <- 0
  predicted5 <- 0
  predicted6 <- 0
  predicted7 <- 0
  predicted8 <- 0
  predicted9 <- 0
  predicted10 <- 0
}
```

```
inter <- 0
false <- 0
notlin <- 0

for (i in 1:simsize)
{
  # x1 - x10 uniformly distributed over [-1,1]
  x1 <- runif(noobs,-1,1)
  x2 <- runif(noobs,-1,1)
  x3 <- runif(noobs,-1,1)
x4 <- runif(noobs,-1,1)
x5 <- runif(noobs,-1,1)
x6 <- runif(noobs,-1,1)
x7 <- runif(noobs,-1,1)
x8 <- runif(noobs,-1,1)
x9 <- runif(noobs,-1,1)
x10 <- runif(noobs,-1,1)
x1.1 <- runif(noobs,30,75)
x12 <- rbinom(noobs,1,0.5)
x13 <- rnorm(noobs,0,1)
x14 <- runif(noobs,0,10)
x15 <- rbinom(noobs,1,0.7)

# error eps normally distributed with mu=0 and sigma=1
eps <- rnorm(noobs,0,sigma)

# Definition of the Response
if(scenario == 1) # scenario A: large coefficients
{
  y <- 10*x1*x2+5*x3+10*sin(4*x4)+log(abs(x5))+sigma*eps
}
```

```

marsdata <- cbind(y,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10)
marsdata <- as.data.frame(marsdata)
}
else if(scenario == 2) # scenario B: small coefficients
{
y <- x1*x2+x3+sin(4*x4)+log(abs(x5))+eps
marsdata <- cbind(y,x1,x2,x3,x4,x5,x6,x7,x8,x9,x10)
marsdata <- as.data.frame(marsdata)
}
else if(scenario == 3) # same relationship as survival data
{
y <- 0.04*(x1.1/10)^2 - 6*x12 + 0.1*x1.1*x12
marsdata <- cbind(y,x1.1,x12,x13,x14,x15)
marsdata <- as.data.frame(marsdata)
}

attach(marsdata)

# call of polymars
library(polspline,T)
if(scenario == 1 || scenario == 2)
{
poly <- polymars(y,cbind(x1,x2,x3,x4,x5,x6,x7,x8,x9,x10),startmodel=
matrix(c(1,NA,NA,NA,2,NA,NA,NA,3,NA,NA,NA,4,NA,NA,NA,5,NA,
NA,NA,6,NA,NA,NA,7,NA,NA,NA,8,NA,NA,NA,9,NA,NA,NA,10,NA,NA,NA),
nrow=10,byrow=T),knot.space=20,gcv=gcv)
}
else if(scenario == 3)
{
poly <- polymars(y,cbind(x1.1,x12,x13,x14,x15),startmodel=

```

```
matrix(c(1,NA,NA,NA,2,NA,NA,NA,3,NA,NA,NA,4,NA,NA,NA,5,NA,
NA,NA),
nrow=5,byrow=T),knot.space=20,gcv=gcv)
}

# predicted values
xmod <- poly$ranges.and.medians[3,]

x1lim <- c(poly$ranges.and.medians[1,1],poly$ranges.and.medians[2,1])
x2lim <- c(poly$ranges.and.medians[1,2],poly$ranges.and.medians[2,2])
x3lim <- c(poly$ranges.and.medians[1,3],poly$ranges.and.medians[2,3])
x4lim <- c(poly$ranges.and.medians[1,4],poly$ranges.and.medians[2,4])
x5lim <- c(poly$ranges.and.medians[1,5],poly$ranges.and.medians[2,5])

pred.values1 <- pred.values2 <- pred.values3 <- pred.values4 <-
pred.values5 <- matrix(nrow=noobs, ncol=ncol(poly$ranges.and.medians),
data=xmod,byrow=T)

mesh1 <- seq(x1lim[1],x1lim[2],(x1lim[2]-x1lim[1])/(noobs-1))
mesh2 <- seq(x2lim[1],x2lim[2],(x2lim[2]-x2lim[1])/(noobs-1))
mesh3 <- seq(x3lim[1],x3lim[2],(x3lim[2]-x3lim[1])/(noobs-1))
mesh4 <- seq(x4lim[1],x4lim[2],(x4lim[2]-x4lim[1])/(noobs-1))
mesh5 <- seq(x5lim[1],x5lim[2],(x5lim[2]-x5lim[1])/(noobs-1))

pred.values1[,1] <- mesh1
pred.values2[,2] <- mesh2
pred.values3[,3] <- mesh3
pred.values4[,4] <- mesh4
pred.values5[,5] <- mesh5
```

```
Y1 <- predict.polymars(poly,pred.values1)
Y2 <- predict.polymars(poly,pred.values2)
Y3 <- predict.polymars(poly,pred.values3)
Y4 <- predict.polymars(poly,pred.values4)
Y5 <- predict.polymars(poly,pred.values5)

predicted1 <- predicted1 + Y1
predicted2 <- predicted2 + Y2
predicted3 <- predicted3 + Y3
predicted4 <- predicted4 + Y4
predicted5 <- predicted5 + Y5

if(scenario == 1 || scenario == 2)
{

  x6lim <- c(poly$ranges.and.medians[1,6],poly$ranges.and.medians[2,6])
  x7lim <- c(poly$ranges.and.medians[1,7],poly$ranges.and.medians[2,7])
  x8lim <- c(poly$ranges.and.medians[1,8],poly$ranges.and.medians[2,8])
  x9lim <- c(poly$ranges.and.medians[1,9],poly$ranges.and.medians[2,9])
  x10lim <- c(poly$ranges.and.medians[1,10],poly$ranges.and.medians[2,10])

  pred.values6 <- pred.values7 <- pred.values8 <- pred.values9 <-
  pred.values10 <- matrix(nrow=noobs, ncol=ncol(poly$ranges.and.medians),
    data=xmod,byrow=T)

  mesh6 <- seq(x6lim[1],x6lim[2],(x6lim[2]-x6lim[1])/(noobs-1))
  mesh7 <- seq(x7lim[1],x7lim[2],(x7lim[2]-x7lim[1])/(noobs-1))
  mesh8 <- seq(x8lim[1],x8lim[2],(x8lim[2]-x8lim[1])/(noobs-1))
  mesh9 <- seq(x9lim[1],x9lim[2],(x9lim[2]-x9lim[1])/(noobs-1))
  mesh10 <- seq(x10lim[1],x10lim[2],(x10lim[2]-x10lim[1])/(noobs-1))
```

```
pred.values6[,6] <- mesh6
pred.values7[,7] <- mesh7
pred.values8[,8] <- mesh8
pred.values9[,9] <- mesh9
pred.values10[,10] <- mesh10

Y6 <- predict.polymars(poly,pred.values6)
Y7 <- predict.polymars(poly,pred.values7)
Y8 <- predict.polymars(poly,pred.values8)
Y9 <- predict.polymars(poly,pred.values9)
Y10 <- predict.polymars(poly,pred.values10)

predicted6 <- predicted6 + Y6
predicted7 <- predicted7 + Y7
predicted8 <- predicted8 + Y8
predicted9 <- predicted9 + Y9
predicted10 <- predicted10 + Y10
}

# requests especially the simulations with interaction of x1, x2 and
# x1.1, x1.2, respectively
if(length(poly[[1]][,1])>1)
{
  for(k in 1:length(poly[[1]][,1]))
  {
    if(poly[[1]][k,1] == 1 && poly[[1]][k,3] == 2 &&
       is.na(poly[[1]][k,2]) && is.na(poly[[1]][k,4]) ||
       poly[[1]][k,1] == 2 && poly[[1]][k,3] == 1 &&
```



```

        is.na(poly[[1]][k,2]) && is.na(poly[[1]][k,4]))
      {inter <- inter + 1}
    }
  }

if(scenario == 1 || scenario == 2)
{
  # requests false results
  falsch <- 0
  if(sum(poly[[1]][,1] == 4) < 2) {falsch <- falsch + 1} # linear x2
  if(sum(poly[[1]][,1] == 5) < 2) {falsch <- falsch + 1} # linear x5
  if(length(poly[[1]][,1])==1) {falsch <- falsch + 1} # null model
  if(any(poly[[1]][,1]>5)) {falsch <- falsch + 1} # nuisance pars found
  # other interactions
  if(any(poly[[1]][,1]==1 && poly[[1]][,3]!=0 && poly[[1]][,3]!=2))
    {falsch <- falsch + 1}
  if(any(poly[[1]][,1]==2 && poly[[1]][,3]!=0 && poly[[1]][,3]!=1))
    {falsch <- falsch + 1}
  if(any(poly[[1]][,3]!=0 && poly[[1]][,3]!=1 && poly[[1]][,3]!=2))
    {falsch <- falsch + 1}
  if(any(poly[[1]][,1]==3 && !is.na(poly[[1]][,2])))
    {falsch <- falsch + 1} # nonlinear x3
  if(any(poly[[1]][,1]==1 && !is.na(poly[[1]][,2])))
    {falsch <- falsch + 1} # nonlinear x1
  if(any(poly[[1]][,1]==2 && !is.na(poly[[1]][,2])))
    {falsch <- falsch + 1} # nonlinear x2

  if(falsch > 0) false <- false + 1
}

```

```

model <- poly[[1]]
#print(model)
if(scenario == 3)
{
  # linear x1.1
  nolin <- F
  for(j in 1:length(model[,1]))
  {
    if(model[j,1] == 1 && !is.na(model[j,2])) nolin <- T
  }
  if(nolin == T) notlin <- notlin + 1

  # false basis functions
  falsch <- 0
  if(sum(model[,1] == 1) < 3) {falsch <- falsch + 1}
                                # linear x1.1 or no interaction
  if(sum(model[,3] == 2) == 0) {falsch <- falsch + 1}
                                # no interaction
  if(sum(model[,1] == 2) == 0) {falsch <- falsch + 1} # no x12
  if(length(model[,1]) == 1) {falsch <- falsch + 1} # null model
  if(sum(model[,1] > 2) > 0) {falsch <- falsch + 1} # nuisance parameters
  if(any(model[,3] != 0 && model[,3] != 1 && model[,3] != 2))
    {falsch <- falsch + 1} # other interaction
  if(length(model[,1]) > 1)
  {
    for(k in 1:length(model[,1]))
    {
      if((model[,1] == 1 && model[,3] != 2) ||
          (model[,1] == 2 && model[,3] != 1)) {falsch <- falsch + 1}
      # false interaction
    }
  }
}

```

```
    }
  }
  if(falsch > 0) false <- false + 1
}

detach(marsdata)
}

false <- false/simsize*100
inter <- inter/simsize*100
if(scenario == 3) {notlin <- notlin/simsize*100}

# number of simulations with false basis functions
listefalsch <- list(paste("simulations with false basis functions:",
                          false,"%"),
                  paste("Simulations with interaction of x1 and x2:",
                          inter,"%"))
if(scenario == 3)
{
  listefalsch <- list(paste("simulations with false basis functions:",
                            false,"%"),
                    paste("Simulations with nonlinear x11:",notlin,"%"),
                    paste("Simulations with interaction of x1 and x2:",
                            inter,"%"))
}

# plot of the mean model
predicted1 <- predicted1/simsize
predicted2 <- predicted2/simsize
predicted3 <- predicted3/simsize
```

```
predicted4 <- predicted4/simsize
predicted5 <- predicted5/simsize
if(scenario == 1 || scenario == 2)
{
  predicted6 <- predicted6/simsize
  predicted7 <- predicted7/simsize
  predicted8 <- predicted8/simsize
  predicted9 <- predicted9/simsize
  predicted10 <- predicted10/simsize
  par(mfrow=c(4,2))
  plot(mesh3,predicted3,type="l",xlab="x3", ylab="y",main="Variable x3")
  plot(mesh4,predicted4,type="l",xlab="x4", ylab="y",main="Variable x4")
  plot(mesh5,predicted5,type="l",xlab="x5", ylab="y",main="Variable x5")
  plot(mesh6,predicted6,type="l",xlab="x6", ylab="y",ylim=c(-6,4),
        main="Variable x6")
  plot(mesh7,predicted7,type="l",xlab="x7", ylab="y",ylim=c(-6,4),
        main="Variable x7")
  plot(mesh8,predicted8,type="l",xlab="x8", ylab="y",ylim=c(-6,4),
        main="Variable x8")
  plot(mesh9,predicted9,type="l",xlab="x9", ylab="y",ylim=c(-6,4),
        main="Variable x9")
  plot(mesh10,predicted10,type="l",xlab="x10", ylab="y",ylim=c(-6,4),
        main="Variable x10")
}
if(scenario == 3)
{
  par(mfrow=c(2,2))
  plot(mesh1,predicted1,type="l",xlab="x11", ylab="y",main="Variable x11")
  plot(mesh3,predicted3,type="l",xlab="x13", ylab="y",ylim=c(-6,4),
        main="Variable x13")
}
```

```
plot(mesh4,predicted4,type="l",xlab="x14", ylab="y",ylim=c(-6,4),
      main="Variable x14")
plot(mesh5,predicted5,type="l",xlab="x15", ylab="y",ylim=c(-6,4),
      main="Variable x15")
}

return(listefalsch)

}
```

B.2 Interaction of a binary and a continuous covariate

This function was used to execute the simulations on the interaction of a continuous and a binary covariate (section 4.2).

```
interact <- function(nobs,simsize,vari=50)
{
  library(foreign,T)
  library(polspline,T)

  rsqohne <- rep(0,simsize) # Rsquare without interaction
  rsqmit <- rep(0,simsize) # Rsquare with interaction

  for(i in 1:simsize)
  {
    age <- trunc(runif(nobs,40,75)) # continuous covariate
    eps <- rnorm(nobs,0,vari) # error
```

```
geschl <- rbinom(nobs,1,0.5)      # binary covariate
ageknot <- age-50                # knot for cont. cov
for(j in 1:max(nobs))
{
  if(ageknot[j]<=0) ageknot[j] <- 0
}

# defintion of the response
y <- ((age-40)^2 - 10*age - 7*ageknot*geschl - 2*geschl
      + eps)/100

simdat <- as.data.frame(cbind(age,geschl,eps,y))
write.dta(simdat, file = sprintf("simdat_%i.dta", i))
attach(simdat)

# polymars with and without interaction
polohne <- polymars(y,cbind(age,geschl),additive=T)
polmit <- polymars(y,cbind(age,geschl))

# predicted values
pred <- predict.polymars(polohne,cbind(age,geschl))
pred1ohne <- pred[geschl==1]
pred0ohne <- pred[geschl==0]
age1 <- age[geschl==1]
age0 <- age[geschl==0]

# plot
basis <- polmit$model[!is.na(polmit$model[,2]),]
basisohne <- polohne$model[!is.na(polohne$model[,2]),]
basis <- basis[basis[,1]==1,2]
```

```
basisohne <- basisohne[basisohne[,1]==1,2]
mesh <- sort(c(basis,min(age),max(age)))
meshohne <- sort(c(basisohne,min(age),max(age)))
geschlvec1 <- rep(1,length(mesh))
geschlohne1 <- rep(1,length(meshohne))
geschlvec0 <- rep(0,length(mesh))
geschlohne0 <- rep(0,length(meshohne))
pred1 <- cbind(mesh,geschlvec1)
predohne1 <- cbind(meshohne,geschlohne1)
pred0 <- cbind(mesh,geschlvec0)
predohne0 <- cbind(meshohne,geschlohne0)
pred.values1 <- predict.polymars(polmit,pred1)
pred.ohne1 <- predict.polymars(polohne,predohne1)
pred.values0 <- predict.polymars(polmit,pred0)
pred.ohne0 <- predict.polymars(polohne,predohne0)
plot(age,y, xlab="z", main="variance = 50")
lines(mesh,pred.values1,col=2)
lines(mesh,pred.values0,col=4)
lines(meshohne,pred.ohne1,col=1)
lines(meshohne,pred.ohne0,col=1)
legend(65,-2,c("w/o inter","x = 0","x = 1"),merge=T,
      fill=c(1,4,2), bty="n")

# goodness of fit
predohne <- predict.polymars(polohne,cbind(age,geschl))
predmit <- predict.polymars(polmit,cbind(age,geschl))
sseohne <- sum((y-predohne)^2)
ssemit <- sum((y-predmit)^2)
sst <- sum((y-mean(y))^2)
rsquareohne <- 1-sseohne/sst
```

```
rsquaremit <- 1-ssemit/sst
rsqohne[i] <- rsquareohne
rsqmit[i] <- rsquaremit

detach(simdat)
}

meanrsqohne <- mean(rsqohne)
meanrsqmit <- mean(rsqmit)
minrsqohne <- min(rsqohne)
minrsqmit <- min(rsqmit)
maxrsqohne <- max(rsqohne)
maxrsqmit <- max(rsqmit)
return(list(paste("Mean Rsquare without interaction =",
                  meanrsqohne),
            paste("Min Rsquare without interation =",
                  minrsqohne),
            paste("Max Rsquare without interaction =",
                  maxrsqohne),
            paste("Mean Rsquare with interation =",
                  meanrsqmit),
            paste("Min Rsquare with interation =",
                  minrsqmit),
            paste("Max Rsquare with interation =",
                  maxrsqmit)))
}
```


B.3 Survival MARS simulations

These functions were used to run simulations on the new approach with martingale and deviance residuals from a Cox null model as response in a common MARS setting.

```
simsurv <- function(nobs = 500, simsize = 10, resi = "mart",gcv=4,distr="expo",
                   shapep=2,scalep=1,probs=0.7)
{
  # probs controls the censoring rate of the data:
  # probs = 0.7 => ca. 50% censoring rate
  # probs = 0.89 => ca. 25% censoring rate
  # probs = 0.26 => ca. 85% censoring rate

  library(survival,T)
  library(polspline,T)
  library(foreign,T)

  # generation of the results list

  models <- list("Models")

  correct <- 0
  notlin <- 0
  false <- 0

  # binary covariate
  x2 <- rbinom(nobs,1,0.5)

  predicted10 <- rep(0,length(x2[x2==0]))
```

```
predicted11 <- rep(0,length(x2[x2==1]))
predicted1 <- rep(0,nobs)
predicted2 <- rep(0,nobs)
predicted3 <- rep(0,nobs)
predicted4 <- rep(0,nobs)
predicted5 <- rep(0,nobs)

diffptest <- rep(0,simsize)

# loops
for(i in 1:simsize)
{
  # other covariate
  x1 <- runif(nobs,30,75)

  # nuisance parameters
  x3 <- rnorm(nobs,0,1)
  x4 <- runif(nobs,0,10)
  x5 <- rbinom(nobs,1,0.7)

  # error
  eps <- rnorm(nobs,0,1)
  u <- runif(nobs,0,1)

  # survival time
  if(distr=="expo")
  {
    time <- -log(u)/(0.5*exp(0.04*(x1/10)^2-6*x2+0.1*x1*x2))*100
  }
  if(distr=="weibull")
```

```
{
  time <- (-log(u)/(scalep*exp(0.04*(x1/10)^2-
                             6*x2+0.1*x1*x2)))^(1/shapep)*100
}

# censoring time
cens.zeit <- runif(nobs,0,quantile(time,probs=probs))

# status variable
status <- rep(0,nobs)

# time variable
for(j in 1:nobs)
{
  if(time[j] <= cens.zeit[j]) status[j] <- 1
}

for(j in 1:nobs)
{
  if(status[j]==0) time[j] <- cens.zeit[j]
}

data <- as.data.frame(cbind(x1,x2,time,status,u))

# Cox null model
coxnull <- coxph.moni(Surv(time,status)~1)

# Deviance and Martingale residuals of the null model
devinull <- residuals(coxnull,type="deviance")
martnull <- residuals(coxnull,type="martingale")
```

```
# proceeding of the MARS model for Martingale and Deviance residuals
if(resi == "mart")
{
  marsmod <- polymars(martnull, cbind(x1,x2,x3,x4,x5),gcv=gcv)
}
else if(resi == "devi")
{
  marsmod <- polymars(devinull, cbind(x1,x2,x3,x4,x5),gcv=gcv)
}
model <- marsmod[[1]]

# predicted values

xmod <- marsmod$ranges.and.medians[3,]
x1lim <- c(marsmod$ranges.and.medians[1,1],
          marsmod$ranges.and.medians[2,1])
x2lim <- c(marsmod$ranges.and.medians[1,2],
          marsmod$ranges.and.medians[2,2])
x3lim <- c(marsmod$ranges.and.medians[1,3],
          marsmod$ranges.and.medians[2,3])
x4lim <- c(marsmod$ranges.and.medians[1,4],
          marsmod$ranges.and.medians[2,4])
x5lim <- c(marsmod$ranges.and.medians[1,5],
          marsmod$ranges.and.medians[2,5])
pred.values1 <- pred.values2 <- pred.values3 <- pred.values4 <-
  pred.values5 <- matrix(nrow = nobs, ncol =
    ncol(marsmod$ranges.and.medians),data=xmod, byrow=T)
mesh1 <- seq(x1lim[1],x1lim[2],(x1lim[2]-x1lim[1])/(nobs-1))
```

```
mesh2 <- seq(x2lim[1],x2lim[2],(x2lim[2]-x2lim[1])/(nobs-1))
mesh3 <- seq(x3lim[1],x3lim[2],(x3lim[2]-x3lim[1])/(nobs-1))
mesh4 <- seq(x4lim[1],x4lim[2],(x4lim[2]-x4lim[1])/(nobs-1))
mesh5 <- seq(x5lim[1],x5lim[2],(x5lim[2]-x5lim[1])/(nobs-1))

pred.values1[,1] <- mesh1
pred.values2[,1] <- mesh2
pred.values3[,1] <- mesh3
pred.values4[,1] <- mesh4
pred.values5[,1] <- mesh5

pred.values10 <- pred.values1[x2==0,]
pred.values11 <- pred.values1[x2==1,]
pred.values10[,2] <- rep(0,length(pred.values10[,1]))
pred.values11[,2] <- rep(1,length(pred.values11[,1]))

Y10 <- predict.polymars(marsmod,pred.values10)
Y11 <- predict.polymars(marsmod,pred.values11)
Y1 <- predict.polymars(marsmod,pred.values1)
Y2 <- predict.polymars(marsmod,pred.values2)
Y3 <- predict.polymars(marsmod,pred.values3)
Y4 <- predict.polymars(marsmod,pred.values4)
Y5 <- predict.polymars(marsmod,pred.values5)

mesh11 <- mesh1[x2==1]
mesh10 <- mesh1[x2==0]

predicted10 <- predicted10 + Y10
predicted11 <- predicted11 + Y11
predicted1 <- predicted1 + Y1
```

```
predicted2 <- predicted2 + Y2
predicted3 <- predicted3 + Y3
predicted4 <- predicted4 + Y4
predicted5 <- predicted5 + Y5

# was the interaction observed?
richtig <- F
for(k in 1:length(model[,1]))
{
  if(model[k,1] == 1 && model[k,3] == 2) richtig <- T
}
if(richtig == T) correct <- correct + 1

# was the non-linearity of x1 observed?
nolin <- F
for(j in 1:length(model[,1]))
{
  if(model[j,1] == 1 && !is.na(model[j,2])) nolin <- T
}
if(nolin == T) notlin <- notlin + 1

rsquare <- rsquare + marsmod[[9]]

# are there any false basis functions?
falsch <- 0
if(sum(model[,1] == 1) < 3) {falsch <- falsch + 1}
  # linear x1 or no interaction x1x2
if(sum(model[,3] == 2) == 0) {falsch <- falsch + 1}
```

```

    # no interaction x1x2
    if(sum(model[,1] == 2) == 0) {falsch <- falsch + 1}
    # no x2
    if(length(model[,1]) == 1) {falsch <- falsch + 1}
    # null model
    if(sum(model[,1] > 2) > 0) {falsch <- falsch + 1}
    # nuisance parameters
    if(any(model[,3] != 0 && model[,3] != 1 && model[,3] != 2))
    {falsch <- falsch + 1}
    # other interaction
    if(length(model[,1]) > 1)
    {
      for(k in 1:length(model[,1]))
      {
        if((model[,1] == 1 && model[,3] != 2) ||
           (model[,1] == 2 && model[,3] != 1)) # false interaction
        {falsch <- falsch + 1}
      }
    }
    if(falsch > 0) false <- false + 1

}

# plot of the mean models
predicted10 <- predicted10/simsize
predicted11 <- predicted11/simsize
predicted1 <- predicted1/simsize
predicted2 <- predicted2/simsize
predicted3 <- predicted3/simsize

```

```
predicted4 <- predicted4/simsize
predicted5 <- predicted5/simsize

par(mfrow=c(3,2))

if(resi == "mart")
{
  residu <- "Martingale residuals"
}
else if(resi == "devi")
{
  residu <- "Deviance residuals"
}

plot(mesh10,predicted10,type="l",col=4,main="Interaction between x1 and x2",
      ylim=c(-1.5,1.5),ylab=residu,lwd=2,xlab="x1")
lines(mesh11,predicted11,col=2,lwd=2)
legend(60,-0.55,c("x2 = 0","x2 = 1"),col=c(4,2),lty=c(1,1),lwd=c(2,2),
      cex=0.8)

plot(mesh1,predicted1,type="l",main="Variable x1",ylab=residu,col=2,lwd=2,
      xlab="x1")

plot(mesh2,predicted2,type="l",main="Variable x2",ylim=c(-3,3),ylab=residu,
      col=2,lwd=2,xlab="x2")

plot(mesh3,predicted3,type="l",main="Variable x3",ylim=c(-3,3),ylab=residu,
      col=2,lwd=2,xlab="x3")

plot(mesh4,predicted4,type="l",main="Variable x4",ylim=c(-3,3),ylab=residu,
      col=2,lwd=2,xlab="x4")

plot(mesh5,predicted5,type="l",main="Variable x5",ylim=c(-3,3),ylab=residu,
      col=2,lwd=2,xlab="x5")

# fraction of models with false basis functions
```



```
false <- false/simsize
# fraction of models with nonlinear x1
notlin <- notlin/simsize
# fraction of models with interaction
correct <- correct/simsize

if(resi == "mart")
{
  ergebnis <- list("Model with martingale residuals:",
                  paste("fraction of models with nonlinear x1:",notlin),
                  paste("fraction of models with interaction x1*x2:",
                        correct),
                  paste("fraction of models with false basis functions:",
                        false))
}
if(resi == "devi")
{
  ergebnis <- list("Model with deviance residuals:",
                  paste("fraction of models with nonlinear x1:",notlin),
                  paste("fraction of models with interaction x1*x2:",
                        correct),
                  paste("fraction of models with false basis functions:",
                        false))
}

return(ergebnis)
}
```

B.4 Function for Survival MARS

This function can be used to execute the MARS approach in the survival setting.

```
#####
#                                                                 #
# MARS approach for survival type data with martingale and deviance #
# residuals of a Cox model as response                               #
#                                                                 #
#####

survmars <- function(data, resi="mart", startmodel=T, gcv=4, maxsize,
                    additive=F, knot.space=3)
{
  # data in the order time variable, status variable, coefficients
  # no missing values allowed
  # type of residual may be "mart" (martingale, default) or "devi" (deviance)

  library(survival,T)
  library(polspline,T)

  ncases <- nrow(data)
  if(missing(maxsize)) maxsize <- ceiling(min(6 * (ncases^(1/3)), ncases/4,
                                             100))

  coxnull <- coxph(Surv(data[,1],data[,2])~1)
}
```

```
# martingale residuals
if(resi == "mart")
{
  resinull <- residuals(coxnull,type="martingale")
}

# deviance residuals
if(resi == "devi")
{
  resinull <- residuals(coxnull,type="deviance")
}

# MARS model
pred <- data[,-c(1,2)]

if(startmodel == T)
{
  n <- ncol(pred)
  sequenc <- seq(1,n,1)
  nas <- rep(NA,n)
  startmodel <- as.matrix(cbind(sequenc,nas,nas,nas))
  marsmod <- polymars(resinull,pred,startmodel=startmodel,gcv=gcv,
                      maxsize=maxsize,additive=additive,knot.space=knot.space)
}
else if(startmodel == F) marsmod <- polymars(resinull,pred,gcv=gcv,
                                             maxsize=maxsize,additive=additive,
                                             knot.space=knot.space)

if(resi == "mart")
  results <- list("Survival MARS with martingale residuals",marsmod)
```

```
if(resi == "devi")
  results <- list("Survival MARS with deviance residuals",marsmod)

return(results)

}
```

Bibliography

- [1] Aalen OO *Statistical inference for families of counting processes*. Ph.D. dissertation (1975), University of California, Berkeley.
- [2] Andersen PK, Borgan Ø, Gill R & Keiding N *Statistical Models Based on Counting Processes* (1993), Springer, New York.
- [3] Barlow WE & Prentice RL *Residuals for relative risk regression*. *Biometrika* 75.1 (1988): 65 - 74.
- [4] Bauer A, Kantelhardt JW, Barthel P, Schneider R, Maekikallio T, Ulm K, Hnatkova K, Schoemig A, Huikuri H, Bunde A, Malik M, Schmidt G *Deceleration capacity of heart rate as a predictor of mortality after myocardial infarction: cohort study*. *The Lancet* 367(2006): 1674 - 1681.
- [5] Bender R, Augustin T, Blettner M *Generating survival times to simulate Cox proportional hazards models*. *Statis-*

- tics in Medicine 24 (2005): 1713 - 1723.
- [6] Breslow NE *Contribution to the discussion on the paper by DR Cox, Regression and life tables*. Journal of the Royal Statistical Society B 34 (1972): 202 - 205.
- [7] Breslow NE *Covariate Analysis of Censored Survival Data*. Biometrics 30 (1974): 89 - 99.
- [8] Cox DR *Regression Models and Life-Tables*. Journal of the Royal Statistical Society B 34 (1972): 187 - 202.
- [9] Cox DR & Oakes D *Analysis of Survival Data* (1984), Chapman & Hall, London.
- [10] Doob JL. *Stochastic Processes* (1953) Wiley, New York.
- [11] Fleming TR & Harrington DP *Counting Processes & Survival Analysis* (1991), John Wiley & sons, New York.
- [12] Friedman J *Multivariate Adaptive Regression Splines*. The Annals of Statistics 19.1 (1991): 1 - 141.
- [13] Hastie T & Tibshirani R *Generalized Additive Models* (1990), Chapman & Hall, New York.
- [14] Klein JP & Moeschberger ML *Survival Analysis - Techniques for Censored and Truncated Data* (1997), Springer, New York.

- [15] Kooperberg C, Stone CJ & Truong YK *Hazard Regression*. Journal of the American Statistical Association 90, Issue 429 (1995): 78 - 94.
- [16] Kooperberg C, Bose S, Stone CJ *Polychotomous Regression*. Journal of the American Statistical Association 92, Issue 437 (1997): 117 - 127.
- [17] LeBlanc M & Crowley J *Adaptive Regression Splines in the Cox Model*. Biometrics 55 (1999): 204 - 213.
- [18] Lown B & Verrier RL *Neural activity and ventricular fibrillation* New England Journal of Medicine 294 (1976): 1165 - 70.
- [19] McCullagh P & Nelder JA *Generalized Linear Models* (1998), Chapman & Hall, London.
- [20] Meyer PA. *Probability and Potentials* (1966) Waltham MA, Blaisdell.
- [21] Parmar MKB & Machin D *Survival Analysis - A Practical Approach* (1996), John Wiley & sons, Chichester.
- [22] Priori SG, Aliot E, Blomstrom-Lundqvist C, et al. *Task Force on Sudden Cardiac Death of the European Society of Cardiology*. European Heart Journal 22 (2001): 1374 - 1450.

- [23] Royston P & Altman DG *Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with discussion)*. Applied Statistics 43 (1994): 429 - 467.
- [24] Royston P & Sauerbrei W *A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials*. Statistics in Medicine 23 (2004): 2509 - 2525.
- [25] Schmidt G, Malik M, Barthel P et al. *Heart-rate turbulence after ventricular premature beats as a predictor of mortality after acute myocardial infarction* The Lancet 353(1999): 1390 - 96.
- [26] Therneau TM & Grambsch PM *Martingale-based residuals for survival models*. Biometrika 77.1 (1990): 147 - 160.
- [27] Therneau TM, Grambsch PM, *Modeling Survival Data - Extending the Cox Model* (2000), Springer, New York .

Lebenslauf

Monika Kriner, geb. Seebauer
geboren am 1.9.1978 in München

Schulbildung

09/1984 - 07/1988 Grundschule an der Pfarrer-Grimm-Straße München
09/1988 - 06/1997 Louise-Schroeder-Gymnasium München
Juni 1997 Abitur

Studium

11/1997 - 09/2003 Diplom-Studiengang Statistik an der Ludwig-Maximilians-
Universität München
05/2000 Vordiplom
Anwendungsgebiet: medizinische Biometrie (Doppelgebiet)
Vertiefungsfach im Hauptstudium: mathematische Stochastik
Diplomarbeit: Bump Hunting zur Identifizierung von
Therapie-Respondern unter der Berücksichtigung von
zeitvariierenden Koeffizienten

Beruflicher Werdegang

05/2001 - 09/2003 studentische Hilfskraft am Institut für medizinische Statistik
und Epidemiologie am Klinikum rechts der Isar
seit 11/2003 wissenschaftliche Mitarbeiterin am Institut für medizinische
Statistik und Epidemiologie am Klinikum rechts der Isar und im
Sonderforschungsbereich 386