Modelling time-varying effects in Cox model under order restrictions

Georgia Salanti, Kurt Ulm

Abstract

The violation of the proportional hazards assumption in Cox model occurs quite often in studies concerning solid tumours or leukaemia. Then the time varying coefficients model is its most popular extension used. The function $\phi(t)$ that measures the time variation of a covariate, can be assessed through several smoothing techniques, such as cubic splines. However, for practical propose, it is more convenient to assess $\phi(t)$ by a step function. The main drawback of this approach is the lack of stability since there is no standard method of defining the cutpoints of the underlined step function.

The variation in the effect of a predictor can be assumed to be monotonic during the observational period. In these cases, we propose a method to estimate $\phi(t)$ based on the isotonic regression framework. Applying the idea of Grambsch and Therneau, where smoothing the Schoenfeld residuals plotted against time reveal the shape of the underlined $\phi(t)$ function, we use the Pooled Adjacent Violators Algorithm as smoother. As a result a set of cutpoints is returned without any a priori information about their location. Subsequently, the corresponding step function is introduced in the model and the standard likelihood-based method is applied to estimate it while adjusting for other covariates. This approach presents the advantage that additional decisions that can effect the result, as the number of knots in cubic splines, do not need to be taken. The performance of the provided PH test and the stability of the method are explored in a simulation study.

1 Introduction and background

The Cox model is by far the most popular procedure for analyzing survival data. Consider the case where P predictors $\mathbf{X} : X_1, X_2, ..., X_P$, have been identified to affect significantly the survival probability. The Cox model specifies the hazard for an individual i as

$$\lambda(t|\mathbf{X}) = \lambda_0(t)e^{\tilde{\beta}\mathbf{X}}. \tag{1}$$

A key assumption of this model is that the ratio of two hazards is independent on time (proportional hazards model or PH model), i.e. the impact of each predictor included in the model does not change during the observation period and therefore the relative risk RR regarding two levels x_i, x_j of an explanatory variable is $exp(\beta(x_i - x_j))$ at any time. However this assumption may not hold for some variables included in the model. In that case the coefficient β_i and therefore the RR are functions of time $\beta = \beta(t)$ and $RR = exp(\beta(t)(x_i - x_j))$.

The application of the Cox model requires validation of the proportional hazards assumption. In this direction, several tests have been proposed so far to check the predictors for time-dependency. In case of evidence, the usual PH model needs transformation, in order to include the dynamic structures.

Many graphical approaches have been proposed in order to check for proportionality. Although the judgment is rather subjective, they can be used as a first guide. Consider again a predictor in categories, a first intuitive way is to check the Kaplan-Meier curves for parallelism. If that is true, proportionality is rather likely to be fullfield. The equivalent multivariate approach would be to fit a Cox model stratified for the factor of interest and plot the survival curves for the mean value of the other predictors. The resulting curves should be parallel but also in agreement with the survival curves estimated non-parametrically (for example the Altschuler-Nelson estimates).

Another more sophisticate graphical estimation of PH assumption can be performed by plotting the log minus log survival functions against time for each level of the predictor [1]. If the proportionality assumption holds, the two curves should be parallel. To assess the survival function in each level of the predictor one has to fit again a stratified Cox model. Alternatively one can use the cumulative Schoenfeld residuals. Under the proportional hazard assumtion each curve should be a random walk starting and ending at 0 (Brownian bridge). All graphical approaches described above present difficulties of visualizing the actual pattern of time-dependency and to reveal the consequences of the underlying violation of proportional hazards.

Alternatively, one can split the data in subgroups that correspond to pre-selected time intervals. In each data set a Cox model is fitted and the coefficients obtained are compared to the confidence interval of the overall coefficient. Moreover, in case of violation, the pattern of interval-coefficients can roughly indicate the form of the time dependency. The time-intervals are usually selected to include enough events, but no further cut-off criteria can be established.

The most accurate approach is to apply *time-varying coefficients model* [3] where the coefficient $\tilde{\beta}$ is allowed to be a function of time $\tilde{\beta}(t)$. It provides a test for proportional hazards and a modeling alternative in case of violation. As special part of this approach, the Grambsch and Therneau test is defined which is based on scaled Schoenfeld residuals. Regarding this approach, a new proposal will be presented in this paper: the incorporation of isotonic regression [6] in the Grambsch and Therneau test to improve power.

The principal motivation for using isotonic regression in modeling time variation in Cox model, is that it provides a changepoint model regarding time. Therefore, optimal cutpoints can be assessed to split time in intervals within which the effect of the variable of interest remains constant. That is an important task in many clinical studies. Isotonic regression provides unbiased estimators for changepoints without any additional requirements.

This paper focuses on combining the benefits from isotonic regression and the flexibility of the varying-coefficients approach. The first part deals with an isotonic

[1]That is because: $S(t) = S_0(t)^{e^{\beta x}} \rightarrow S_l(t) = S_k(t)^{RR} \rightarrow ln(S_l(t)) = RRln(S_k(t)) \rightarrow ln(-ln(S_l(t))) = ln(RR) + ln(-ln(S_k(t)))$

version of the Grambsch and Therneau test [2, 9]. Further, we will present how one can use the isotonic smoother to model the function $\tilde{\beta}(t)$ in a time-varying Cox model. The gain of introducing isotonic regression in testing and modeling PH departures will be outlined and a simulation study will be performed to assess the properties of the approach. Finally we will present an application in data set containing children with acute lymphoblastic leukemia.

Notation:
As D is denoted the total number of events and t the random variable for the survival time. As t_j, $j = 1, ..., J$ are denoted the unique failure times with $d_j > 0$ individuals failing at t_j and R_j the observations having $t > t_j$.

2 The time-varying coefficients Cox Model

As pointed out in the introduction section one easily expressed alternative to proportional hazards is provided by applying models with a time-dependent coefficient. That is simply an extension of the Cox model where the time consistency assumption on $\tilde{\beta} = (\beta_p, p = 1, ..., P)$ is relaxed and is allowed to be a function of time $\beta_p(t) = \beta_{0p} + \beta_{1p} f_p(t)$. Model (1) takes the following form:

$$\lambda(t) = \lambda_0(t) e^{\tilde{\beta}(t) X} \tag{2}$$

where $\tilde{\beta}(t)$ is the vector $(\beta_1(t), \beta_2(t), ..., \beta_P(t))$. If the predictor is a binary variable, $\beta_p(t)$ measures the difference in log(relative risk) between the two groups as a function of time. The advantage of this approach is twofold: On one hand it offers a straightforward way to investigate time-dependent structures, by testing for $\beta_{1i} = 0$. On the other hand, in case of PH rejection, it provides automatically an alternative model that fits adequately the data.

In case that all coefficients in $\tilde{\beta}$ vary ($\tilde{\beta} = \tilde{\beta}(t) = (\beta_1(t), \beta_2(t), ..., \beta_P(t))'$) the usual partial likelihood of the model takes the form:

$$L(\beta_1(t), \beta_2(t), ..., \beta_P(t); t) = \prod_{j=1}^{J} \frac{exp(\sum_{l=1}^{d_j} X_l \tilde{\beta}(t_j))}{[\sum_{s \in R_j} exp(X_s \tilde{\beta}(t_j))]^{d_j}} \tag{3}$$

where X_l is the covariate vector corresponding to lth failure at time j.

For one predictor p a function f_p is used so that $\tilde{\beta}_p(t) = \beta_0 + \beta_p f_p(t)$. The adequacy of this approach depends clearly on the choice of function f_p. There are several proposals about how to estimate the appropriate function f_p. The two main methods that can be used - smoothing splines and fractional polynomials - are shortly presented in section 3.1 together with a new method using isotonic regression. For isotonic estimation, fitting algorithms and tests, see [6].

3 Detecting PH departures under order restriction

Assume that if there is any PH violation, it follows a monotonic pattern. Starting from a time-varying Cox model (equation 2), the Schoenfeld residuals provide a useful tool in detecting time-variation for the predictors of interest. That can be accomplished either graphically, or by applying a specific test as outlined below.

3.1 Smoothing Schoenfeld residuals scatterplot

The Schoenfeld residuals are defined at each unique failure times. In absence of ties they are equal to the difference between the observed covariate vector for an event at time $t_j, j = 1, ..., J$ and its expected value.

$$\tilde{r}_j = \tilde{X}_j - E(\tilde{X}_j | R_j) \rightarrow \tilde{r}_j = \tilde{X}_j - \frac{\sum_{l \in R_j} \tilde{X}_l e^{\tilde{\beta} X_l}}{\sum_{l \in R_j} e^{\tilde{\beta} X_l}}. \tag{4}$$

In the presence of P covariates, the Schoenfeld residuals $\tilde{r}$ can be presented as a $J \times P$ matrix.

Assume that for each variable p we have one estimated coefficient for each event time i.e. β_{pj}. Grambsch and Therneau [2] showed that if β_p is the coefficient from an ordinary PH Cox model, then

$$E(r^*_{pj}) + \beta_p \approx \beta_{pj}(t) \tag{5}$$

where $r^* = V_{\tilde{\beta}}^{-1} r$ are the *scaled Schoenfeld residuals* and $V_{\tilde{\beta}}$ is the variance matrix for the estimated coefficients $\tilde{\beta}$. This suggests to plot $r^*_{pj} + \beta_p$ versus time, to reveal the functional form of time variation. In case that the PH assumption holds, the residuals should form approximately a horizontal line at the constant coefficient β_p from model (1). One can used any kind of smoother for this purpose.

A popular choice are the *natural cubic splines.* The principal idea is to split the time-axis by selecting an appropriate number of *nodes* and to fit piecewise polynomials. The choice of number of nodes (which determines the degrees of freedom) can affect the result, and no specific functional form is given. *Fractional polynomials* [7, 8] provide an interesting alternative, and result in a functional estimation of the time variation, but again one has to choose a set of exponents and maximal number of components.

The *isotonic smoother* provides an alternative to standard used smoothers. It requires monotonic trend, which is true for many prognostic factors. For example, consider a long-time therapy in which younger people respond better, but its prognostic value decreases with age. Additional considerations as for example the number of nodes need not be taken. The main advantage is that it detects jumps in risk the for the time axis. Without any a priori information, the procedure returns some cutpoints, and segments the observational time in homogenous groups. The risk within each group is considered to be constant.

3.2 Grambsch and Therneau test and its isotonic version

Next to this graphical approach, Grambsch and Therneau introduced a version of the score test based on the weighted Schoenfeld residuals. Assume that all P predictor variables are time-dependent. The coefficient for the p variable has the time-depended form $\beta_p(t) = \beta_{0p} + \beta_{1p}(f_p(t) - \bar{f}_p)$ where $\bar{f}_p$ is the mean of $f_p(t)$ over time. Then, the PH hypothesis implies that $H_0: \ \beta_{1p} = 0$.

Using matrix notation the test statistic takes the following form

$$GTtest_{Px1} = \frac{[(\tilde{t} - \bar{t})' r^*]^2}{diag(V_{\tilde{\beta}}) D \sum (t_i - \bar{t})^2} \tag{6}$$

where $V_{\tilde{\beta}}$ is the variance-covariance matrix for the estimated coefficients $\tilde{\beta}$. Each one of the resulting values corresponds to a variable and tests for time-dependency. This test is approximately χ^2 distributed with one degree of freedom for each tested coefficient.

This test can be thought of as a generalization of the least-square statistic for estimating $\beta(t)$ given equation (5). Under the assumption of monotonic trend, one can substituted function $f(t)$ by the isotonic function $is(t)$: if $is(\beta_p(t))$ is a consistent estimator of $\beta_p(t)$ then

$$is(r^*_{pj}) + \beta_p \approx \beta_{pj} \tag{7}$$

where $is(r^*_{pj})$ is the residual matrix divided in blocks that correspond to time intervals. Substituting r^* by the isotonic estimation $is(r^*)$ in equation 6 results to an *isotonic version of the GT test.*

Note that the idea to use piecewise constant and non overlapping time intervals to estimate $f(t)$ was first proposed by O'Quigley and Pessione [5]. However, as noted in their paper, the investigator has to choose the partition of the time axis. Although the authors introduce some useful guidelines, the choice of the cutpoints remains rather subjective. Applying isotonic regression this disadvantage is bypassed. In section 5 the performance of isotonic transformation in the residuals is assessed and compared to the standard Grambsch and Therneau test.

4 Fitting the generalized additive model using isotonic smoothing techniques

Fitting smoothing splines in estimating $\beta(t)$ within the Cox model requires maximization of the *penalized likelihood function*. The result is a natural cubic spline, having nodes at each failure time point. The oscillation of the fitted spline increases as the penalty parameter decreases. This parameter need to be pre-specified and defines the degrees of freedom. With fractional polynomials, one has to fit a stratified Cox model where the unique failure time points $t_j, j = 1, ..., k$ determine the strata. At each such strata the corresponding covariate values are attributed and the new observational time is set to $t_{j+1} - t_j$. Using then X and $\tilde{f}(t)X$ as predictors, the stratified Cox model applied in the new data set will provide $\tilde{\beta}(t)$.

With step functions modeling time-varying effects is easier. Once the time-intervals are estimated the varying coefficients model (2) shall be estimated. Assuming that PAVA returns m time cutpoints regarding the effect of a variable, the time-varying coefficient for this variable takes the form:

$$\beta(t) = \beta_0 + \alpha_1 I_{t_1}(t) + \alpha_2 I_{t_2}(t) + ... + \alpha_m I_{t_m}(t) \tag{8}$$

$$I_{t_j}(t) = \begin{cases} 0 & \text{if } t <= t_j \\ 1 & \text{if } t > t_j \end{cases} .$$

The functional form of $\beta(t)$ has to be introduced in the model in order to estimate $\tilde{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_m)$. Standard Likelihood based methods are applied for this purpose. Thereafter the usual Score test or the Likelihood Ratio test with m degrees of freedom can be applied to compare the PH model to the dynamic model, by testing all time-specific coefficients to be zero:

$$H_0 : \alpha_1 = ... = \alpha_m = 0. \tag{9}$$

The parameter α_i measures the increase (or decrease) in the risk from time t_{i-1} to time t_i on a logit scale.

It is very often the case that the time axis seems oversegmented. Some of the observed cutpoints do not correspond in an important increase (or decrease) in risk.

One has to proceed to *a backward elimination* of the level sets. First the time groups containing few events (less than 10% of the total number of events) are deleted. Once those groups are eliminated, the likelihood ratio test can be applied to test one by one the coefficients $\alpha_i = 0$ in order to define the neighboring level sets that do not differ significantly. The deletion of a coefficient a_i and its time-interval I_{t_i} is equivalent to its union to the previous interval. The elimination proceeds by such time-interval unions, re-fits the Cox model and stops when all α_i are found to be significant. The $(1-\alpha)\%$ confidence band for a time varying predictor is expressed by

$$CI_{1-\alpha} = \tilde{\beta} \pm \sqrt{X^2_{df,1-\alpha/2} diag(ZV_{\tilde{\beta}}Z')} \tag{10}$$

where $V_{\tilde{\beta}}$ is the large sample variance-covariance matrix for $\tilde{\beta} = (\beta_0, \alpha_1, \alpha_2, ..., \alpha_m)$.

When more than one covariate is time-varying, the backfitting strategy is applied to fit the model. The general idea is to fit the time-varying coefficients allowing variation at one variable at time while the rest covariates remain time-independent

$$\beta^{it=1}(t) = (\beta_0^{it=1} + \tilde{\alpha}_1^{it=1} f_1(t), \beta_2^{it=1}, ..., \beta_P^{it=1})$$

where $f(t)$ is a step function. The likelihood ratio test will assess the gain in the fit i.e will test ${\alpha_\mathbf{1}}^{it=1} = 0$. In case of evidence $f(t)$ is retained. In the next step all coefficients are reestimated, allowing now variation for the first two variables

$$\beta^{it=2}(t) = (\beta_0^{it=2} + \tilde{\alpha}_1^{it=2} f_1(t), \beta_0^{it=2} + \tilde{\alpha}_1^{it=2} f_2(t), ..., \beta_P^{it=2})$$

where only $f_1(t)$ is estimated from the previews step and held constant in step 2. The procedure goes on like that updating in each iteration only the coefficients. Such loops are repeated until a small change in the likelihood is achieved.

5 Simulation study

A simulation study was conducted to explore the properties of the new proposal for testing proportional hazards applying the isotonic version of Grambsch and Therneau test (equation 6). This section focuses on revealing the advantages of the isotonic GT test against the conventional test. When forming assumptions about the functional form of the regression, we tried to be as consistent as possible with situations frequently observed in clinical studies. The simulations are designed to avoid ties.

Only the case of a simple binary predictor is considered. One proportional and three non proportional hazard models are analyzed. In the baseline group the covariate has been set $X = 0$ and the hazard $\frac{e^{-4}}{1+e^{-4}}$. The treatment group has $X = 1$ and hazard $\frac{e^{-4+\beta(t)}}{1+e^{-4+\beta(t)}}$. Each group contains 100 observations.

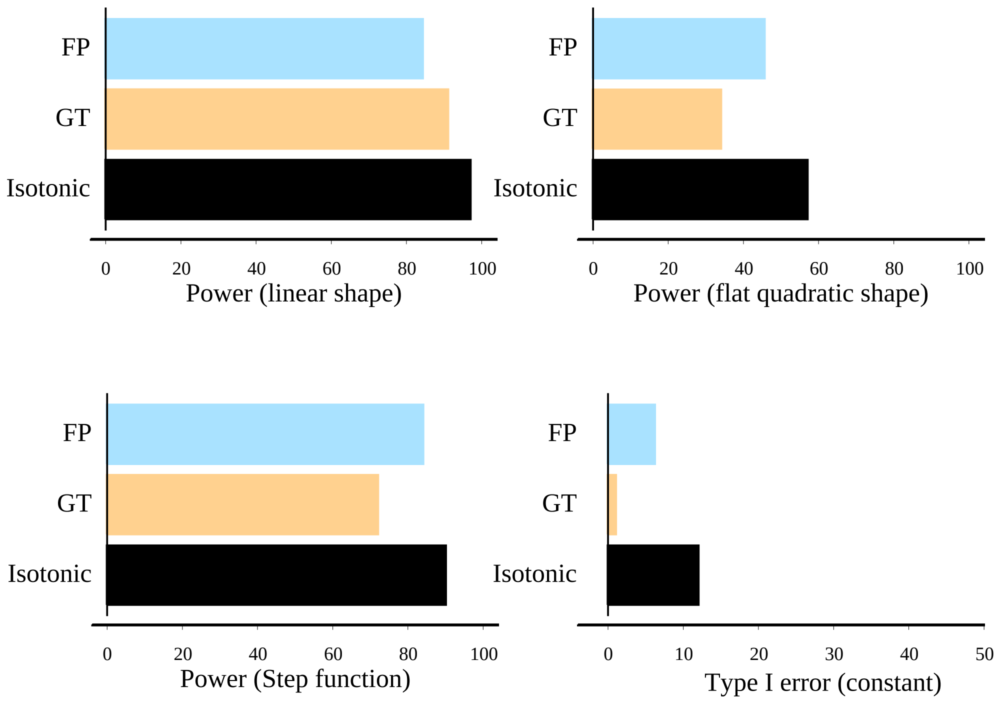

Figure 1: *Simulations study for survival data. Compare in terms of power (first three figures) and type I error (last figure) the Grambsch and Therneau test (GT test), the fractional polynomials test and the isotonic version of GT test.*

To generate the data sets, we proceed separately in each group (treatment or baseline) as follows: starting from time=1 the number of failures is calculated using the hazard function. For the observations remaining at risk, the number censored observations is calculated, as a random binary process. The procedure is repeated for time=2 and stops when no more observations remain at risk. The censoring probability used here was 0.5%. To model dynamic structures that decrease with time the following scenarios are made:

Linear: a decreasing linear time-dependency where $\beta(t) = -0.02t + 1$.

Quadratic: where $\beta(t) = -0.04t - 0.004t_2 - 1$ representing a decreasing umbrella shape

Step function: having shift at t=24 and $\beta(t) = \begin{cases} 1.5 & t \leq 24 \\ 0 & t > 24 \end{cases}$.

Constant: $\beta(t) = 1$ for estimating the properties in case where the PH assumption is not violated.

Simulations under the first three functions will give information about the power of the compared tests, whereas with the constant function the type I error will be assessed. Three test are compared: a) a test based on fractional polynomials model described in [1] b) the GT test (6) assuming linear transformation for time and c) the isotonic version of GT test (5). The results are presented in figure 1.

The isotonic test presents the best power for all non-constant functions, whereas the conventional GT test gives the lowest power. For every shape the power from fractional polynomial is lower that this from isotonic regression. One would expect that this advantage of the isotonic test is eliminated in case of a non-monotonic function. The more flexible approach as the fractional polynomials should present a better performance in case of the flat quadratic function. This is not the case, as outlined in figure 1: isotonic regression gives higher power for this shape as it gives for a step function. However, the price one has to pay for the increasing power in the isotonic test is a higher type I error.

6 Case Study in time-varying Cox model

The data set used to illustrate the above approaches contains 141 observations from children having acute leukemia (ALL). The endpoint was overall survival time. The probability to die within a period of 7 days to about 10 years follow up, has been found to be dependent upon the following binary variables:
- Remission after the first induction (REMI, 1: yes)
- ALL relapse after the first Chemotherapy (RELP, 1: yes)
- The size of massive spleen below the rib (MSPS, 1: > 1 cm)
- White blood cell count (WBC, 1: $> 60.6\ 10^9/L$)

The survival time is measured in YEARS. The main of the study was to estimate if there is a time variation in the effect of MSPS, and in case of evidence to describe this variation. The sample is characterized by a high event rate (122/141), and the value of deviance in absence of any predictor is estimated to be 1037.59.

The Cox PH model with forward LR selection has been applied and table 1 shows the estimated coefficients. Time variation in the predictive value of MSPS has been tested applying Grambsch and Therneau test, Kaplan-Meier curves (figure 6) and smoothing the Schoenfeld residuals using splines, fractional polynomials and isotonic regression (figure 3).

Table 1: *Acute lymphoblastic leukemia study: The PH Cox model. The deviance is 957.08 with 5 degrees fo freedom.*

Variables	Coefficients	SE	p-value
RELP	0.507	0.219	0.021
REMI	-0.991	0.387	0.010
MSPS	0.549	0.203	0.007
WBC	0.785	0.232	0.000
CONTS	-0.974	0.362	0.007

These different methods are more or less in agreement: there is a dynamic effect for MSPS. The Grambsch and Therneau test results in a test value 3.930 and the

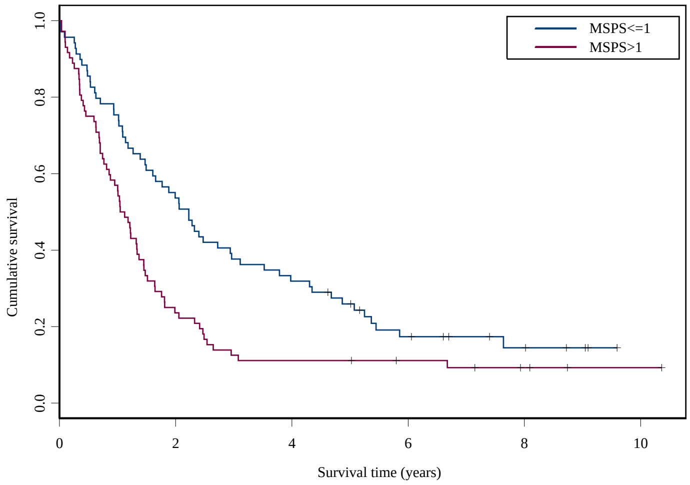

Figure 2: *Kaplan-Meier cumulative survival curves for MSPS.*

corresponding p-value is 0.047. Fitting the varying coefficients model using splines, the constant predictor lies out of the confidence bands for more than 10% of the total number of events (figure not shown). There is a decreasing positive prognostic value for MSPS. Children that do not have massive spleen have better prognosis that decreases progressively, and after about four years the direction of the prognosis changes. This conclusion is quite strange and against any biological plausibility. However a possible explanation could be the following: perhaps many children get a very intensive chemotherapy that is effective against the tumor but is also too burdensome. So, it may cause a preliminary death to many children. But once a child overcomes that crucial period and does not relapse, it has the best chances to survive.

By isotonizing the Schoenfeld residuals (figure 3) the appropriate time-cutpoints are revealed. The confidence intervals correspond to fractional polynomials. However some of the resulting steps contain very few events and therefore do not offer a lot of information while increasing the degrees of freedom. Each group is restricted to contain at least 10% of the total number of events. After elimination of those

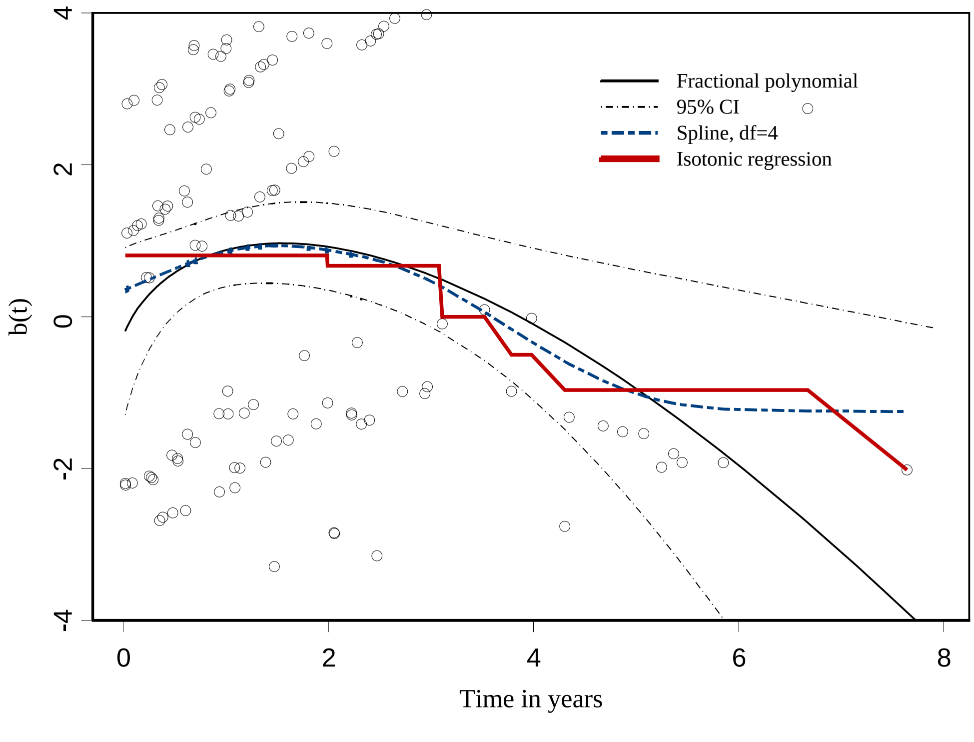

Figure 3: *Smoothing the scaled Schoenfeld residuals for MSPS.*

groups, model 8 can be written for the resulting time-cutpoints:

$$\beta(t) = \beta_0 + \alpha_1 \cdot I_{1.98}(t) + \alpha_3 \cdot I_{3.52}(t). \tag{11}$$

The time-stratified Cox model can now be fitted again to estimate whether some of the $I_t(t)$ variables are non significant predictors and to delete them. Recall that any coefficient α that is found to be non significant corresponds in a union of the above defined time-level sets (table 2). Note that p-value correction has to be considered because of the multiple comparisons i.e. $a = 1 - \sqrt[c]{0}.95$, c the number of time-segments.

Both time-interval variables $I_{1.98}$, $I_{3.52}$ are significant. The fitted function with the corresponding confidence bands are presented in figure 4. The dynamic form $\beta(t)$ for MSPS is:

$$\beta_{MSPS}(t) = 1.57 - 1.18 \cdot I_{1.98}(t) - 2.85 \cdot I_{3.52}(t). \tag{12}$$

Table 2: *Elimination of the time level sets for MSPS dynamic coefficient.*

Coefficient	Deviance	p-value
β_0	957.08	0.0000
α_1	928.59	0.0000
α_2	915.64	0.0013

The final achieved deviance have been estimated 915.64, that yields an overall LR test for PH of 41.44 ($p < 0.001$). Finally the model containing all the significant predictors and their time dependent effects is:

$$h(t) = h_0(t)e^{\beta(t,x)}$$

where

$$\beta(t,x) = 0.490 \cdot RELP - 1.105 \cdot REMI +$$
$$+[1.568 - 1.184 \cdot I_{1.98}(t) - 2.851 \cdot I_{3.52}(t)] \cdot MSPS + 0.235 \cdot WBC - 0.604 \cdot CONTS$$

7 Extensions

One can image implementations of isotonic regression in several approaches regarding survival settings. John O'Quigley [5] for example introduced a test for proportional hazards based on the model: $\lambda(t) = \lambda_0(t)exp[(\tilde{\beta} + \Psi\tilde{\theta})'X]$ The matrix $\Psi = diag(\tilde{\psi}_1, \tilde{\psi}_2, ..., \tilde{\psi}_P)$ is a score matrix determined by the user. Obviously if $\tilde{\theta} = 0$ the proportional hazards model is recovered. The model is fitted using the stratified Likelihood, where arbitrary time cutpoints define the strata, and a sort of score test is applied to test for $\tilde{\theta} = 0$. Isotonic regression can be easily introduced into this context and improve the performance of this approach.

Another assumption undertaken by the Cox model is that each variable enters the model linearly, assumption that may also be violated. This case entails that both coefficient and RR depend on the variable ($\beta = \beta(X), RR = exp(\beta(x_i) - \beta(x_j))$). The adequacy of the linear form of a predictor in the Cox model can be visualized by smoothing the martingale residuals plotted against the predictor. If the shape seems

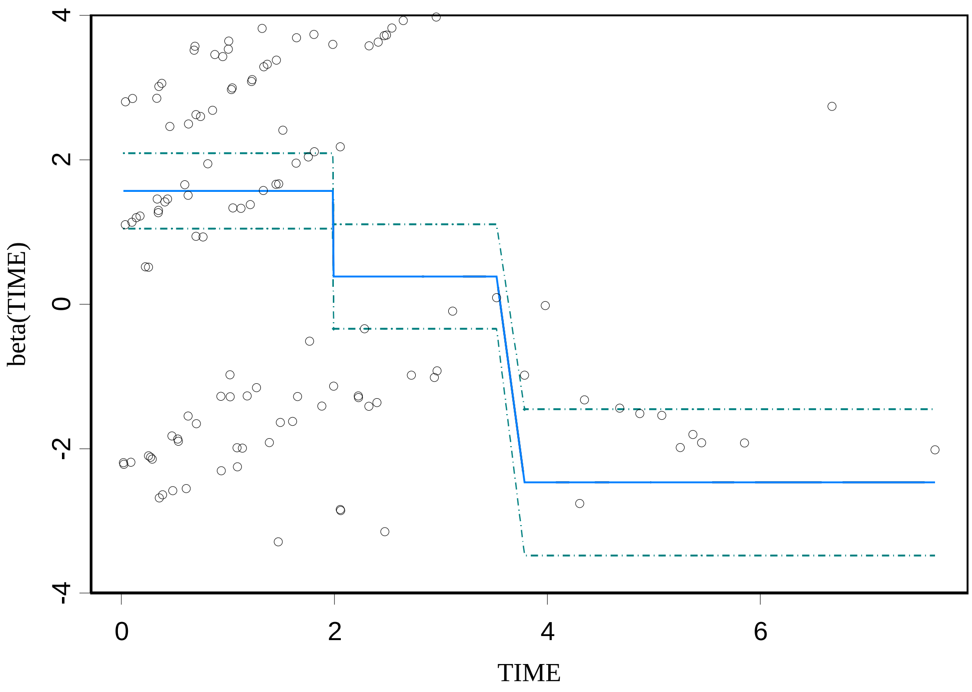

Figure 4: *Isotonic fit for time-dependent coefficient for MSPS.*

not linear, the predictor has to be transformed. An approach similar to this used for modeling time variation can be applied to model properly non linear predictors.

An alternative approach that uses step functions in modeling dynamic structures is accomplished with CART [10]. The main advantage provided is that the time-cutpoints are not prespecified, but the pruning parameter has to be calculated through cross validation. The PAVA algorithm can modify the splitting criteria, to include monotonicity restrains if so required.

References

[1] Berger U., Gerein P., Ulm K., Schfer J. (2000) On the use of fractional polynomials in dynamic cox models. *Discussion Paper No. 207* Sonderforschungsbereich 386 Ludwig-Maximilians-University, Munich.

[2] Grambsch P., Therneau T. (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**: 515-526

[3] Hastie T, Tibshirani R. (1990) *Generalized additive models.* Chapman & Hall, London.

[4] Holm S. (1979) Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association*, **92**: 299-306.

[5] O'Quigley J., Pessione F. (1989) Score tests for homogeneity of regression effect in the proportional hazards model. *Biometrics*, **45**: 135-144

[6] Robertson T., Wright FT. and Dykstra RL. (1988) *Order Restricted Statistical Inference.* Wiley, New York.

[7] Royston P., Altman DG. (1994) Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling (with discussion). *Applied Statistics*, **43**: 429-467

[8] Sauerbrei W., Royston P. (1999) Building multivariable prognostic and diagnostic models: transformation of the predictors using fractional polynomials. *Journal of the Royal Statistical Society*, Series A **162**: 71-94

[9] Therneau T., Grambsch P. (2001) *Modelling survival data.* Springer, New York.

[10] Xu R., Adak S. (2002) Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics*, **58**(2): 305-315