Berger, Gerein, Ulm, Schäfer:

On the use of Fractional Polynomials in Dynamic Cox Models

Sonderforschungsbereich 386, Paper 207 (2000)

Online unter: http://epub.ub.uni-muenchen.de/

On the use of Fractional Polynomials in Dynamic Cox Models

Ursula Berger *
Pia Gerein
Kurt Ulm
Juliane Schäfer

Institut für medizinische Statistik und Epidemiologie,
Technische Universität München, Germany

Abstract

Despite a sophisticated research on modelling of survival data in the last years, the most popular model used in practice is still the proportional hazards regression model proposed by Cox (1972). This is mainly due to its exceptional simplicity. Nevertheless the fundamental assumption of the Cox model is the proportionality of the hazards, which particularly implies that the covariate effects are constant over time. For many applications this assumption is, however, doubtful. Other, more flexible approaches, which are able to cope with non-proportional hazards usually require non-standard estimation techniques, which are often rather complex and thus not favoured in application. Moreover, the selection of an appropriate test-statistic, to examine the improvement of the fit, is not obvious. In this paper we propose a flexible, yet simple method for modelling dynamic effects in survival data within the Cox framework. The method is based on Fractional Polynomials as introduced by Royston and Altman (1994). This allows for a transformation of the dynamic predictor which leads back to the conventional Cox model and hence fitting is straightforward using standard estimation techniques. In addition, it offers the possibility to easily verify the existence of time-variation. We describe a model selection algorithm which enables to include time-varying effects only when evidence is given in the data, in order to construct a model, which is just as complex as needed. We illustrate the properties of the approach in a simulation study and an application to gastric carcinoma data and compare it with other methods (e.g. the residual score test and smoothed Schoenfeld residuals of Grambsch and Therneau, 1994; natural smoothing splines of Hastie and Tibshirani, 1993).

Key words: Cox model, time-varying effects, PH-test, Fractional Polynomials.

*Ursula.Berger@imse.med.tu-muenchen.de

1 Introduction

The statistical analysis of censored failure time data, (e.g. to investigate the prognostic impact of different covariates on survival), is in practice usually performed by the regression model of Cox (1972). This model is particularly advantageous in situations where only the relative impact of a covariate on the hazard, but not the true hazard function itself is of primary interest.

1.1 The Cox PH-model

For ease of notation, we assume that inference about a single covariate is of interest. The Cox model defines the hazard function for failure as:

$$\lambda(t|\mathrm{X}) = \lambda_0(t)\, \exp\{\beta \mathrm{X}\}, \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard and β is the regression coefficient describing the effect of the covariate X. The power of the Cox model is in its semi-parametric character, since no particular shape for the hazard function is specified, i.e. the baseline hazard $\lambda_0(t)$ can be any arbitrary non-negative function of time. However it is presumed that the log hazard ratio is additively associated to the covariates by the linear predictor βX. This in turn leads to the assumption of proportional hazards (PH-assumption), which implies that the ratio of two hazards, i.e. the relative risk $\mathbf{RR}(\mathrm{X}) = \lambda(t|\mathrm{X})/\lambda_0(t) = \exp\{\beta \mathrm{X}\}$, is assumed to be independent of time. With other words, the effect of a covariate measured at a baseline time-point (e.g. surgery, first diagnosis, beginning of a treatment) is supposed to stay unchanged during the whole observation period. It is obvious that this condition is doubtful in numerous practical situations, e.g. a treatment effect may vanish over time or the impact of a covariate may react with some delay, possibly disappearing after a while. The consequence of incorrectly assuming proportional hazards is a questionable model inference where the importance of a covariate might be misinterpreted or overlooked.

1.2 The Dynamic Cox model

To describe the dynamic development of an effect, the Cox PH-model can be modified to a Varying-Coefficient Model (Hastie and Tibshirani, 1993). This leads to the *dynamic* Cox model, where the effect is allowed to vary with time:

$$\lambda(t|\mathrm{X}) = \lambda_0(t)\, \exp\{\beta(t) \mathrm{X}\}, \tag{2}$$

and hence the relative risk $\mathbf{RR}(\mathrm{X}, t) = \exp\{\beta(t)\mathrm{X}\}$ becomes a function in X and time t.
A variety of methods has been suggested to estimate the varying coefficient $\beta(t)$. The first proposal already describes Cox in his original paper (Cox, 1972), where he extends his model of proportional hazards by adding to the constant coefficient β_0 a function of time:

$$\lambda(t|\mathrm{X}) = \lambda_0(t)\, \exp\left\{(\beta_0 + \beta_1\varphi(t))\mathrm{X}\right\}, \tag{3}$$

with $\varphi(t)$ is some arbitrary prespecified transformation of time t and β_1 is a further regression coefficient. On the one hand this extension offers a simple way to investigate consistency with the PH-assumption by testing $H_0 : \beta_1 = 0$. In addition, the approach directly provides an alternative model when the PH-assumption is rejected. However, the power of the test respectively the goodness of the fit distinctly depends on the choice of the function $\varphi(t)$, where typically only simple constructions are used.
A direct and common alternative is to fit a piecewise constant model in prespecified disjunctive time intervals, which results in a step function for $\beta(t)$. To obtain a smooth estimation, Hess (1994) suggests to substitute the constant pieces of $\beta(t)$ by basis functions of cubic regression splines for a prespecified grid of knots.
Several authors introduced smooth estimators using penalty functions together with a smoothing parameter, in order to control the trade-off between fit and roughness (e.g. Gray, 1992; Verweij and van Howelingen, 1995). The most common approach in this class are the natural smoothing splines of Hastie and Tibshirani (1993), where the partial likelihood, measuring the fit, is penalised by the squared second derivative of the effect function. This results in an cubic spline function with knots at each failure time-point, (Zucker and Karr, 1990).
An alternative method is given by Grambsch and Therneau (1994), who propose to study scatter plot smoothers of the scaled Schoenfeld residuals (Schoenfeld, 1982) versus time.
A Bayesian approach is given by Sargent (1997), who defines a hierarchical Cox model with state-space structure and uses MCMC-methods to estimate time-varying coefficients.

A main problem with all those proposals is, that they demand prespecifications, which do noticeably influence the shape of $\hat{\beta}(t)$, (e.g. the number and location of knots, a smoothing parameter or priori assumptions on the dynamic structure). Moreover, most of the non-parametric methods require non standard estimation techniques and their application is not straightforward. In addition, the selection of an appropriate test-statistic to verify the chosen structural form is often not obvious, so that it is difficult to decide, whether the improvement

of the fit using (2) justifies the increase of the model's complexity. However, to ensure that a model reaches acceptance in practice, it is important to keep it as parsimonious as possible and include time-variation only, when evidence is given in the data.

1.3 Testing for time-variation

A number of graphical and test-based procedures for exploring possible dynamic structures, i.e. violations of the PH-assumption, have been proposed in literature. (For a broad collection on reviews and references see e.g. Hess, 1994; Hess, 1995.)

An informal but simple method is to examine, whether the constant estimator of the PH-model lies within the standard error bands of a dynamic estimation, e.g. received by smoothing splines ("SE-method").

Beside graphical methods formal goodness of fit tests for $H_0 : \beta(t) = \beta$ are of particular interest. Following the proposal of Cox, this could be done by estimating the modified predictor $\eta(t) = (\beta_0 + \varphi(t)\beta_1)\mathrm{X}$ of model (3) and testing the hypothesis $H_0 : \beta_1 = 0$, using the common test statistics, (e.g. the likelihood-ratio statistic). Omnibus goodness-of-fit tests, which compare the expected and observed frequencies of failures for a given partition of time and covariate space, are suggested by Schoenfeld (1980), Moreau, O'Quigley and Mesbah (1985) and Moreau, O'Quigley and Lellouch (1986). Hess (1994) notes that the cubic regression spline approach based on fixed knots allows for formally testing the PH-assumption, too. Harrell (1986) recommends to test for correlation between residuals and failure-time, where for non-monotonic time-dependencies an appropriate transformation of time has to be used. Grambsch and Therneau (1994) present a weighted Schoenfeld residuals score test, regarding $\beta(t) = \beta_0 + \beta_1\varphi(t)$ and testing $H_0 : \beta_1 = 0$. They show, that their test is equivalent to a generalised least square test, and that its computation only requires the fit of the PH-model under H_0 together with the corresponding Schoenfeld residuals. The disadvantage however is, that the suspected departure from time-constancy has to be prespecified in $\varphi(t)$.

An overview of different tests including a discussion of their performance is found in Ng'Andu (1997). However, since all these tests require again a prespecification of the time-variation by a functional form respectively a partitioning of the time-axis, they are in some way only appropriate for testing on a certain structure of time-dependency. And even though some of the tests are fairly thorough so that the prespecifications might not need to match the exact time-structure of the effects to identify PH-violation, none of them provides an acceptable

alternative once departures from the PH-assumption have been detected. An alternative is proposed by Gray (1994), who considers cubic regression splines together with penalty functions, and presents a formal test on the dynamic effect structure, for which he derives an asymptotic distribution. In broad simulation studies, he shows that the test is rather powerful and robust to the number of knots. However, estimation and testing is based on non-standard methods so that its application is not straightforward.
Consequently survival analysis within the Cox framework is usually realised in two separate steps: First the validity of the PH-assumption is investigated based on simple prespecifications. Then, in absence of proportional hazards, an additional, more flexible model must be determined to adequately fit the data.

In this paper we present a simple, yet flexible method based on Fractional Polynomials (FPs), which enables to detect and simultaneously model dynamic effect-structures. FPs have been introduced by Royston and Altman (1994), who use them to smoothly fit predictors of the form $\eta = f(\mathrm{X})$ in the Generalised Additive Models context. They are an extension of common polynomials, constructed out of terms of the form $\theta^{(p)}$, with the exponents (p) selected out of a set of integer and non-integer values. This definition ensures a large variety of possible shapes, including linear, bounded and asymptotic courses, (for details see Royston and Altman, 1994). We propose the use of Fractional Polynomials in the context of time-varying effects within the Cox model framework. Because of their flexibility, the FP approach allows the identification even of rather complex departures from the PH-assumption. At the same time it provides a good fitting alternative when significant time-variation exists. Furthermore, this approach preserves the linear structure of the predictor, and hence, implementation and inference is straightforward using the standard estimation techniques.

The paper is organised as follows: In the next section we give a formal definition of Fractional Polynomials. We describe the use of these functions in order to estimate a dynamic Cox model and identify temporal dependencies of the effects, using the usual partial likelihood approach and the likelihood-ratio statistic. In Section 3 we describe an algorithm for multivariable modelling. This procedure takes a data-driven decision on the optimal dynamic structure and additionally allows for an iterative variable selection. Simultaneously, the hypothesis of a change in the effect over time is tested. In Section 4 we illustrate the

properties of the FP approach with regard to its capability of identifying and modelling dynamic structures. In a simulation study we analyse its testing-performance and compare it a) with Cox's proposal using simple time-transformations, b) with the residual score test of Grambsch and Therneau and c) with the graphical SE-method using natural smoothing splines of Hastie and Tibshirani. In an application to gastric-cancer data we use the proposed algorithm to define a multivariable prognostic system and compare the outcome a) to the Cox PH-model, b) to the Schoenfeld residuals smoother and c) to the natural smoothing splines. Finally in Section 5 we discuss the advantages of the method, as well as its limitations.

2 Method

The advantage of Cox's modification of the PH-model lies in its simplicity of estimating and testing dynamic effects. The extension for non-proportional hazards by

$$\lambda(t|\mathrm{X}) = \lambda_0(t)\, \exp\{\beta_0 \mathrm{X} + \beta_1 \varphi(t) \mathrm{X}\} \tag{4}$$

can be regarded as the introduction of a new, time-dependent covariate $\mathrm{Z}(t) := \varphi(t)X$, constructed as an interaction term of the covariate X with time. The model can therefore be estimated with the usual partial likelihood approach for models with time-dependent covariates. Nonetheless, its flexibility depends on the function $\varphi(t)$. Stablein et al. (1981) proposes quadratic polynomials in order to model treatment effects that rise in the beginning and decrease later. Gore et al. (1984) employ exponential functions to describe the exponential decay of the relevance of certain covariates in a breast cancer study. Typically, the choices are less subtle, e.g. $\ln(t)$, t or $rank(t)$, and hence less flexible.
It is therefore a general objective to substitute $\varphi(t)$ by a more flexible function, which provides an appropriate and smooth fit for $\beta(t)$ and can be generated by a data-driven algorithm. In addition the advantages of using standard estimation and testing techniques should be retained. This is achieved by Fractional Polynomial functions.

2.1 Fractional Polynomials

A Fractional Polynomial of degree m for a single continuous variable $\theta > 0$ is given by

$$\phi_m(\theta, \mathbf{p}) = \beta_0 + \sum_{j=1}^{m} \beta_j \theta^{(p_j)}, \tag{5}$$

where m is a positive integer, β_j are regression coefficients and $p_1 \leq \ldots \leq p_m$ are any positive or negative real-valued exponents, (Royston and Altman, 1994). The logarithmic function $\theta^{(p_j)} := \ln \theta$ is included for $p_j = 0$.

The definition of FPs assures a wide range of flexible shapes. For $p_j \in I\!N$ the function ϕ_m is a conventional polynomial, while powers $p_j < 0$ render asymptotic courses. The possibility of repeated powers $p_i = \ldots = p_k$, $i < k \leq m$ additionally involve combinations with $\ln \theta$. E.g. a FP of degree $m = 5$ with powers $p = (-0.5, 0, 2, 2, 2)$ is defined as:
$\phi_5(\theta, (-0.5, 0, 2, 2, 2)) = \beta_0 + \beta_1 \frac{1}{\sqrt{\theta}} + \beta_2 \ln \theta + \beta_3 \theta^2 + \beta_4 \theta^2 \ln \theta + \beta_5 \theta^2 (\ln \theta)^2$.

Royston and Altman (1994) introduce these functions in order to model additive structures. In particular they give an example for fitting survival data in an additive Cox model. Also Sauerbrei and Royston (1999) use the FP approach to study prognostic factors for oncological data within the Generalized Additive Models-framework. We propose the use of Fractional Polynomial functions in order to model time-varying coefficients in the dynamic Cox model (2) by defining $\beta(t) = \phi_m(t) = \beta_0 + \sum_{j=1}^{m} \beta_j t^{(p_j)}$. Since the FP approach preserves the linear structure of the predictor, it leads to the time-dependent covariate model

$$\lambda(t|\mathrm{X}) = \lambda_0(t) \, \exp\left\{\phi_m(t)\mathrm{X}\right\} = \lambda_0(t) \, \exp\left\{\beta_0 \mathrm{X} + \sum_{j=1}^{m} \beta_j \mathrm{Z}_j(t)\right\}, \tag{6}$$

with constructed time-dependent components $\mathrm{Z}_j(t) := t^{(p_j)}\mathrm{X}$. As survival time t is positive, $t^{(p_j)}$ is well defined for any $p_j \in I\!R$. Model (6) can again be estimated by the known partial likelihood approach.

2.2 Estimation and verification of the dynamic structure

For the additive Cox model based on FP functions Royston and Altman could perform the estimation of the regression coefficients directly, after computation of $\mathrm{X}^{(p_j)}$, using the partial likelihood method as for the Cox PH-model (1). In dynamic Cox models (2) the determination of (6) corresponds to the estimation of constant effects for the time-dependent covariate $\mathrm{Z}_j(t) := t^{(p_j)}\mathrm{X}$. This can be done by restructuring the data set, for which all

the risk sets at every failure time-point are matched with the appropriate values for the time-dependent covariate and thus a "pooled" data set is obtained. Finally a regular, stratified Cox PH-model is estimated by taking the different failure times as a stratification factor.

Fitting a model with FPs additionally requires to determine the optimal values of degree m and of the powers $p_1, \ldots, p_m$. A convenient and practical way to overcome this problem is to set an upper limit m_{max} for the degree and choose the "best" powers out of a prefixed set of possible powers $\mathcal{P}$ due to some goodness-of-fit criterion. Here we use the p-value of the likelihood ratio statistic as a non-linear function of complexity and fit. Out of all possible combinations of $p_j \in \mathcal{P}$ for all $m \leq m_{max}$ the model with the smallest p-value is selected. Note, that the restriction of the powers to a carefully chosen set $\mathcal{P}$ not only speeds up computation, but also assures reasonable interpretations.

The selected "optimal" FP model is then compared to the PH-model, i.e. the PH-assumption is verified by testing $H_0 : \beta(t) = \beta_0$. Since comparison with the PH-model is a nested hypothesis testing problem, the likelihood ratio test for $H_0 : \beta_1 = \ldots = \beta_m = 0$ can be applied. We consider $2m$ degrees of freedom (df), counting one df for each FP coefficient β_j and one for each selected exponent p_j. (For details on the degrees of freedom see Royston and Altman, 1994.) The general effect of a covariate can be verified by testing $H_0 : \beta(t) = 0$, using $2m + 1$ degrees of freedom. In addition to the tests, the resulting FP function can be plotted along with the confidence bands to visualise the nature of the time dependency. Due to the linearity of the FP-predictor $\eta = \beta_0 \mathrm{X} + \sum_{j=1}^{m} \beta_j t^{(p_j)} \mathrm{X}$, also confidence bands for FPs can be readily computed applying standard estimation techniques:
Using matrix notation, let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_m)'$ be the regression coefficient estimates and $\mathbf{Z} = (\mathrm{X}, t^{(p_1)}\mathrm{X}, ..., t^{(p_m)}\mathrm{X})'$ be the matrix of values for the covariate and the time-by-covariate interactions, so that $\hat{\eta} = \hat{\boldsymbol{\beta}}'\mathbf{Z}$. The $(1-\alpha)$-confidence band for the FP-predictor is given by

$$\mathrm{CI}_{(1-\alpha)} = [\hat{\boldsymbol{\beta}}'\mathbf{Z} \pm (\chi^2_{df,\frac{\alpha}{2}} \mathbf{Z}'\Sigma\mathbf{Z})^{1/2}]$$

where Σ is assumed to be a large-sample covariance matrix for $\hat{\boldsymbol{\beta}}$ and $\chi^2_{df,\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-fractile of the χ^2-distribution with $df = 1 + 2m$ degrees of freedom.

3 A model selection procedure for multivariable analyses

For ease of presentation we have regarded so far only univariate predictors. An extension of the model function to multivariate problems is however basically a question of notation. Suppose that $\mathbf{X}$ is a set of q covariates in arbitrary order. In the multivariable dynamic Cox model with FPs the hazard rate is defined as:

$$\lambda(t|\mathbf{X}) = \lambda_0(t)\,\exp\{\sum_{i=1}^{q}\beta_i(t)\mathrm{X}_i\} = \lambda_0(t)\,\exp\{\sum_{i=1}^{q}[\beta_{i0}\mathrm{X}_i + \sum_{j=1}^{m_i}\beta_{ij}t^{(p_{ij})}\mathrm{X}_i]\}\,, \tag{7}$$

where the functions $\beta_i(t) = \phi_m(t,\mathbf{p})$, $i = 1,\ldots,q$ describe the time interactions of each covariate effect. Model estimation can be realised following the backfitting-strategy proposed by Hastie and Tibshirani (1990), obtaining the coefficients β_{ij} and the exponents p_{ij} iteratively. The backfitting-type procedure starts with fitting the predictor

$$\eta_1 = \beta_{10}\mathrm{X}_1 + \sum_{j=1}^{m_1}\beta_{1j}t^{(p_{1j})}\mathrm{X}_1 + \sum_{i=2}^{q}\beta_{i0}\mathrm{X}_i,$$

by selecting an optimal FP for X_1, choosing m_1 and p_{1j} $(j = 1,\ldots,m_1)$ as described above for the univariate case. The effects of $\mathrm{X}_2,\ldots,\mathrm{X}_q$ are left time-constant in the first step, and only $\beta_{20},\ldots,\beta_{q0}$ are additionally estimated. Then the likelihood ratio test is applied to assess the gain of fitting time-variation, i.e. the hypothesis $H_0 : \beta_1(t) = \beta_{10}$ is tested. If the time-varying effect of X_1 is distinct, the p_{1j}'s of the first FP are fixed and the predictor

$$\eta_2 = \beta_{10}\mathrm{X}_1 + \sum_{j=1}^{m_1}\beta_{1j}t^{(p_{1j})}\mathrm{X}_1 + \beta_{20}\mathrm{X}_2 + \sum_{j=1}^{m_2}\beta_{2j}t^{(p_{2j})}\mathrm{X}_2 + \sum_{i=3}^{q}\beta_{i0}\mathrm{X}_i,$$

is fitted by selecting m_2, p_{2j} for X_2 as above. Note, that in this step only the powers p_{1j} are fixed, while all the coefficients $\beta_{10},\ldots,\beta_{1m_1}$, $\beta_{20},\ldots,\beta_{2m_2}$ and $\beta_{30},\ldots,\beta_{q0}$ are re-estimated. This is continued until η_q is achieved and optimal Fractional Polynomial terms are obtained for all q covariates.

In the next iterations the FP functions are similarly updated for each covariate X_i fixing the FPs of the remaining covariates. The algorithm can be stopped when the selected powers p_j, respectively the fit, do not change from one iteration to the next. If simultaneously a selection of covariates is required, additionally the likelihood ratio test of $H_0 : \beta_i(t) = 0$, $i = 1,\ldots,q$, can be investigated in each iteration. A covariate can then be omitted in one iteration, if it does not provide sufficient improvement to the fit, and be re-entered and re-tested in later iterations.

The order of the covariates is generally irrelevant, if the covariates are independent. For dependent covariates however, the results of the algorithm may depend, like any iterative procedure of this type, on the order of selection. In this case, further analysis of the finally selected FP model - especially of its dynamic structure - has to be performed on a substance matter basis, which possibly requires a reselection of the model. This is, however, a general problem in multivariable analysis.

4 Application

To estimate the FP functions for the dynamic Cox models in all our applications, we restricted the degree of the polynomials to $m \leq 2$ and fix the set of relevant powers to $\mathcal{P} = \{-2, -1, -0.5, 0, 0.5, 1, 2\}$. This follows a proposal of Royston and Altman (1994), who give examples of the large variety of possible shapes, which FPs of this set can take. Our own experiences confirmed, that for the description of time-variations ($\beta(t)$) this setting is sufficient for most practical applications.

4.1 Simulation Study

To analyse the reliability of the FP method, especially regarding its capability to detect special violations of the PH-assumption, we performed a simulation study, where we generated 1000 samples of failure-time data out of five different scenarios. All samples consist of two groups, a "baseline group" with X = 0 and a "risk group" with X = 1, of 100 observations each. To allow a precise simulation of temporal effect structures, the data were generated in a logistic setting, where the "baseline hazard of failure" at each time-point was $\lambda(\mathrm{X}|t_t) = \lambda_0 = \frac{exp\{-4\}}{1+exp\{-4\}}$. We simulated dynamic effects of the group variable X using five different functions (compare Figure 1 (a)):

(1) a null model with no risk-effect: $\beta(t) = 0$, (not plotted),

(2) a constant model with proportional hazards: $\beta(t) = 1$,

(3) a linear time-dependency: $\beta(t) = -0.02t + 1$,

(4) a steep quadratic time-dependency: $\beta(t) = -0.5 + 0.08t - 0.0008t^2$,

(5) a flat quadratic time-dependency: $\beta(t) = -0.5 + 0.04t - 0.0004t^2$.

In addition, we considered a permanent probability for (right) censoring of $P_c = 0.005$, which resulted in a total censoring rate of less then a fourth. These settings produced survival data with few bindings and rather long survival times, so that there were no objections against employing a Cox model. (For the asymptotic equivalence of the discrete logit model and the continuous Cox model, see e.g. Fleming and Harrington, 1991.) Moreover, the structure of the data is very much the same as what is known from clinical studies.

For the simulated data sets (2)-(5) we determined the optimal FP functions and tested for dynamic effect structures (see subsection 2.2) at a significance level of $\alpha_{PH} = 0.05$. We then compared the results with various other proposals:

- The extension (3) proposed by Cox using a) a linear transformation $\varphi(t) = t$, b) a logarithmic transformation $\varphi(t) = \ln(t)$ and c) a quadratic transformation $\varphi(t) = t^2$.
- The residual score test of Grambsch and Therneau (1994), using d) the Kaplan-Meier transformation, and again e) the linear transformation, f) the logarithmic transformation and g) the quadratic transformation.
- The informal "SE-method" using the smoothing spline estimators from Hastie and Tibshirani (1993), checking if the estimated effect β_0 of the constant Cox PH-model lies outside the $2\cdot$ standard errors bands h) for any event point, and i) for more than 10% of the event points.

Figures 1 (b)-(e) show the proportion of cases, where the hypothesis of constant effects, i.e the PH-assumption, was rejected. Hence, Figure 1(b) illustrates the consistency of the different tests. It shows, that the FP approach is slightly too liberal ($\alpha_{PH\_emp} = 0.066$), while the test of Grambsch and Therneau is too stringent. The test of Cox meets the nominal α-level the best. In contrast, the heuristic SE-method clearly far too often wrongly assumes temporal structures, and even the relaxation to a 10% -limit causes an error of 60.9%. This is not surprising, as it is well known that point-wise confidence bands do not allow for a global interpretation. The degree of violation of the significance level is, however, remarkable.
Figures 1 (c)-(d) present the power of the different tests for the three dynamic settings. In the linear setting all tests detect the PH-violation about equally well. In the simulations, where we used a steep quadratic function with a modus at time point 50, the power of the different tests do noticeably differ. With the FP approach, in 84.3% of the situations significant

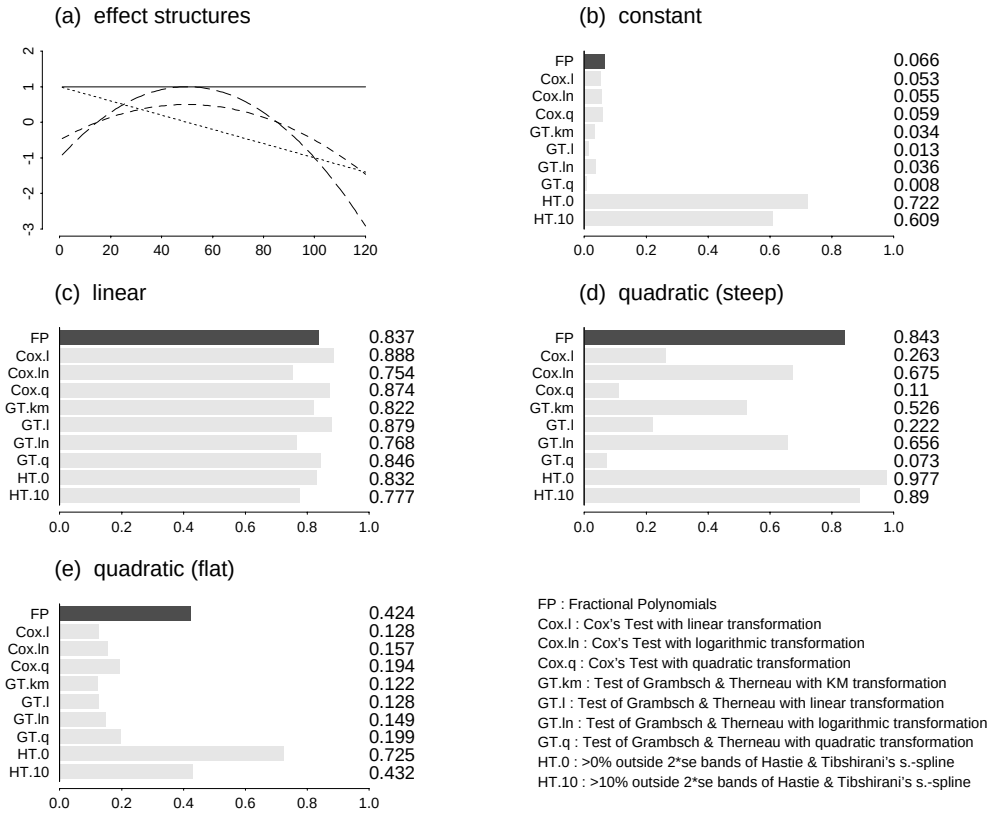

Figure 1: Simulated effect structures and the corresponding powers resp. error rates of different PH-tests at a significance level of $\alpha = 0.05$.

time-variation was detected, while the test based on Cox's proposal as well as the score-test of Grambsch and Therneau even in the best case only found significant PH-violation in about 2/3 of the samples. Note, that the reason of the extremely low powers of these tests when using the quadratic transformation $\varphi(t) = t^2$, is due to the missing linear term, which should have been exactly prespecified to shift the modus of the effect-function from 0 to 50. Of particular interest are the results of the simulations based on a flat quadratic temporal function. Here the FP approach with its flexible functions is clearly superior (42.4%) to the test of Cox and the score-test of Grambsch and Therneau, which could hardly detect the time-variation with the predefined time transformations (power: 12.2% - 19.9%). The interpretation of the SE-method for the dynamic settings is inappropriate due to missing consistency.

Table 1: Power resp. error rates for $H_0 : \beta(t) = 0$ at a significance level of $\alpha = 0.05$

setting	FP model	Cox PH-model
Null model	0.092	0.052
Linear time-dependency	0.98	0.735
Steep quadratic time-dependency	0.924	0.548
Flat quadratic time-dependency	0.464	0.112

Table 1 shows for the null model and the three dynamic settings, how often the FP model found any significant group-effect at a level of $\alpha_{total} = 0.05$, and opposes it to the results of the Cox PH-model. While the FP approach is slightly too liberal when testing for a general effect ($\alpha_{total\_emp} = 0.092$), the Cox PH-model with its constant coefficient often oversees the general group effect, when it varies over time. In the linear setting it was successful in only 73.5% of the samples, while the FP approach modelled an effect in 98%. In the simulations based on steep quadratic time-dependencies, the PH-model succeeded only in 54.8%, while the FP model showed an significant effect in 92.4%, and when the quadratic effect was flat, the PH-model even found it in only 11.2% of the samples, compared to 46.4% for the FP approach.

4.2 Gastric cancer data

For a study on prognostic factors and risk-group stratification in gastric carcinoma, at the Klinikum Rechts der Isar of the Technische Universität München, the survival of gastric cancer patients was followed up after complete resection of the tumour. One major interest of this study was to investigate whether the new tumour-biological factors uPA and PAI-1, assessed in extracts of cancer tissue, provide additional information to prognosis beside established factors, like age, percentage of positive lymph nodes, local tumour invasion and metastasis (Nekarda et al., 1994). The urokinase-type plasminogen activator uPA and its type-1 inhibitor PAI-1 belong to the plasminogen activator system, which has been reckoned to play an important role in tumour cell migration.

295 gastric cancer patients were enrolled between the years 1987 and 1996. 108 of them died during follow up. Time to death is used as failure time, measured in months. The median follow-up time is 41 months. Table 2 gives a short description of the prognostic factors used in the analysis. To simplify the clinical interpretation we restrict the analysis to binary or binary coded factors. NOD.RATIO, T.SUB, uPA and PAI-1 were coded as binary factors

Table 2: Prognostic factors analysed in the gastric cancer study

factor	range	coding	interpretation
AGE	28-90	0: ≤ 65 1: > 65	Age at surgery
NOD.RATIO	0-97	0: < 20 1: ≥ 20	Percentage of positive lymph nodes
T.SUB	1-7	0: ≤ 4 1: > 4	local tumour invasion (Japanese staging system)
METAS	yes/no	0: no 1: yes	distant metastasis
uPA	0.02-20.57	0: ≤ 5.94 1: > 5.94	urokinase-type Plasminogen Activator
PAI-1	0.02-264.62	0: < 4.13 1: ≥ 4.13	Plasminogen Activator Inhibitor Type 1

using cutpoints selected by optimisation of the log-rank-statistics, age was dichotomised at the median.

Univariate analysis

The results of the univariate Cox PH-model for each binary factor are summarised in Table 3, giving the factor's effect β, the relative risk **RR**, its p-value from the likelihood ratio test and the p-value for the test of Grambsch and Therneau.

Table 3: Results of univariate Cox PH-models

factor	Cox PH-fit			Schoenfeld residual test
			$H_0 : \beta = 0$	$H_0 : \beta(t) = \beta$
	β	**RR**	p-value	p-value
AGE	0.18	1.20	0.36	0.002
NOD.RATIO	1.88	6.55	<0.001	0.223
T.SUB	1.55	4.71	<0.001	0.795
METAS	1.37	3.94	<0.001	0.880
uPA	1.09	2.97	<0.001	0.835
PAI-1	1.25	3.49	<0.001	0.975

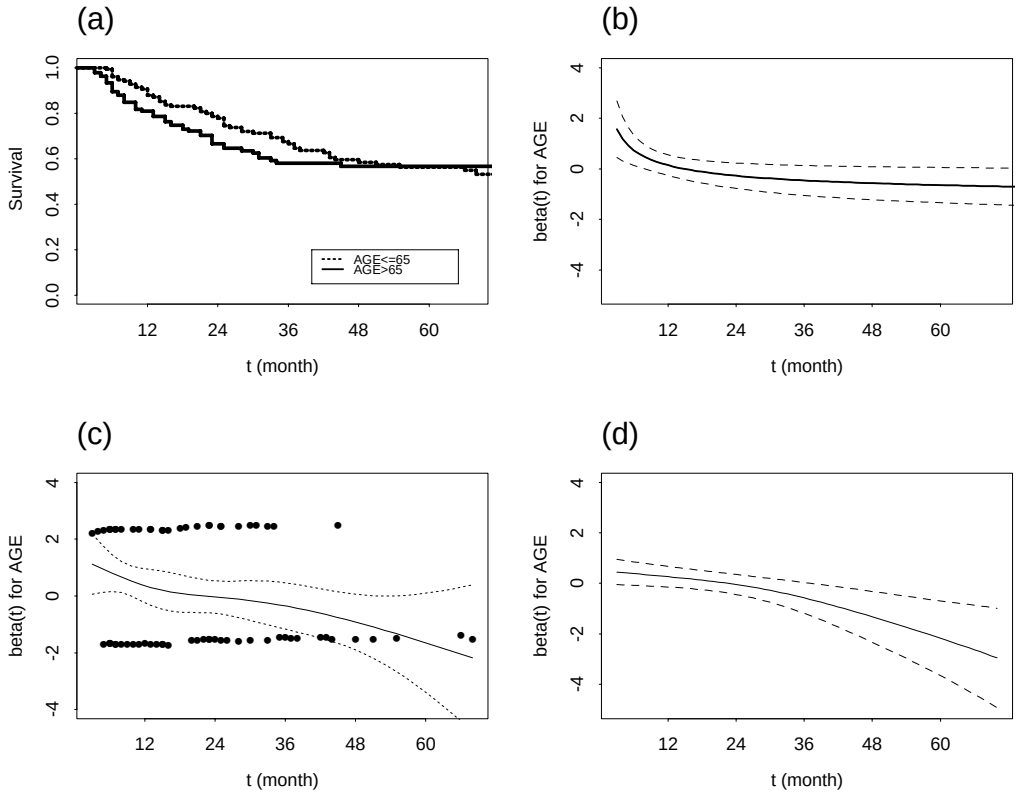

Figure 2: (a) Estimated Kaplan-Meier survival curves of the groups with AGE ≤ 65 and Age > 65. (b) FP function for AGE with ± 2 standard error bands. (c) Raw and spline-smoothed scaled Schoenfeld residuals for AGE, with ± 2 standard errors. (d) The natural smoothing spline estimator of $\beta(t)$ for AGE with 95%-confidence limits.

All factors, except age, show a statistically significant impact on survival at a level of $\alpha = 0.05$. The Schoenfeld residual test, however, identifies significant time-variation for the effect of age ($p = 0.002$).

The univariate FP approach confirms a significant decreasing effect for age ($p_{PH} = 0.005$, $p_{total} = 0.01$). Its FP function $\beta(t) = -1.28 + 4.94 \cdot t^{-0.5}$ is shown in Figure 2(b). Short after surgery the older patients (65 years and older) have a higher mortality rate. This difference is declining with time and after about two years of follow-up the younger patients seem to have a higher risk. For all other factors the FP approach results in time-constant effects. Figure 2(c) gives the scaled Schoenfeld residuals, together with a scatter-plot spline-smoother and ± 2 standard error bands, which yield a similar dynamic structure. For further comparison we additionally determined the natural smoothing spline estimator of $\beta(t)$ for age, given in Figure 2(d), which shows an even sharper decline.

Table 4: Results of a multivariable Cox PH model

factor	Cox PH-fit			Schoenfeld residual test
			$H_0 : \beta_i = 0$	$H_0 : \beta_i(t) = \beta_i$
	β_i	**RR**	p-value	p-value
AGE	0.25	1.28	0.230	<0.001
NOD.RATIO	1.50	4.47	<0.001	0.197
T.SUB	1.11	3.05	<0.001	0.054
METAS	0.07	1.08	0.770	0.408
uPA	0.95	2.57	0.002	0.029
PAI-1	0.88	2.4	0.004	0.413

Multivariable analysis

Table 4 shows the results of the Cox PH-model comprising all six binary covariates. The percentage of positive lymph nodes and local tumour invasion turn out to be the strongest prognostic factors, increasing the risk of death by 4.47 and 3.05. In addition, both proteolytic factors uPA and PAI-1 show statistically significant impact on survival.

Table 5: FP fit in multivariable analysis

factor	$\beta_i(t)$	$H_0 : \beta_i(t) = \beta_{i0}$	$H_0 : \beta_i(t) = 0$
		p-value	p-value
AGE	$\beta_1(t) = 0.9 - 0.001 \cdot t^2$	<0.001	<0.001
NOD.RATIO	$\beta_2(t) = 1.51$	—	<0.001
T.SUB	$\beta_3(t) = 1.19$	—	<0.001
METAS	$\beta_4(t) = -0.13 + 6.63 \cdot t^{-2}$	0.016	0.037
uPA	$\beta_5(t) = 1.82 - 7.19 \cdot t^{-1}$	0.039	0.005
PAI-1	$\beta_6(t) = 0.93$	—	<0.001

The results for the time-varying coefficients model based on FPs are given in Table 5. Beside age also the effects of uPA and distant metastases show a significant change over time. If these dynamic structures are taken into account, all six factors considered provide additional information to prognosis. The test based on scaled Schoenfeld residuals identifies a significant time-variation for the effect of AGE and uPA.

Figure 3 compares the FP estimators of $\beta_i(t)$ for AGE, uPA and METAS to the natural smoothing spline estimator (with 4 df) from a multivariable analysis. Both estimators indicate a distinct decrease of the effect of age, although the FP function has a sharper slope

than the non-parametric estimator in the second half of follow-up. For uPA the FP curve describes an increasing influence with $\beta_{uPA}(t) = 1.82 - 7.19 \cdot t^{-1}$, while the smoothing spline has a slight non-monotonic trend. However, the FP function remains between the confidence limits of the smoothing spline. Some disagreement can be observed between these two estimators for the effect of distant metastasis. While the FP function stays constant after one year, the smoothing spline estimator shows a strong decreasing effect. The reason for this decline could be the small number of events towards the end of follow-up in the rather shrunken risk group "METAS=1", (see Figure 3(c)).

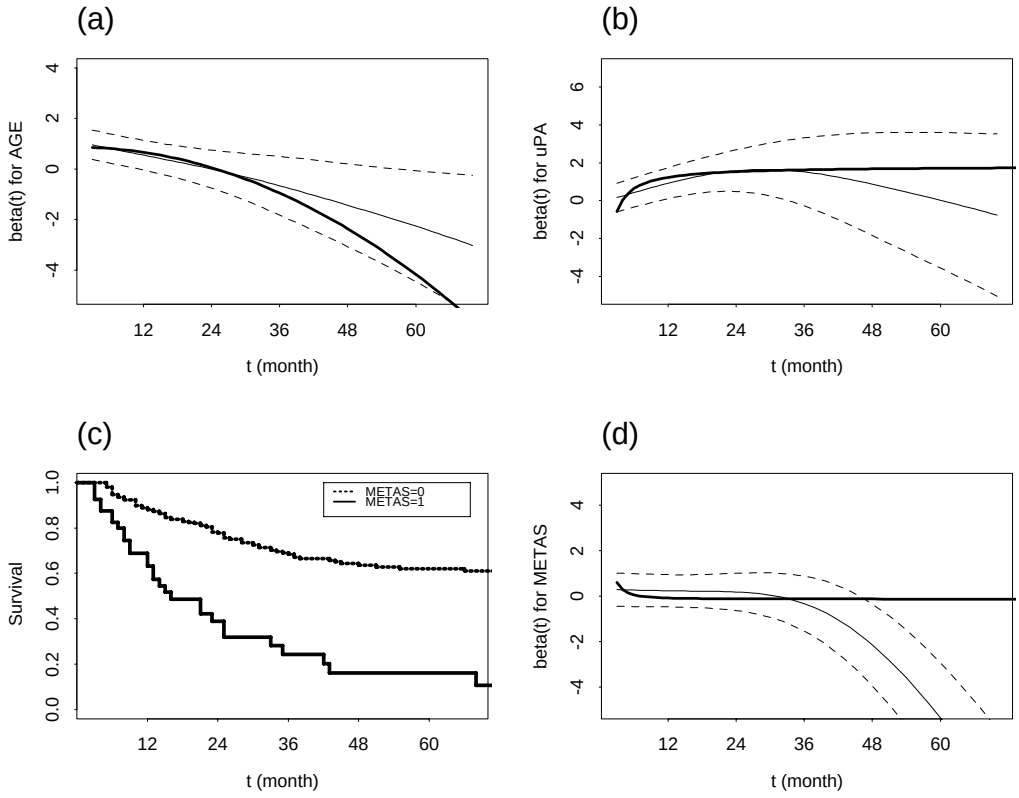

Figure 3: Comparison between FP estimators (thick curve) and smoothing spline estimators of $\beta(t)$ (—— , with 95%-confidence limits - - - -) for (a) AGE, (b) uPA and (d) METAS. (c) Estimated Kaplan-Meier survival curves of the groups with METAS=0 and METAS=1.

5 Discussion

In this paper we propose the use of Fractional Polynomials in order to detect and model dynamic effects in survival data within the Cox model framework. Since the FP functions are linear in the regression coefficients, the estimation problem reduces in formulation (6) to fitting a Cox PH-model with time-dependent covariates. Therefore, model estimation can be performed straightforward, following the strategy described in Section 2.2, where standard methods of inference are used, which are readily available in most statistical packages.

In addition, the FP approach allows formal testing for time-variation in the effects using standard methods, as the likelihood ratio test. In our simulation study this test shows an acceptable consistency with the nominal significance level, and it yields high power for detecting time-variation. In particular, it turns out to be superior to the other PH-validation methods which are widely used. The test of Cox and the score-test of Grambsch and Therneau, which use predefined time-transformations, are not flexible enough to properly identify time-variation of complex structure. The graphical method for checking the PH-assumption, based on a comparison of the constant Cox estimator and point-wise confidence bands of natural smoothing splines highly violates consistency, underlining the fact, that it is not usable for global assessments.

The example of gastric cancer data illustrates clearly the general importance of correctly specifying existing dynamic effect-structures. The prognostic impact of the factor age on survival was not identifiable in the PH-model. Fractional Polynomials offer a useful tool for analysing survival data, when no pre-information about the dynamic effect structure is given and proportionality of the hazards seems doubtful. They describe the relationship between the effect of the factor and time with simple functions, which are rather stable and easily communicated. By allowing for variable selection and verification of the dynamic structure, the multivariable analysis results in a parsimonious model, where only effects are modelled in a dynamic way, when evidence is given in the data.
However, although Fractional Polynomials provide a rather flexible fit their global (non-local) definition still suffers from the same restrictions as other non-local definition for smooth function-estimation. If a fine drawing of the variation is essential, in a hybrid-like algorithm the FP approach could be used to select the important covariates and their effect-structure,

while afterwards a further semi-parametric modelling method could be applied. Yet, care has to be taken, not to result in artefacts.

Royston (2000) recently presented a strategy verifying, whether the global, parametric fit of a FP model misses important information. In order to model the effect of a continuous covariate as parsimonious as possible, but still ensuring a sufficient goodness of fit, he suggests to determine a non-parametric model, e.g. using natural smoothing splines, and test it against the best parametric (FP) model. The test he proposes can directly be transferred to our dynamic effect problem.

Acknowledgement

The authors would like to thank Dr. Hjalmar Nekarda for permission to use the gastric cancer data and helpful discussions regarding the data sets. This work was supported by the DFG within the SFB386-project.

References

[1] Cox, D.R., "Regression Models and Life-Tables" (with discussion), *Journal of the Royal Statistical Society*, B, 34, 1972, pp. 187-220.

[2] Fleming, T.R., Harrington, D.P., *Counting Processes and Survival Analysis*, Wiley Series, New York, 1991.

[3] Gore, S.M., Pocock, S.J., Gillian, R.K., "Regression Models and Non-proportional Hazards in the Analysis of Breast Cancer Survival", *Applied Statistics*, 33, 2, 1994, pp. 176-195.

[4] Grambsch, P. and Therneau, T.M., "Proportional hazards tests and diagnostics based on weighted residuals", *Biometrika*, 81, 3, 1994, pp. 515-526.

[5] Gray, R.J., "Flexible Methods for analyzing survival data using splines, with application to breast cancer prognosis", *Journal of the American Statistical Association*, 87, 1992, pp. 942-951.

[6] Gray, R.J., "Spline-based test in survival analysis", *Biometrics*, 50, 1994, pp. 640-652.

[7] Harrell, F.E., Lee, K.L., "Verifying assumptions of the Cox proportional hazards model", it SUGI II: Proceedings of the Eleventh Annual SAS Users Group International Conference, 1986, pp. 823-828.

[8] Hastie, T.J. and Tibshirani, R.J., *Generalized Additive Models*, Chapman and Hall, London, 1990.

[9] Hastie, T.J. and Tibshirani, R.J., "Varying-Coefficient Models (with discussion)", *Journal of the Royal Statistical Society*, B, 55, 4, 1993, pp. 757-796.

[10] Hess, K.R., "Assessing Time-by-Covariate Interactions in Proportional Hazards Regression Models using Cubic Spline Functions", *Statistics in Medicine*, 13, 1994, pp. 1045-1062.

[11] Hess, K.R., "Graphical Methods for Assessing Violations of the Proportional Hazards Assumption in Cox Regression", *Statistics in Medicine*, 14, 1995, pp. 1707-1723.

[12] Moreau, T., O'Quigley, J. and Mesnah, M., "A Global Goodness-of-fit Statistic for the Proportional Hazard Model", *Applied Statistics*, 34, 3, 1985, pp. 221-218.

[13] Moreau, T., O'Quigley, J. and Lellouch, J., "On D. Schoenfeld's approach for testing the proportional hazards assumption", *Biometrika*, 73, 2, 1986, pp. 513-515.

[14] Nekarda, H., Schmitt, M., Ulm, K., Wenninger, A., Vogelsang, H., Becker, K., Roder, J.D., Fink, U., Siewert, J.R. "Prognostic Impact of Urokinase-type Plasminogen Activator uPA and its Inhibitor PAI-1 in Gastric Cancer with complete Resection (R0-Category, UICC)", *Cancer Research*, 54, 1994, pp. 2900-2907.

[15] Ng'Andu, N.H., "An Empirical Comparison of Statistical Tests for Assessing the Proportional Hazards Assumption of Cox's Model", *Statistics in Medicine*, 16, 1997, pp. 611-626.

[16] Royston, P., "A strategy for modelling the effect of continuous covariates in medicine and epidemiology", *Statistics in Medicine*, 19, 2000, pp. 1831-1847.

[17] Royston, P. and Altman, D.G., "Regression using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling", *Applied Statistics*, 43, 1994, pp. 429-467.

[18] Sargent, D.J., "A Flexible Approach to Time-varying Coefficients in the Cox Regression Setting", *Lifetime Data Analysis*, 3, 1997, pp. 13-25.

[19] Sauerbrei, W. and Royston, P, "Building multivariable prognostic and diagnostic models: transformation of the predictors by using fractional polynomials", *Journal of the Royal Statistical Society*, B, 162, 1999, pp. 77-94.

[20] Schoenfeld, D., "Chi-squared goodness-of-fit tests for the proportional hazards regression model", *Biometrika*, 67, 1, 1980, pp. 145-153.

[21] Schoenfeld, D., "Partial residuals for the proportional hazards regression model", *Biometrika*, 69, 1, 1982, pp. 239-241.

[22] Stablein, D.N., Carter, W.H.Jr., Novak, J.W., "Analysis of Survival Data with Nonproportional Hazard Functions", *Controlled Clinical Trials*, 2, 1981, pp. 149-159.

[23] Verweij, P., van Howelingen, H., "Time-dependent Effects of Fixed Covariates in Cox Regression", *Biometrics*, 51, 1995, pp. 1550-1556.

[24] Zucker, D. and Karr, A., "Non-parametric survival analysis with time-dependent covariate effects: a penalized likelihood approach", *Annals of Statistics*, 18, 1990, pp.329-352.