



Universiteit  
Leiden  
The Netherlands

## Modelling long term survival with non-proportional hazards

Perperoglou, A.

### Citation

Perperoglou, A. (2006, October 18). *Modelling long term survival with non-proportional hazards*. Retrieved from <https://hdl.handle.net/1887/4918>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4918>

**Note:** To cite this publication please use the final published version (if applicable).

## Chapter 2

# Reduced-rank hazard regression

### Abstract

*The Cox proportional hazards model is the most common method to analyze survival data. However, the proportional hazards assumption might not hold. The natural extension of the Cox model is to introduce time varying effects of the covariates. For some covariates such as (surgical)treatment non-proportionality could be expected beforehand. For some other covariates the non-proportionality only becomes apparent if the follow-up is long enough. It is often observed that all covariates show similarly decaying effects over time. Such behavior could be explained by the popular (Gamma-) frailty model. However, the (marginal) effects of covariates in frailty models are not easy to interpret. In this Chapter we propose the reduced-rank model for time varying effects of covariates. Starting point is a Cox model with  $p$  covariates and time varying effects modeled by  $q$  time functions (constant included), leading to a  $p \times q$  structure matrix that contains the regression coefficients for all covariate by time function interactions. By reducing the rank of this structure matrix a whole range of models is introduced, from the very flexible full rank model (identical to a Cox model with time varying effects) to the very rigid rank one model that mimics the structure of a gamma frailty model, but is easier to interpret. We illustrate these models with an application to ovarian cancer patients.*

### 2.1 Introduction

For the last 30 years the field of survival analysis has been dominated by the Cox *Proportional Hazards* (PH) model [24]. The term proportional hazards refers to the fact that covariates have a multiplicative effect on the hazard, implying that the ratio of the hazards for different individuals is constant over time. For some covariates such as (surgical)treatment non-proportionality could be expected beforehand. For some other covariates the non-proportionality only

## Reduced-rank hazard regression

becomes apparent if the follow-up is long enough.

A large number of researchers have worked on generalizations of the PH model. Cox proposed to include an interaction term of the covariates with time as a test for proportionality [24]. Other approaches include a class of dynamic models which use sequential analysis based on a factorization of the likelihood over the time intervals (Gamerman [33]), varying coefficients model, which allow the coefficient of a covariate to vary depending on time or on the value of another covariate (Hastie and Tibshirani [43]), adaptive models that allow stepwise selection of variables and their possibly non linear or time varying effects (Kooperberg, Stone and Truong [58]), smoothed time varying coefficients based on a penalized likelihood (Verweij and van Houwelingen [103]) and many more.

All the above generalizations deal with the problem of non-proportionality but they can be difficult to estimate. The number of free regression parameters can get quite large if many covariates are included with flexible time varying effects. The number of free parameters could be reduced by taking into account mechanisms that can lead to non-proportional hazards such as frailty models. The latter can be thought to arise from omitted covariates or from latent variables as “cured/not-cured”.

A simple frailty model leads to a (marginal) model in which all covariates show similarly decaying effects over time ( a good introduction on frailty models can be found in Aalen [1], [3]). The slightly more complicated cure model allows to model the relation between covariates and frailty as well. The drawback of frailty models is that the effects of the covariates in the marginal models are not additive any more, which makes the models harder to interpret.

As an alternative we propose the use of reduced-rank models in survival analysis. By introducing a structure matrix in a Cox model with many covariates and time varying effects that contains the regression coefficients for all covariate by time function interactions, we can reduce the rank of the model and thus the number of parameters needed to be estimated. The reduced-rank models run from the very flexible full rank model (identical to a Cox model with time varying effects) to the very rigid rank one model that mimics the structure of a gamma frailty model, but is easier to interpret.

The Chapter is organized as follows: In Section 2.2 we start with the basic notation ways to handle non-proportional hazards and the marginal gamma frailty/Burr model . Subsection 2.2.3 provides an introduction to the general idea of reduced-rank regression, followed by the specific reduced-rank theory for modelling survival data with time varying effects. The theory will then be

## 2.2. Time varying effects and frailty models

applied to a data set of ovarian cancer patients in Section 2.3. We will discuss the results and compare the models. The Chapter closes with a discussion.

## 2.2 Time varying effects and frailty models

Let  $t$  denote time until the event of interest,  $X$  a row vector of covariates of dimension  $p$  and  $\beta$  a  $p$ -vector of covariate coefficients  $(\beta_1, \dots, \beta_p)'$ . The Cox model is defined as:

$$h(t|X) = h_0(t) \exp(X\beta) \quad (2.1)$$

where  $h(t|X)$  is the hazard at time  $t$  for individual with covariate vector  $X$  and  $h_0$  is an unspecified non-negative function of time called the *baseline hazard*. The model assumes that the hazard ratio between two subjects with fixed covariates is constant and that is why it is also known as *proportional hazards* model. When the assumption of proportionality does not hold, alternative models have to be considered.

### Cox model with time varying effects of the covariates

For simple categorical covariates apparent non-proportionality of the effects can be handled by stratification. However, this is impossible for continuous covariates and impractical for a large number of categorical covariates. To model non-proportional hazards we follow the line of Cox, who proposed to include an interaction term of the covariates with a time function  $f(t)$  or several interactions with more than one time function. He considered simple transformations of time, such as  $\log(\text{time})$ ,  $\text{time}^2$ ,  $\sqrt{\text{time}}$ . In that case the assumption of proportionality is relaxed by introducing one single function to describe the time varying effect.

More generally, we can construct more complex models by choosing a set of  $q$  “base functions”  $(f_1(t), f_2(t), \dots, f_q(t))$  such as polynomials or splines [40]. The time varying effect of covariate  $X_i$  can then be described as  $\beta_i(t) = \theta_{i1}f_1(t) + \dots + \theta_{ip}f_p(t)$ . The choice of time functions, as well as the number is arbitrary and may depend on the functional form of the data or the nature of the problem. To assure that this model also contains the basic PH-model, the constant is one of the basis time functions or is contained in the linear subspace spanned by the time functions. We will come back to an appropriate choice of the time-functions later on.

Let  $F(t)$  be the row vector containing the time functions, that is  $F(t) = (f_1(t), f_2(t), \dots, f_q(t))$ . Let  $\Theta_{p \times q}$  be the matrix of coefficients whose  $j$ -th column

## Reduced-rank hazard regression

consists of the unknown regression coefficients for the interaction between the covariates and the  $j$ -th time function. The non-proportional Cox model can then be written as

$$h(t|X) = h_0(t) \exp(X\Theta F'(t)) \quad (2.2)$$

The parameters can be estimated by software for time dependent covariates by considering the  $p \times q$  time dependent covariates  $X_i f_j(t)$ . The estimation of this model can be unstable, especially when there are many covariates present and many interactions with time functions. In the latter case there are too many parameters to estimate, and a dimension reduction method as introduced in this Chapter could be very useful.

### Frailty models

Frailty models help us to understand how non-proportional hazards may arise and lead to very parsimonious extensions of the simple Cox model and a way of testing for the presence of non-proportionality. One of the most commonly used frailty model is the Gamma frailty, where the random frailty effect  $Z$  is assumed to follow a  $\Gamma(1/\zeta, 1/\zeta)$  distribution with *mean* = 1 and *variance* =  $\zeta$ . Since the random effect  $Z$  is not observable at the individual level we can consider the model at the population level, or *marginal* model, which can be viewed as the “averaged” hazard function for individuals with same covariate values and corresponds to what can actually be observed. The marginal hazard is:

$$h(t|X) = \frac{h_0(t) \exp(X\beta)}{\{1 + \zeta \exp(X\beta) H_0(t)\}} \quad (2.3)$$

This model can be seen as a model with time varying effects of the covariates, which are assumed to behave in the same way throughout the follow up period, with  $\zeta H_0(t)$  a factor to describe the time dependent pattern. In this model, the baseline hazard is connected to the estimation of  $\beta$ , which makes the model hard to fit. Moreover, the time varying effects are difficult to interpret, because the effects of the covariates in the marginal model are not additive anymore.

The general behavior of the gamma-frailty model was the inspiration to formulate the reduced-rank regression models for survival analysis that will be introduced in the next section.

## 2.2. Time varying effects and frailty models

### Reduced-rank regression

Anderson [10] was the first to apply reduced-rank methods to multivariate linear regression. Consider a multivariate linear regression model with a  $p$ -dimensional predictor  $x$  and a  $q$ -dimensional outcome  $y$ . The linear model could be written as

$$y = a + x\Theta + \epsilon$$

Here,  $a$  is a row vector of unknown constants,  $x$  and  $y$  are row vectors for the independent and responses variables, respectively, and  $\Theta_{p \times q}$  is a matrix whose columns are the unknown regression coefficients for each response and  $\epsilon$  is the row vector of errors with  $E(\epsilon) = 0$ . We call  $\Theta$  the *structure matrix*. To avoid the problem of estimating a large number of parameters and over-fitting the data, Anderson proposed to put a *rank = r* restriction on the structure matrix, that can be written as  $\Theta = B\Gamma'$ , with  $B$  a  $p \times r$  matrix and  $\Gamma$  a  $q \times r$  matrix. In that way the number of parameters needed to be estimated could be reduced, resulting in a more stable and parsimonious model, depending on the rank  $r$ .

Anderson's idea can be used to reduce the dimension of a general Cox model. As discussed earlier a model with time varying effects of the covariates can be described by the structure matrix  $\Theta$  that contains the coefficients for the covariates and their interactions with the time functions. The structure matrix could be factorized in different ways, resulting in a matrix of lower rank  $r$ . Consider  $B$  a  $p \times r$  matrix and  $\Gamma$  a  $q \times r$  matrix of coefficients, that factorize the  $\Theta$  matrix as  $\Theta = B\Gamma'$ . The rank  $r$  model is written

$$h(t|X) = h_0(t) \exp(XB\Gamma'F'(t))$$

or alternatively

$$h(t|X) = h_0(t) \exp \left\{ \sum_{k=1}^r (X\beta_k)(F(t)\gamma_k) \right\} \quad (2.4)$$

with  $\beta_k$  the  $k$ th column of  $B$  and  $\gamma_k$  the  $k$ th column of  $\Gamma$ . In this model  $F(t)\gamma_k, k = 1, \dots, r$  is a set of  $r$  linear combinations of the time functions and  $X\beta_k, k = 1, \dots, r$  a set of  $r$  linear combinations of the covariates. Depending on the data and the nature of the problem the researcher can choose the rank of the model, with maximum rank  $r = \min(p, q)$ . The model is over-specified with  $r(p + q)$  parameters and the number of free parameters is  $r(p + q - r)$ .

## Reduced-rank hazard regression

When  $\Theta = B\Gamma'$  is of full rank, no special structure is imposed and the model is identical to the Cox non-proportional model. When the rank equals one, there are close similarities to the marginal frailty model. A  $rank = 1$  model is an extension of the simple proportional hazards model, where there is just *one* linear combination of time functions  $F\gamma$  to describe the time varying effects of the covariates, which acts in a common way for *all* the covariates. Note that this is similar to the approximation of a Gamma frailty model. It can be shown that the marginal hazard of a frailty model is similar to a Cox model with time-varying effects of the covariates, in which there is one factor  $\xi H_0(t_i)$  to describe the time dependency of the model. The main difference between the gamma-frailty model and the  $rank = 1$  model is that the function that regulates the time-dependency is linked to the base-line hazard in the gamma-frailty model and is totally free in the  $rank = 1$  model.

### Estimation

Consider information on  $n$  individuals with  $p$  covariates that formulate the  $X_{n \times p}$  covariate matrix.  $X_j$  is the row vector of covariates for individual  $j$ . Let  $D$  be the total number of event time points with  $t_1 < t_2 < \dots < t_D$ . Let  $F_i$  be the row vector of time functions at time point  $t_i$ .

The reduced-rank model can be fitted with standard software for Cox regression with time varying effects. The partial log likelihood is given as

$$pl(\beta, \gamma) = \sum_{i=1}^D \sum_{k=1}^r (X_{(i)}\beta_k)(F_i\gamma_k) - \sum_{i=1}^D \ln \left[ \sum_{j \in R_i} \exp \left\{ \sum_{k=1}^r (X_j\beta_k)(F_i\gamma_k) \right\} \right] \quad (2.5)$$

where  $X_{(i)}$  is the covariate vector of a person with event at time  $t_i$ , and  $R_i$  is the set of individuals at risk at time  $t_i$ . We start the estimation procedure giving initial values for the  $r$  vectors  $\beta_k$ . For instance for  $\beta_1$ , the values from a simple Cox model could be used, and when the rank of the model is greater than one a random perturbation of the initial estimate of  $\beta_1$  can be used for the other  $\beta$ 's. Then estimating the  $\gamma_k$ 's, with the  $\beta$ 's fixed, is actually estimating the  $q \times r$  coefficients in a model with time varying effects as the interactions between the linear combinations  $X_j\beta_k$ ,  $k = 1, \dots, r$  and the time functions  $f_1(t), \dots, f_q(t)$ . The next step is the estimation of  $\beta$ 's, where there are  $p \times r$  interactions of the covariates and  $(F_i\gamma_k)$ ,  $k = 1, \dots, r$ . The process then alternates between the two steps until the partial likelihood stabilizes.

As usual, an estimate of the baseline hazard function can be obtained from the Breslow estimator.

## 2.2. Time varying effects and frailty models

This iterative procedure does not yield the full information matrix of the  $\beta$  and  $\gamma$  parameters. The appendix describes how the information matrix can be obtained. Since the reduced-rank is over-specified, the information matrix is singular. A generalized inverse matrix should be used to obtain the variance-covariance matrix of estimable parameters of interest. We use the Moore-Penrose but any generalized inverse would do.

The standard errors of the coefficients  $B$  and  $\Gamma$  can not be estimated, since  $B$  and  $\Gamma$  are not identifiable. The delta method can be used to find standard errors and draw confidence bands for estimable (linear) functions of  $B$  and  $\Gamma$ , like the coefficients of  $\Theta$  and the time varying effects of all covariates. Details can be found at the appendix.

As one referee pointed out, some restrictions could be imposed to the estimation algorithm in order to make the parameters more easily interpretable. The interpretation is linked to the choice of time functions to be discussed in the next subsection.

### Choice of rank and time functions

To determine the optimal rank, one could follow a forward-type algorithm, beginning with a rank one model and moving on by increasing the rank. The final choice could be based on the Akaike information criterion. (It is well-known that the  $\chi^2$  distribution does not apply when testing the rank of the structure matrix).

Another problem to consider is the choice of time functions. There is a series of papers dealing with modelling strategies and time varying effects. The approach of dynamic modelling by Berger, Schäfer and Ulm [13] extended for reduced-ranks, could provide a strategy for deciding both the rank and the time functions of the model. Another choice could be to model the time-dependency of the covariates by the use of splines, which can be simple or more sophisticated to reveal more complex relationships.

The simplest strategy would be to divide the time axis in a number of intervals and to take the indicator functions of those intervals as basic functions. A straight forward extension is take a set of B-splines with the interior knots at convenient positions. In order that the B-functions be estimable, the intervals between the knots should contain enough events. This strategy allows the time functions to be very flexible and lets the model (based on the choice of the rank) to reveal the appropriate functional form of the time varying pattern.

In any choice of time functions though, we impose the condition that  $f_1(0) = 1$  (the simplest choice is  $f_1(t) \equiv 1$ ) and  $f_k(0) = 0$  for  $k = 2, 3, \dots$ , so one could



Table 2.1: Definitions of variables and patients frequencies

$X_k$	0	1	2	3	4
Karnofsky	100	90	80	70	< 70
$n$	137	108	47	46	20
$X_f$	0	1			
Figo	III	IV			
$n$	262	96			
$X_d$	0	1	2	3	4
Diameter	Micro	< 1	1-2	2-5	> 5
$n$	29	67	49	68	145

see the effect of the covariates easily at the start of the follow up period. If one imposes the condition that  $\Gamma_{11} = 1$  and  $\Gamma_{1j} = 0$  for  $j = 2, \dots, r$ , the first column of B gives the effects of the covariates at time  $t = 0$ . These values can easily be compared with the estimates coming from other models, such as the simple Cox model and the gamma frailty model. These restrictions do not resolve the identifiability problems, but further restrictions become quite artificial.

### 2.3 Application to ovarian cancer patients

The data consist of survival data of 358 patients with advanced ovarian cancer originally analyzed by Verweij and van Houwelingen [104]. We have information on three covariates, the diameter of the residual tumor after the operation (coded as 0 up to 4, from small to big) the Karnofsky performance status that measures the patients ability on a scale from 0 to 100 (coded as 0 for value 100, up to 4 for Karnofsky less than 70) and the Figo index which expresses the site of the metastases as III or IV (coded as 0 for Figo=III and 1 for Figo=IV). The median follow up was 24.5 months (1 up to 91) and 266 patients died during the study period. The covariates were treated as linear, following the original analysis of Verweij and van Houwelingen, since their effects are nearly linear. A summary of the data can be found in table A.1.

We fitted a proportional hazards model to the data with estimated coefficients  $\hat{\beta}_{Diam} = 0.200$  ( $se = 0.050$ ),  $\hat{\beta}_{Figo} = 0.542$  ( $se = 0.135$ ) and  $\hat{\beta}_{Karn} = 0.172$  ( $se = 0.053$ ) with a partial log-likelihood of -1382.62 on 3 degrees of freedom.

### 2.3. Application to ovarian cancer patients

A visual test of proportionality was proposed by Grambsch and Therneau

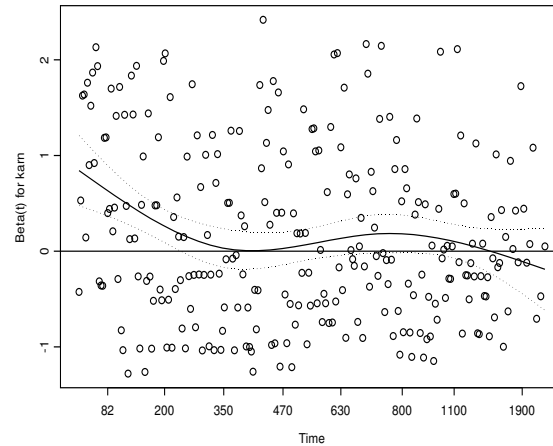


Figure 2.1: Test of proportionality based on scaled Schoenfeld residuals along with a spline smooth with 90% confidence intervals for karn variable.

[39] based on *scaled Schoenfeld residuals*. They suggested plotting *Schoenfeld residuals + Cox model estimate* versus time, or some function of time, as a method of visualizing the nature and extent of non-proportional hazards. In figures 2.1, 2.2 and 2.3 we present the plot of the scaled residuals versus time along with a spline smooth and 90% confidence intervals around it, for each covariate. The time varying effect of the Karnofsky status can be seen. The fit suggests that a low score increases the risk of an event early in time, however, the effect diminishes over time and it goes to 0 after approximately 350 days (figure 2.1). The variable Figo does not show much of a time varying pattern - the smooth spline is more or less constant-, which is also the case for the tumor diameter. The overall likelihood ratio test had a p-value 0.021 indicating departure from proportionality. The test for Karnofsky gave a p-value 0.0071, Figo 0.947 and tumor diameter 0.57.

In our analysis B-splines were used to model the time varying behavior of the covariates. We used second degree splines with the time axis divided into 5 intervals, with four interior knots placed at the end of the first, second,

## Reduced-rank hazard regression

third and fourth year, and the boundary knots at time 0 and at the time of the last event occurred. The B-spline functions are defined in such a way that at  $t = 0$  all the B-spline functions are zero. In the F-matrix, the first function is the constant and the columns 2 to 7 correspond with the six B-spline functions. Since there are only 20 events in the last interval, the last of the six B-splines might be hard to estimate.

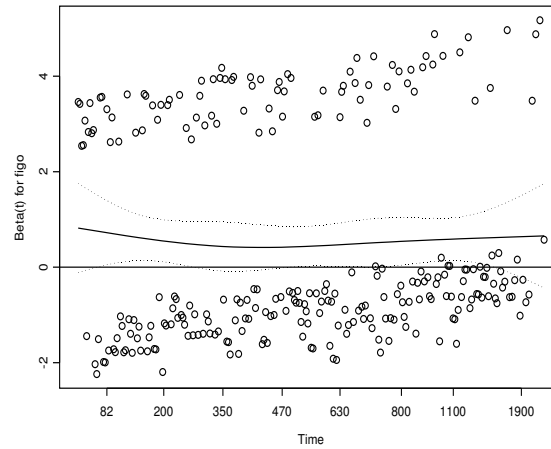


Figure 2.2: Test of proportionality based on scaled Schoenfeld residuals along with a spline smooth with 90% confidence intervals for figo variable.

The Cox model with unrestricted time varying effects was estimated with 21 degrees of freedom and had a partial likelihood of -1365.54. The estimated coefficients are of little interest because they depend very much on the choice of the time functions. The model could be best understood from the plot of the time varying effects of all covariates, which are shown in figure 2.4. The Karnofsky effect follows the pattern suggested by the smoothed scaled Schoenfeld residuals, and drops to 0 after approximately 250 days. On the other hand, the effect of tumor staging (Figo) has big fluctuations as it drops during the first year and then rises again for time up to 1000 days but then drops rapidly again until the last days of the time period, resulting to an irrational pattern that could be due to data overfitting. Confidence intervals

### 2.3. Application to ovarian cancer patients

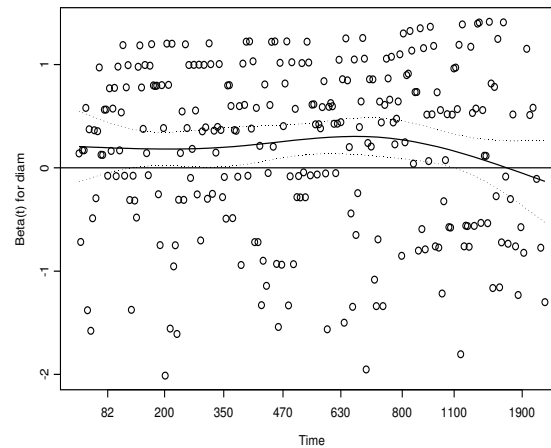


Figure 2.3: Test of proportionality based on scaled Schoenfeld residuals along with a spline smooth with 90% confidence intervals for diam variable.

around the estimated curves were very wide during the last days of follow up (data not shown). The effect of Diameter has smaller fluctuations and the pattern indicates that the effect does not vary much over time. Note the wild behavior of all three curves after the 4th year (1500 days), where the number of events is small, and the behavior of spline functions tend to be erratic near the boundaries.

In a rank 2 model there are a total of 16 parameters to estimate. The  $\Theta$  matrix is parametrized as  $B\Gamma'$  with  $B_{3 \times 2}$  and  $\Gamma_{7 \times 2}$ . The partial likelihood is -1368.19. The resulting model is very similar to the full rank model, as indicated by Figure 2.5, although the behavior of Figo looks a bit smoother, as well as the behavior of all three covariates at the end of follow up. Staging of the cancer drops its effect rapidly at the start of the follow up, but does not reach 0 until 3 years have elapsed, where the number of patients at risk is smaller and the fit is less reliable. The effect of Karnofsky status drops again to 0 after 250 days, indicating an effect consistent with medical knowledge, as one would expect that the Karnofsky performance at the time of diagnosis would be important at a first period but then it would wash away with time.

## Reduced-rank hazard regression

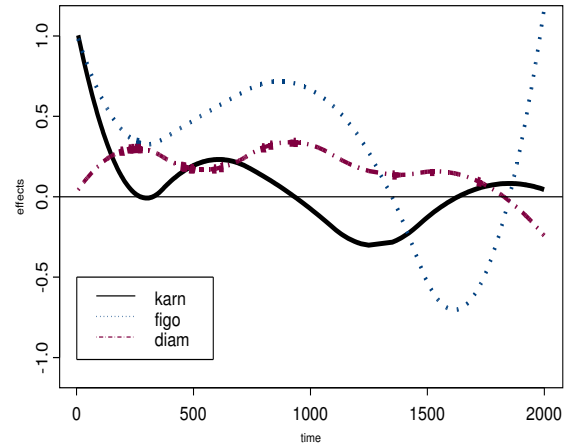


Figure 2.4: Estimated effects of the covariates over time, for the full rank model

The more parsimonious rank 1 model was fitted with partial likelihood -1373.80 on 9 degrees of freedom. In figure 2.6 we see that the effects of Karnofsky and Diameter are very similar, because the time functions are forced to be proportional, and the differences are due to the  $\beta$ 's, which in a simple Cox model were also similar.

From Table 2.2 we learn that the rank=1 model performs best according to the AIC criterion. That is in line with our observation that the full rank and the rank=2 model show some clear signs of over-fitting. This is partly related to our choice of time functions in the F-matrix. We will come back to that in the discussion. Fitting a model with Karnofsky as the only time varying covariate and the rest fixed resulted in a log-likelihood of -1370.79, on 9 degrees of freedom. Although this model seems better in terms of the AIC it is more data driven than the rank=1 model.

We also fitted a gamma frailty model to the data. With the inclusion of a random effect the estimates of the coefficients become larger. We assume a random parameter following a gamma distribution which gave a significant variance 1.11 and estimates  $\hat{\beta}_{Diam} = 0.340$ ,  $\hat{\beta}_{Figo} = 0.782$  and  $\hat{\beta}_{Karn} = 0.326$ .

### 2.3. Application to ovarian cancer patients

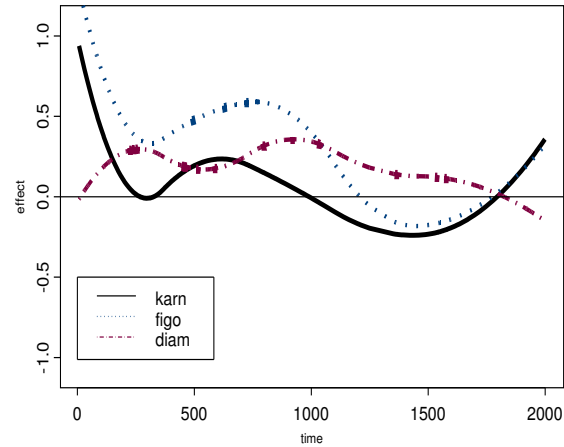


Figure 2.5: Estimated effects of the covariates over time, for rank=2 model

Table 2.2: Partial log-likelihoods of the models along with the degrees of freedom and the Akaike information criterion AIC

	log-lik	df	AIC
Cox PH	-1382.26	3	-1385.26
frailty	-1379.20	?	?
r=1	-1373.80	9	-1382.80
r=2	-1368.19	16	-1384.19
r=3	-1365.54	21	-1386.54

The pseudo-partial likelihood, which is estimated in S-plus software as the log-partial likelihood with the frailty terms integrated out, was -1379.2. Since the partial likelihood also involves the cumulative baseline hazard  $H_0(t)$  as can be seen from formula (4), it is unclear what the degrees of freedom are and, therefore, AIC cannot be computed and the gamma frailty cannot be compared to the other models on this criterion.

## Reduced-rank hazard regression

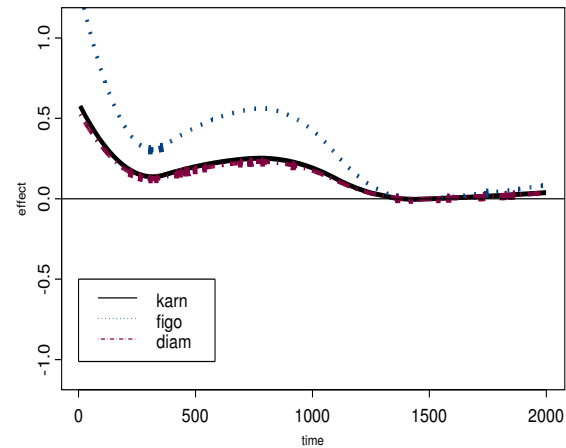


Figure 2.6: Estimated effects of the covariates over time, for rank=1 model

Under the gamma frailty model there is no such thing as a simple additive time varying effect of a specific covariate. In the marginal model of formula (4) there is implicit time varying interaction between the covariates. Therefore, it is not easy to compare the gamma frailty model with additive time varying effects models. An elaboration is given in the next subsection.

### Comparison between the different models

In order to compare the different models we consider the two groups of best and worst prognosis according to the simple Cox model. The best prognosis group contains 17 patients with all covariate at the lowest values, that is 0 for all, and the worst prognosis contains the 8 patients with the highest values for all covariates (Karnofsky =4, Figo= 1 and Diameter =4).

The graphs are presented in Figures 2.7 and 2.8. Observe that the different models give very similar graphs for both groups. Apparently, the models completely agree which are the good prognosis and bad prognosis patients. The shape of the survival curves differ slightly between the models. The gamma frailty slightly deviates from the time varying effects models. In Section 2.2

#### 2.4. Discussion

we expressed the expectation that the rank=1 model and the gamma frailty would be very similar. That is contradicted by the findings of our example. The rank=1 model does not bridge the gap between the full rank model and the gamma frailty model.

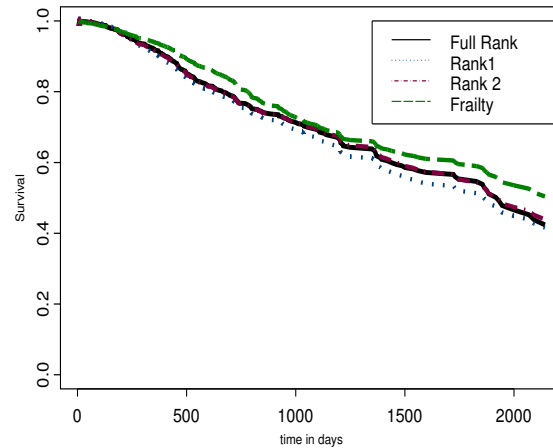


Figure 2.7: Survival for low risk patients under the rank=1, rank=2, full rank and frailty model

#### 2.4 Discussion

We have introduced a family of reduced-rank models in the context of survival data, as a simple device to model time varying effects. Depending on the rank of the structure matrix, reduced-rank models can offer a whole range of models, from the very flexible full rank model to the rigid rank one model, which resembles the structure of a gamma frailty model.

We illustrated, using a small data set, how the different choices of ranks affect the fit of the model, and how they perform compared to a gamma frailty and a full rank model. The covariates in the application were treated as continuous, with the assumption that the time varying effect of all outcomes are identical. We proposed a reduction of the number of free parameters, and



## Reduced-rank hazard regression

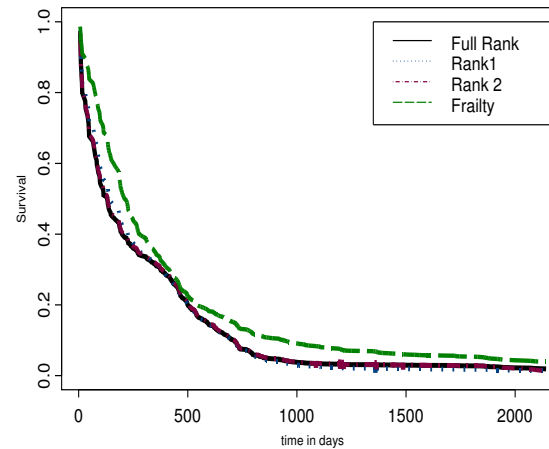


Figure 2.8: Survival for high risk patients under the rank=1, rank=2, full rank and frailty model

choose a rank=1 as the best model according to AIC.

The rank=1 model turns out to be best model from the AIC point of view because it is much more parsimonious than the full model or the rank=2 model. Parsimony could also be achieved by reducing the dimension of the time function space or by penalizing the time varying effects for each covariate as in Verweij and Van Houwelingen [103]. There is no clear-cut answer to what is the best strategy. Nevertheless, using data based functions to model effects of time-varying covariates has downfalls. In our example it might be best to let one covariate, namely Karnofsky, have a time-varying effect and keep the other effects fixed. The choice of Karnofsky is clearly data driven. In data sets with more (or categorical) covariates it might be less clear-cut which covariates show time-varying effects and which not.

The choice of time functions based on residual plots may result in small  $p$ -values but also guarantees over fitting, and breaks the assumption that the model should be specified independently of the data values actually realized. Thus, letting the rank of the model decide for the functional form of the effects could be a useful approach. In any case, reduced-rank models that assume

#### *2.4. Discussion*

similarities between time varying patterns could be one useful way to reduce the number of parameters and to fight over-fitting.

If the fitted model is of rank  $r > 1$  rotations and bi-plots like in factor-analysis might be helpful to find groups of covariates which show similar patterns over time.

We believe that reduced-rank models should be considered when modelling time varying data, not only for their flexibility but also for their simplicity. The estimation is still done via a Newton-Raphson algorithm, which is a familiar and easy technique, as opposed to the more complicated EM algorithm that is used for estimation of frailty models. Fitting can be done using standard software, although the authors intend to publish an R routine for fitting reduced-rank models more efficiently.

Albeit our reduced-rank model was inspired by the gamma frailty model and it has the same flavor of simultaneously converging hazards, the resulting survival curves are different. Understanding this phenomenon and trying to find other parsimonious models that are close to the gamma frailty model is the topic of current research at our department.

## Appendix

The data is based on a sample of size  $n$ . We assume that censoring is not informative and that given  $X_j$  the event and censoring time for the  $j$ th patient are independent. Also, assume that there are no ties between the event times. Let  $t_1 < t_2 < \dots < t_D$  denote the ordered event times and  $X_{(i)}$  the covariate vector associated with the individual whose failure time is  $t_i$ . Define the risk set at time  $t_i$ ,  $R_i$ , as the set of all individuals who are still under study at time just prior to  $t_i$ ,  $F$  is the  $D \times q$  matrix of time functions with  $F_i$  the vector of time functions at time  $i$ . The rank  $r$  model, with maximum rank  $r \leq \min(p, q)$  is

$$h(t_i|X_j) = h_0(t_i) \exp\left(\sum_{k=1}^r (X_j \beta_k)(F_i \gamma_k)\right) \quad (2.6)$$

The partial likelihood is defined as

$$L(\beta, \gamma) = \prod_{i=1}^D \frac{\exp\{\sum_{k=1}^r (X_{(i)} \beta_k)(F_i \gamma_k)\}}{\sum_{j \in R_i} \exp\{\sum_{k=1}^r (X_j \beta_k)(F_i \gamma_k)\}} \quad (2.7)$$

and  $pl = \log L$  The first derivative of the partial log-likelihood with respect to  $\beta$  gives the scores

$$U(\beta_k) = \frac{\partial pl(\beta, \gamma)}{\partial \beta_k} = \sum_{i=1}^D (F_i \gamma_k) [X_{(i)} - \bar{X}_{(i)}]' \quad (2.8)$$

where

$$\bar{X}_{(i)} = \frac{\sum_{j \in R_i} X_j \exp\{\sum_{l=1}^r (X_j \beta_l)(F_i \gamma_l)\}}{\sum_{j \in R_i} \exp\{\sum_{l=1}^r (X_j \beta_l)(F_i \gamma_l)\}} \quad (2.9)$$

and the  $\gamma$  scores are

$$U(\gamma_k) = \frac{\partial pl(\beta, \gamma)}{\partial \gamma_k} = \sum_{i=1}^D ([X_{(i)} - \bar{X}_{(i)}] \beta_k F_i') \quad (2.10)$$

The Hessian matrix is

$$\mathbf{H} = \begin{pmatrix} \frac{\partial^2 pl}{\partial \beta_k \partial \beta_l} & \cdots & \frac{\partial^2 pl}{\partial \beta_k \partial \gamma_l} & \cdots \\ \vdots & \ddots & \vdots & \ddots \\ \frac{\partial^2 pl}{\partial \gamma_k \partial \beta_l} & \cdots & \frac{\partial^2 pl}{\partial \gamma_k \partial \gamma_l} & \cdots \\ \vdots & \ddots & \vdots & \ddots \end{pmatrix}$$

#### 2.4. Discussion

where

$$\frac{\partial^2 pl(\beta, \gamma)}{\partial \beta_k \partial \beta_l} = - \sum_{i=1}^D (F_i \gamma_k) (F_i \gamma_l) Cov_i \quad (2.11)$$

$$\frac{\partial^2 pl(\beta, \gamma)}{\partial \gamma_k \partial \gamma_l} = - \sum_{i=1}^D (\beta'_k Cov_i \beta_l) (F'_i F_i) \quad (2.12)$$

When ( $k \neq l$ )

$$\frac{\partial^2 pl(\beta, \gamma)}{\partial \beta_k \partial \gamma_l} = - \sum_{i=1}^D (F_i \gamma_k) (Cov_i \beta_l) F_i \quad (2.13)$$

and when ( $k=l$ )

$$\frac{\partial^2 pl(\beta, \gamma)}{\partial \beta_k \partial \gamma_l} = \sum_{i=1}^D (X_{(i)} - \bar{X}_{(i)})' F_i - \sum_{i=1}^D (F_i \gamma_k) (Cov_i \beta_l) F_i \quad (2.14)$$

with

$$Cov_i = \frac{\sum_{j \in \mathbf{R}_i} X_j' X_j \exp\{\sum_i (X_j \beta_k) (F_i \gamma_k)\}}{\sum_{j \in \mathbf{R}_i} \exp\{\sum_i (X_j \beta_k) (F_i \gamma_k)\}} - \bar{X}'_{(i)} \bar{X}_{(i)} \quad (2.15)$$

The Hessian matrix is not invertible. However standard errors for estimable functions can be obtained by using a generalized inverse of  $H$ . To get confidence bands around the time varying effects of the estimates we use the delta method. For a specific covariate  $X_1$  the time varying effect is estimated by:  $\hat{y}_1(t) = \sum_{k=1}^r X_1 \beta_k (\sum_{l=1}^d \gamma_{lk} f_l(t))$ . The variance of  $\hat{y}_1(t)$  is given by:

$$var(\hat{y}_1(t)) = DH^{-1}D'$$

where the first term in the right hand side of the equation is a row vector with the derivatives of  $\hat{y}_1(t)$  with respect to  $\beta$  and  $\gamma$ ,

$$D = \left( \frac{\partial \hat{y}_1(t)}{\partial \beta_{11}}, \dots, \frac{\partial \hat{y}_1(t)}{\partial \beta_{pr}}, \frac{\partial \hat{y}_1(t)}{\partial \gamma_{11}}, \dots, \frac{\partial \hat{y}_1(t)}{\partial \gamma_{qr}} \right)$$

