

A method for checking regression models in survival analysis based on the risk score

Jon Ketil Grønnesby and Ørnulf Borgan

April 14, 1996

Abstract

We propose to perform model check for the Cox and Aalen regression models using martingale residual processes grouped after the risk score. Asymptotic distributions of the grouped martingale residual processes are deduced, so both formal and graphical model check can be performed. The method is validated by stochastic simulation. A data example with patients with primary biliary cirrhosis of the liver is discussed.

1 Introduction

Regression models, especially Cox' proportional hazards model, have become increasingly popular in survival analysis. Recently linear alternatives have been proposed; both nonparametric (Aalen, 1980,1989) and semiparametric (Lin and Ying, 1994; McKeague and Sasieni, 1994).

The available methods for checking model fit are, however, still insufficient, even for the Cox model. For linear normal models, residual methods are the primary tools in the assessment of the model fit, see Weisberg (1985) or Cook and Weisberg (1982) for a general treatment. For regression models in survival analysis diagnostic methods based on (martingale) residuals have been proposed, see e.g. Fleming and Harrington (1991), but the interpretation of the individual residuals is difficult. Aalen (1993) suggests to plot grouped martingale residual processes, a related method is due to Arjas (1988). Lin *et al.* (1993) propose methods to check the model based on the maximum deviation from the fitted model. General goodness of fit tests are given e.g. in Schoenfeld (1980) for the Cox model, and by Hjort (1990) for parametric proportional hazards models.

In this paper we discuss methods for checking the overall fit of the regression models based on martingale residual processes grouped after the risk score. We consider both graphical methods and formal goodness of fit tests. By stochastic simulation, we study how well the methods are able to detect error in the model specification. The simulations are supplemented by a real data example of patients with primary biliary cirrhosis (PBC) of the liver (Fleming and Harrington, 1991).

2 Review of regression models for survival data

We use standard counting process notation and results, see Andersen *et al.* (1993) for a general reference. For simplicity we only consider time-fixed covariates, and we assume only right censoring. Let individual i have covariate vector $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^T$ and at risk indicator $Y_i(t)$. Let further $N_i(t)$ count the number of failures for individual i in $[0, t]$.

2.1 The Cox model

The intensity process for $N_i(t)$ is

$$\lambda_i(t) = h_0(t) \exp(\mathbf{z}_i^T \boldsymbol{\beta}) Y_i(t). \quad (1)$$

The nonparametric form of the baseline $h_0(t)$ gives the model a large flexibility even though the modelling of the effects of the covariates is quite restrictive. Notice however that the structure of the Cox model is not retained when covariates are excluded from the model.

The parameter vector $\boldsymbol{\beta}$ is estimated by maximizing Cox' partial likelihood, and the cumulative baseline function $H_0(t) = \int_0^t h_0(s) ds$ is estimated by the Breslow (1972) estimator

$$\hat{H}_0(t) = \int_0^t \frac{dN_{\cdot}(u)}{S^{(0)}(\hat{\boldsymbol{\beta}}, u)}, \quad (2)$$

where $S^{(0)}(\hat{\boldsymbol{\beta}}, u) = \sum_{i=1}^n Y_i(u) \exp(\mathbf{z}_i^T \hat{\boldsymbol{\beta}})$ and $N_{\cdot}(u) = \sum_{i=1}^n N_i(u)$. For a general reference see e.g. Cox and Oakes (1984) or Andersen *et al.* (1993, Sections VII.2-VII.3).

2.2 Nonparametric linear model

Introduce $\mathbf{Y}_i^T(t) = Y_i(t)(1, \mathbf{z}_i^T) = (Y_i(t), z_{i1}Y_i(t), \dots, z_{ip}Y_i(t))$, and let $\mathbf{Y}(t)$ be the *design matrix* with rows $\mathbf{Y}_i^T(t)$. Aalen (1980,1989) proposes a nonparametric linear model with intensity process

$$\lambda_i(t) = \mathbf{Y}_i^T(t) \boldsymbol{\alpha}(t) \quad (3)$$

for $N_i(t)$. In this model, $\boldsymbol{\alpha}(t) = (\alpha_0(t), \dots, \alpha_p(t))^T$ are regression *functions*, not *parameters*. Note that the linear structure does not guarantee that the intensities will be positive (see also Sections 4 and 5).

The vector of integrated regression functions $\mathbf{A}(t)$ is estimated by

$$\hat{\mathbf{A}}(t) = \int_0^t \mathbf{Y}^-(s) d\mathbf{N}(s)$$

where $\mathbf{Y}^-(s)$ is any generalized inverse of $\mathbf{Y}(s)$ and $\mathbf{N}(s) = (N_1(s), \dots, N_n(s))^T$. We will in this paper use only Aalen's original choice for inverse, $\mathbf{Y}^-(s) = (\mathbf{Y}^T(s) \mathbf{Y}(s))^{-1} \mathbf{Y}^T(s)$, though other inverses are possible, see Huffer and McKeague (1991). Inference is only possible when \mathbf{Y} has full rank.

The linear model is still valid when covariates are omitted from the model, at least when the omitted covariate is independent of the others or if it follows a normal distribution, see Aalen (1989). However the interpretation of the baseline term changes.

2.3 Semiparametric linear model

Lin and Ying (1994) propose a semiparametric linear model

$$\lambda_i(t) = \mathbf{Y}_i^T(t) (\alpha_0(t), \boldsymbol{\alpha})^T = (\alpha_0(t) + \mathbf{z}_i^T \boldsymbol{\alpha}) Y_i(t). \quad (4)$$

In contrast to (3), only the baseline is nonparametric. The estimators are deduced in a similar way as the Cox estimators, see Lin and Ying (1994), but we get an explicit solution for the parameter:

$$\hat{\boldsymbol{\alpha}} = \sum_1^n \int_0^\infty \{\mathbf{z}_i(t) - \bar{\mathbf{z}}(t)\} dN_i(t) \left/ \left[\sum_1^n \int_0^\infty Y_i(t) \{\mathbf{z}_i - \bar{\mathbf{z}}(t)\}^2 dt \right] \right., \quad (5)$$

where $\bar{\mathbf{z}}(t) = \sum_{i=1}^n Y_i(t) \mathbf{z}_i / Y_{\cdot}(t)$ with $Y_{\cdot}(t) = \sum_{i=1}^n Y_i(t)$, the total number at risk. Below we will use the model (4) only as a way to group the data according to 'risk scores' in an Aalen model, see Section 3.2, and the discussion below will only concern the Cox and Aalen models.

3 Model fit

The model fit problem consists of determining the coding of the covariates, checking the link function and checking whether the effect of the covariates changes over time. Our method gives an overall model check.

3.1 Martingale residuals, Grouped martingale residuals, the Arjas plot

If $\lambda_i(s)$ is defined according to the models (1) or (3),

$$M_i(t) = N_i(t) - \int_0^t \lambda_i(s) ds$$

is a martingale under the appropriate model. The *martingale residual processes* $\hat{M}_i(t)$ are defined by replacing $\lambda_i(s)ds$ with the respective estimates $d\hat{\Lambda}_i(s)$, yielding respectively

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp(\mathbf{z}_i^T \hat{\beta}) \frac{dN_i(s)}{S^{(0)}(\hat{\beta}, s)} \quad (6)$$

and

$$\hat{M}_i(t) = N_i(t) - \int_0^t \mathbf{Y}_i^T(s) \mathbf{Y}^-(s) d\mathbf{N}(s) \quad (7)$$

for the Cox and Aalen models. Put $\hat{M}_i = \hat{M}_i(\infty)$.

Note that for the Cox model (1), the martingale residuals are only approximate martingales, while for the Aalen model (3) they are exact martingales, see Aalen (1993). Remark further that the sum of the martingale residual processes is equal to 0 for all t for both models (1) and (3).

Fleming and Harrington (1991) discuss extensively the use of the individual martingale residuals \hat{M}_i in the analysis of model fit. Some smoothing procedure (like LOWESS, see Cleveland (1979)) may then be used to assess whether the covariate is appropriately coded. However, these martingale residual methods disguise the *time* or *process* aspect. Aalen (1993) proposes to group the residual processes $\hat{M}_i(t)$. For this purpose, we introduce a $g \times n$ grouping matrix \mathbf{K} where the element $K_{l,k}$ is equal to 1 if individual k is in group l and 0 otherwise, thus $\sum_l K_{l,k} = 1$ for all k . Now the grouped martingale residual process $\mathbf{K}\hat{\mathbf{M}}(t)$ can be graphed against time. A related method is to plot *observed* ($\mathbf{K}\mathbf{N}(t)$) against *expected* ($\mathbf{K}\hat{\Lambda}(t)$) number of events for the grouped individuals, see Arjas (1988). When the model is correct, the curves of the Arjas plot should approximately follow straight lines with unit slope, whilst the curves of the grouped martingale residual process plot should fluctuate around 0.

The grouped martingale residual processes give the absolute model deviations and it is shown on a time scale. The Arjas plot gives the relative model deviations, and the scale is the number of failures. The two methods are thus supplementary. In the literature the grouping is typically done according to one covariate. A more systematic way, suggested in the next subsection, is to group after the risk score.

3.2 Checking deviations from the Cox model

Lin *et al.* (1993) propose a goodness of fit test for the Cox model based on the *risk score* $\mathbf{z}_i^T \hat{\beta}$. They consider the process

$$W(x) = \sum_{i=i} I(\mathbf{z}_i^T \hat{\beta} \leq x) \hat{M}_i. \quad (8)$$

If the model holds, $W(x)$ will fluctuate around 0. In order to check graphically whether the observed fluctuations are compatible with the model, they compare $W(x)$ to some realizations of the

Gaussian limiting process when the model holds. More formal goodness of fit tests (Kolmogorov-Smirnov like tests) can be performed similarly.

When β is one-dimensional, (8) just gives a method of checking the coding of the covariate. When β is of higher dimension we check the overall model fit. If the link is other than proportional this should be detected. However $W(x)$ does not account for the dynamic aspect, so it will not detect whether a covariate effect becomes weaker (or disappears) with time.

As a simple modification of (8), we therefore propose an approach similar to the one used for logistic regression (Hosmer and Lemeshow, 1989). The idea is to group the individuals after the risk score, i.e. replacing $I(\mathbf{z}_i^T \hat{\beta} \leq x)$ with $I(\mathbf{z}_i^T \hat{\beta} \in J_l)$ in (8) for some interval J_l (below J_l will be the quartile groups). Further we consider the martingale residual process $\hat{M}_i(t)$ in the place of the martingale residual \hat{M}_i . For this purpose, we introduce a $g \times n$ grouping matrix \mathbf{K} where $K_{l,k}$ equals $I(\mathbf{z}_k^T \hat{\beta} \in J_l)$.

Because the regression parameters have to be estimated, the martingale residuals for the Cox model are only approximately martingales. The asymptotic distribution is deduced in the Appendix. Under the model (1) and under mild regularity conditions, the grouped martingale residual process $\mathbf{H}_C(\cdot) = \mathbf{K}\hat{\mathbf{M}}(\cdot)$ properly normalized converges to a mean zero multivariate Gaussian process. Further the covariance between the components $H_{J_i}(s)$ and $H_{J_j}(t)$ of $\mathbf{H}_C(\cdot)$ can be estimated by

$$\hat{\sigma}_{ij}(s, t) = \hat{\phi}_{J_i, J_j}(s \wedge t; \hat{\beta}) - \hat{\psi}_{J_i}(s; \hat{\beta})^T \mathcal{I}^{-1}(\hat{\beta}) \hat{\psi}_{J_j}(t; \hat{\beta}), \quad (9)$$

where

$$\begin{aligned} \hat{\phi}_{IJ}(t; \hat{\beta}) &= \int_0^t \frac{S_I^{(0)}(\hat{\beta}, u)}{S^{(0)}(\hat{\beta}, u)} \left(\delta_{IJ} - \frac{S_J^{(0)}(\hat{\beta}, u)}{S^{(0)}(\hat{\beta}, u)} \right) dN(u), \\ \hat{\psi}_J(t; \hat{\beta}) &= \int_0^t \left(\frac{\mathbf{S}_J^{(1)}(\hat{\beta}, u)}{S^{(0)}(\hat{\beta}, u)} - \frac{S_J^{(0)}(\hat{\beta}, u) \mathbf{S}^{(1)}(\hat{\beta}, u)}{(S^{(0)}(\hat{\beta}, u))^2} \right) dN(u) \end{aligned}$$

and $\mathcal{I}(\hat{\beta})$ is the observed information matrix at $\hat{\beta}$, $S^{(0)}(\hat{\beta}, u) = \sum_{i=1}^n Y_i(u) \exp(\mathbf{z}_i^T \hat{\beta})$, $\mathbf{S}^{(1)}(\hat{\beta}, u) = \sum_{i=1}^n Y_i(u) \mathbf{z}_i \exp(\mathbf{z}_i^T \hat{\beta})$, and $S_J^{(0)}(\hat{\beta}, u)$, $\mathbf{S}_J^{(1)}(\hat{\beta}, u)$ are the same entities summed over the subset J of $\{1, \dots, n\}$. Remark that $\sum_{J_i} H_{J_i}(t) = 0$, so we will have to omit one of the elements to make a test. So with $\hat{\Sigma}(t) = \{\hat{\sigma}_{ij}(t, t)\}_{i,j=1}^{g-1}$,

$$T_C(t) = (H_{J_1}(t), \dots, H_{J_{g-1}}(t)) \hat{\Sigma}^{-1}(t) (H_{J_1}(t), \dots, H_{J_{g-1}}(t))^T$$

is approximately χ_{g-1}^2 distributed when the model (1) holds. Below, we will write T_C for $T_C(\infty)$.

3.3 Checking deviations from the nonparametric linear model

For the nonparametric linear model, $\hat{\mathbf{A}}(t)$ depends on t , and the grouping method above cannot be implemented directly. At least two modifications are possible. Either we can choose one t_0 and group according to the corresponding $\mathbf{z}_i^T \hat{\mathbf{A}}(t_0)$, or we can fit the semiparametric linear model (4) and group after the size of $\mathbf{z}_i^T \hat{\alpha}$. The latter method will be adopted below, and the process will be denoted $\mathbf{H}_L(s) = \mathbf{K}\hat{\mathbf{M}}(s)$ where \mathbf{K} is the grouping matrix.

Writing (7) on matrix form, we get

$$\hat{\mathbf{M}}(t) = \int_0^t (\mathbf{I} - \mathbf{Y}(s) \mathbf{Y}^{-}(s)) d\mathbf{N}(s). \quad (10)$$

Here \mathbf{I} is the $n \times n$ identity matrix. When model (3) holds this is a vector valued martingale with predictable variation process

$$\langle \hat{\mathbf{M}} \rangle(t) = \int_0^t (\mathbf{I} - \mathbf{Y}(s) \mathbf{Y}^{-}(s)) \text{diag}(\lambda(s) ds) (\mathbf{I} - \mathbf{Y}(s) \mathbf{Y}^{-}(s)) \quad (11)$$

(see Aalen, 1993). Here diag of a vector \mathbf{a} is the diagonal matrix with \mathbf{a} on the diagonal. Aalen's original proposal for variance estimator is the *optional variation process*, replacing $\lambda(s)ds$ in (11) by

$$d\mathbf{N}(s). \quad (12)$$

Other variance estimators are possible (Hjort, 1993; Grønnesby, 1996), and will be discussed briefly in Section 6.

Since the sum of the grouped martingale residual processes is zero for all t , we introduce $\bar{\mathbf{H}}_L(t)$ as the vector of the $g-1$ first terms. $\bar{\mathbf{H}}_L(t)$ will be an asymptotic Gaussian martingale with estimated variance process

$$\hat{\Sigma}_L(t) = \widehat{\text{VAR}}\bar{\mathbf{H}}_L(t) = \bar{\mathbf{K}}\widehat{\text{VAR}}\hat{\mathbf{M}}(t)\bar{\mathbf{K}}^T. \quad (13)$$

Here $\widehat{\text{VAR}}\hat{\mathbf{M}}(t)$ is calculated by putting (12) into the formula (11). Now

$$T_L(t) = \bar{\mathbf{H}}_L(t)^T \hat{\Sigma}_L^{-1}(t) \bar{\mathbf{H}}_L(t)$$

is asymptotically χ_{g-1}^2 -distributed. Below, we write T_L for $T_L(\infty)$.

4 Simulation of the goodness of fit methods

In this section, the methods described in Section 3 are explored in simulation experiments. We simulate random censoring with constant rate such that 50 % fail.

When the Cox and linear models are fitted to the same dataset, it is not at all obvious that the same covariate coding gives the best fit for both models. In addition to the original covariate coding, we therefore also show results for transformed covariates in the tables below.

One evident way to choose cutpoints for the grouping after risk score is to assign approximately the same number of individuals in each group. Below we will divide into $g = 4$ groups. In the analyses of the power of T_L and T_C , the nominal level of the tests will be 0.05. All entries in the tables are based on 1000 simulations.

4.1 Simulated Cox data

The simulations are done with constant baseline $h_0(t) \equiv 1$, and 2 standard normally distributed covariates and covariate effect $\beta_1 = \beta_2 = \beta$. Figure 1 displays a typical outcome of the \mathbf{H}_L - and \mathbf{H}_C -processes, pointwise confidence intervals for the lower quartile risk score group and the corresponding Arjas plot for $\beta = 1$ and $n = 100$. Since the rates are constant, i.e. the survival and censoring times are exponential, there will be some extreme (censored) survival times. All figures therefore only show the processes until 10% remain at risk.

The grouped martingale residual processes and the Arjas plot clearly show that the linear model does not fit, at least for $t \leq 0.4$. The true intensities are underestimated for the groups with highest and lowest risk score, as one could expect. Note in particular from the Arjas plot the initial negative number of expected failures, and thereby negative estimated integrated intensity, for the lowest risk score group for the linear model. For this dataset, more than half of all failures before 0.4 are due to the highest risk score group and the intensity from these high covariate values are extrapolated to lower values. The linear structure of the Aalen model has no restrictions to force the data only to admit positive intensities. The insufficient fit for the linear model is confirmed by the goodness of fit tests $T_L = 11.98$ (p-value = .007) and $T_C = 2.38$ (p-value = .497).

Table 1 shows the simulated power for T_L and T_C for different values of β . For the fit of the linear model, both untransformed and exponentially transformed covariates are displayed.

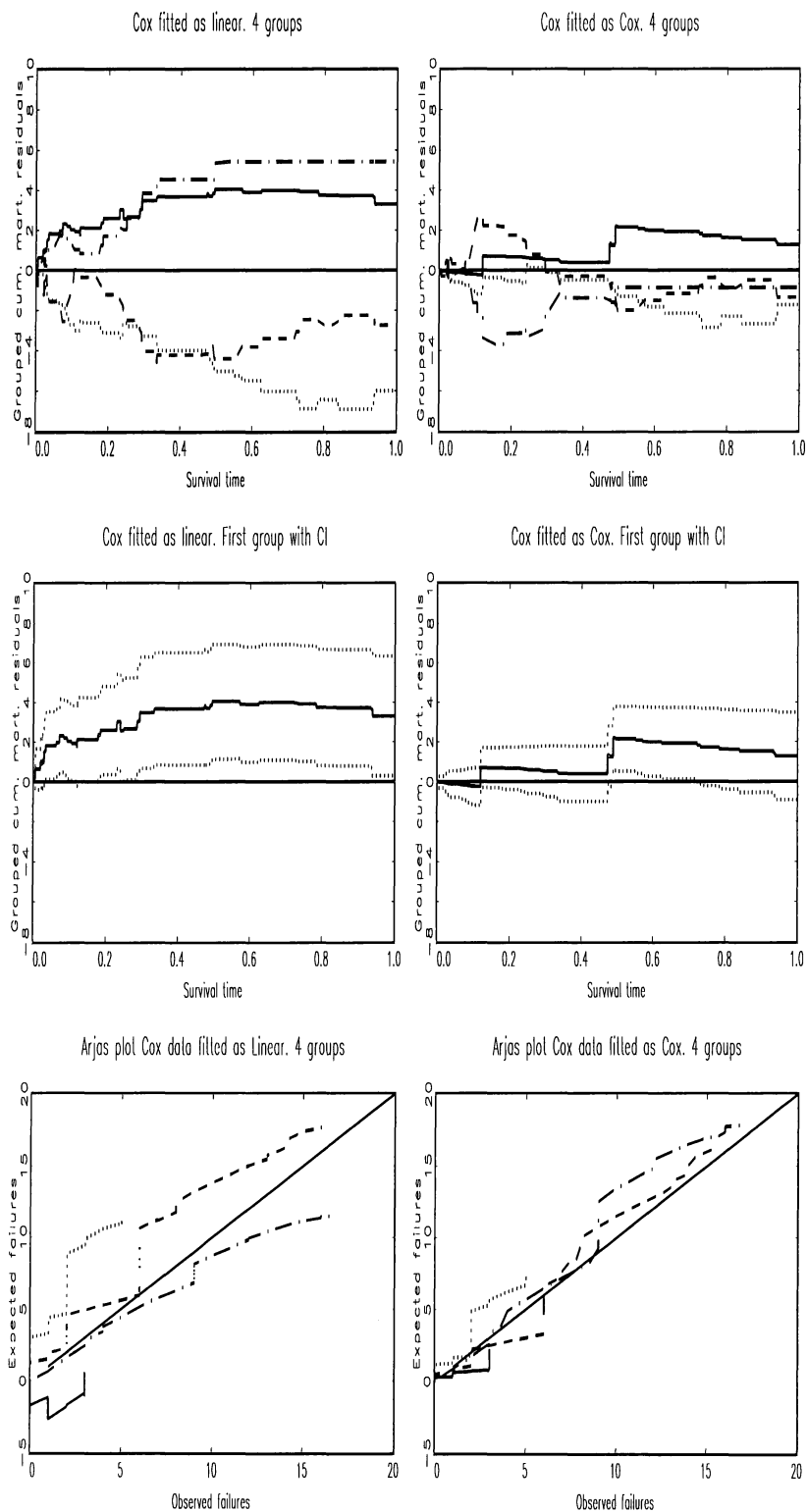


Figure 1: Cox data with 2 normal covariates fitted as linear (left) and Cox (right). Upper curves: estimated grouped martingale residual (H_{\cdot}) processes (All 4 categories). Lower curves: the corresponding Arjas-plot. Middle curves: estimated H_{\cdot} -processes for group I with confidence interval. Legend for the upper and lower figures. Solid: group I (lowest risk score); dotted: group II; stiped: group III; dotted/stiped: group IV

Table 1: Simulated power for T_L and T_C for some Cox datasets

β	$n = 100$			$n = 200$		
	Cox	Linear	Linear transformed	Cox	Linear	Linear transformed
0	.064	.037	.037	.053	.049	.049
.1	.059	.048	.039	.051	.051	.051
.25	.065	.051	.058	.066	.065	.089
.5	.060	.103	.050	.059	.200	.095
.75	.064	.290	.063	.060	.616	.096
1	.071	.571	.072	.050	.953	.120
2	.061	.999	.422	.046	1.000	.628

The obtained level for T_C is slightly too high, but still the approximation is fair. Simulations where we fix the cutpoints to the correct covariate fractiles (not displayed) reduce the achieved level, but it is still higher than the nominal.

For small β , the linear and proportional models are quite similar, even for untransformed covariates. So the power of T_L is monotonously increasing with β , but is still quite low, even when the sample size is 200. With β as high as 1, the fitted models are quite different, because the hazards between the individuals differ so much.

Exponentially transformed covariates yield substantially better fit of the linear model, i.e. lower power of T_L . Only for very high β , the fit is unsatisfactory. Exponential transformation brings the covariates 'on the same scale' in the two models, so the differences between linear and proportional models are smaller, and the negative estimated intensities are avoided for moderate β .

4.2 Simulated linear data with constant effects

In this subsection, the data are drawn from the linear model (4) with $\alpha_0, \alpha_1, \alpha_2$ constants, and further $\alpha_1 = \alpha_2 = \alpha$. It then suffices to vary the baseline because we get essentially the same model by scaling up all the effects, so α is set to 1.

The simulation in Figure 2 is done with two standard lognormally distributed covariates and $n = 100$. We put $\alpha_0(t) \equiv 0$ to maximize the deviation between linear and Cox models. Again we display the grouped martingale residual processes, pointwise confidence interval for the lowest risk score group and Arjas plot.

Here $T_C = 7.83$ (p-value = .050) and $T_L = 2.22$ (p-value = .528), but for the fitted Cox model only the martingale residual process for the lowest risk score group seems to deviate consistently from 0. The Cox model initially overestimates the intensities for the lowest risk score group. This is consistent with the findings in the previous subsection.

As an alternative to the asymptotic interval for H_C based on (9) one could bootstrap the pointwise confidence intervals, either parametrically or nonparametrically. These confidence intervals yield similar results as the asymptotic ones and are therefore not displayed.

Table 2 shows the power of the tests when the Cox and linear models are fitted to simulated linear data with different values of α_0 , and otherwise the setup as in Figure 2. We have also fitted a Cox model with log-transformation of the covariates.

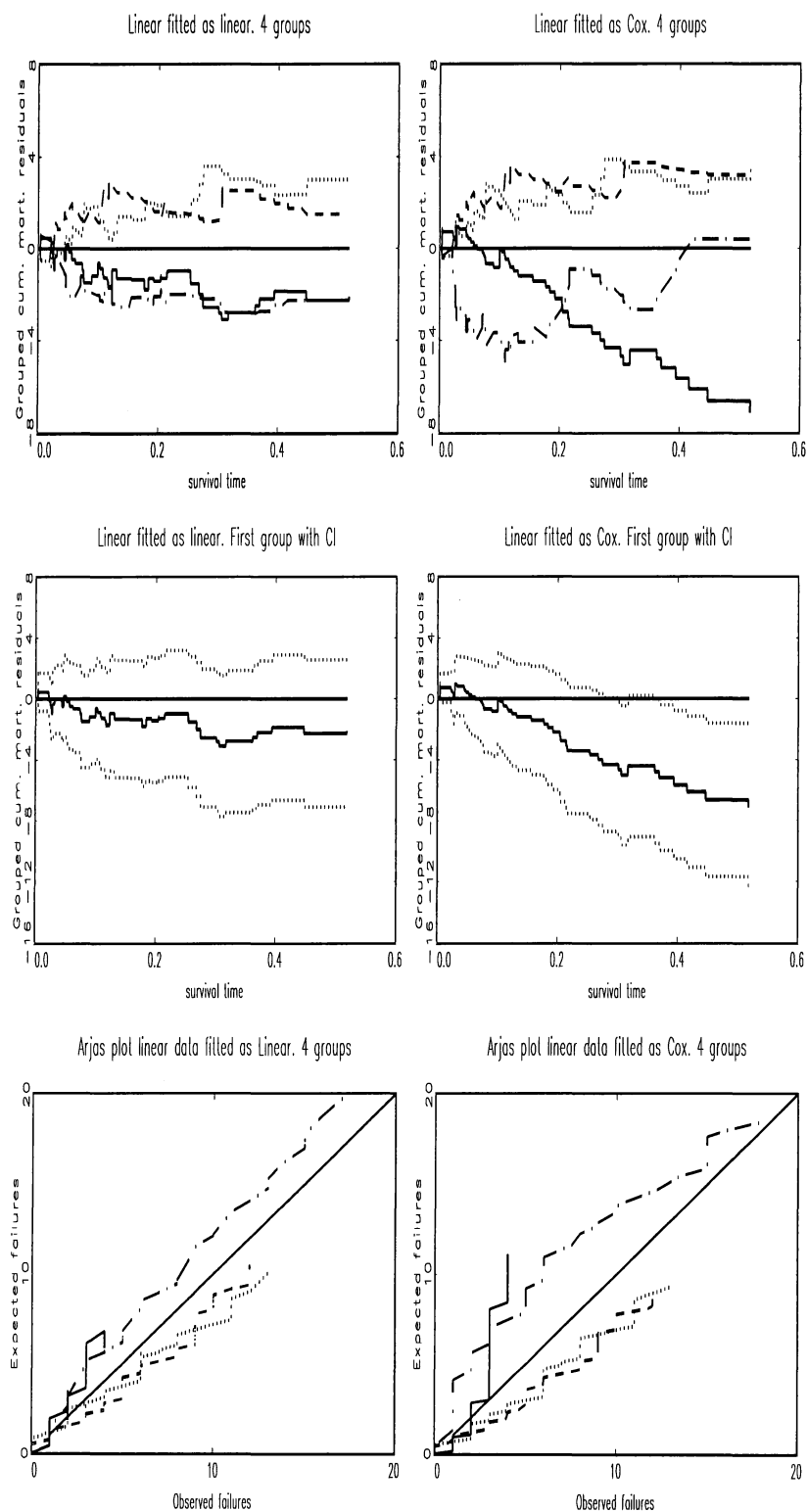


Figure 2: Data are linear with 2 standard lognormal covariates fitted as linear (left) and Cox (right). Upper curves: estimated grouped martingale residual (\mathbf{H}_t) processes (All 4 categories). Lower curves: the corresponding Arjas-plot. Middle curves: estimated \mathbf{H}_t -processes for group I with confidence interval.

Legend. Solid: group I (lowest risk score); dotted: group II; stiped: group III; dotted/stiped: group IV

Table 2: Simulated power for T_L and T_C for some linear data

α_0	$n = 100$			$n = 200$		
	Linear	Cox	Cox transformed	Linear	Cox	Cox transformed
0	.045	.282	.082	.043	.628	.052
.25	.037	.223	.072	.051	.485	.062
.5	.048	.178	.080	.048	.408	.052
.75	.038	.158	.086	.049	.316	.068
1	.040	.143	.078	.052	.270	.075

The achieved level of the T_L -test is systematically slightly below the nominal level. Table 2 further indicates that the Cox model fits well with logtransformed covariates, even with untransformed covariates the fit is fair when $n = 100$.

4.3 Simulation of a true linear data with vanishing effects

One advantage of the linear model is that it permits time varying covariate effects. The model is therefore particularly useful when the effect of covariates vanish after some time. The standard Cox model with time-fixed covariates does not cover this situation, but if the model is extended to time-dependent covariates, the Cox structure can be restored. In practice, however, it is often difficult to detect time-dependent covariate effects for the Cox model.

We choose to let the effect of the first covariate vanish at about the median failure time, and the results are displayed in Figure 3 for $\alpha_0 = 0$ and in Table 3 for different values of α_0 .

Table 3: Simulated power for T_L and T_C for some simulated linear data sets with effect of the first covariate vanishing after some t_0 .

α_0	$n = 100$			$n = 200$		
	Linear	Cox	Cox transformed	Linear	Cox	Cox transformed
0	.029	.323	.072	.053	.692	.046
.25	.033	.234	.055	.046	.490	.064
.5	.042	.189	.078	.045	.374	.076
.75	.048	.145	.088	.040	.302	.071
1	.041	.115	.073	.049	.245	.105

The difference from Section 4.2 is surprisingly small; one could expect a higher power for T_C because the Cox model does not allow a change in the relative risks. But we only achieve higher simulated power for the smallest values of α_0 . The Cox model still fits very well for logtransformed covariates. Some of the reason for the low power for the largest α_0 values is that fewer are at risk after t_0 than in Section 4.2.

The underestimation of the level for T_L is slightly clearer than for Table 2.

5 Real data example: the PBC data

312 patients with primary biliary cirrhosis (PBC) of the liver participated in a double-blind randomized clinical trial at the Mayo clinic comparing survival of patients receiving D-penicillamine or placebo. Liver transplantation was regarded as censoring. Since the drug proved to be ineffective, we can consider the data to provide a description of the 'natural history' of PBC. The data have been extensively analysed using Cox regression by Fleming and Harrington (1991), and we use the data as cited there with corrections for individuals 107 and 253 (explained in op. cit., pages 186-89). We will only consider the covariates of their final model: age, albumin (a protein in

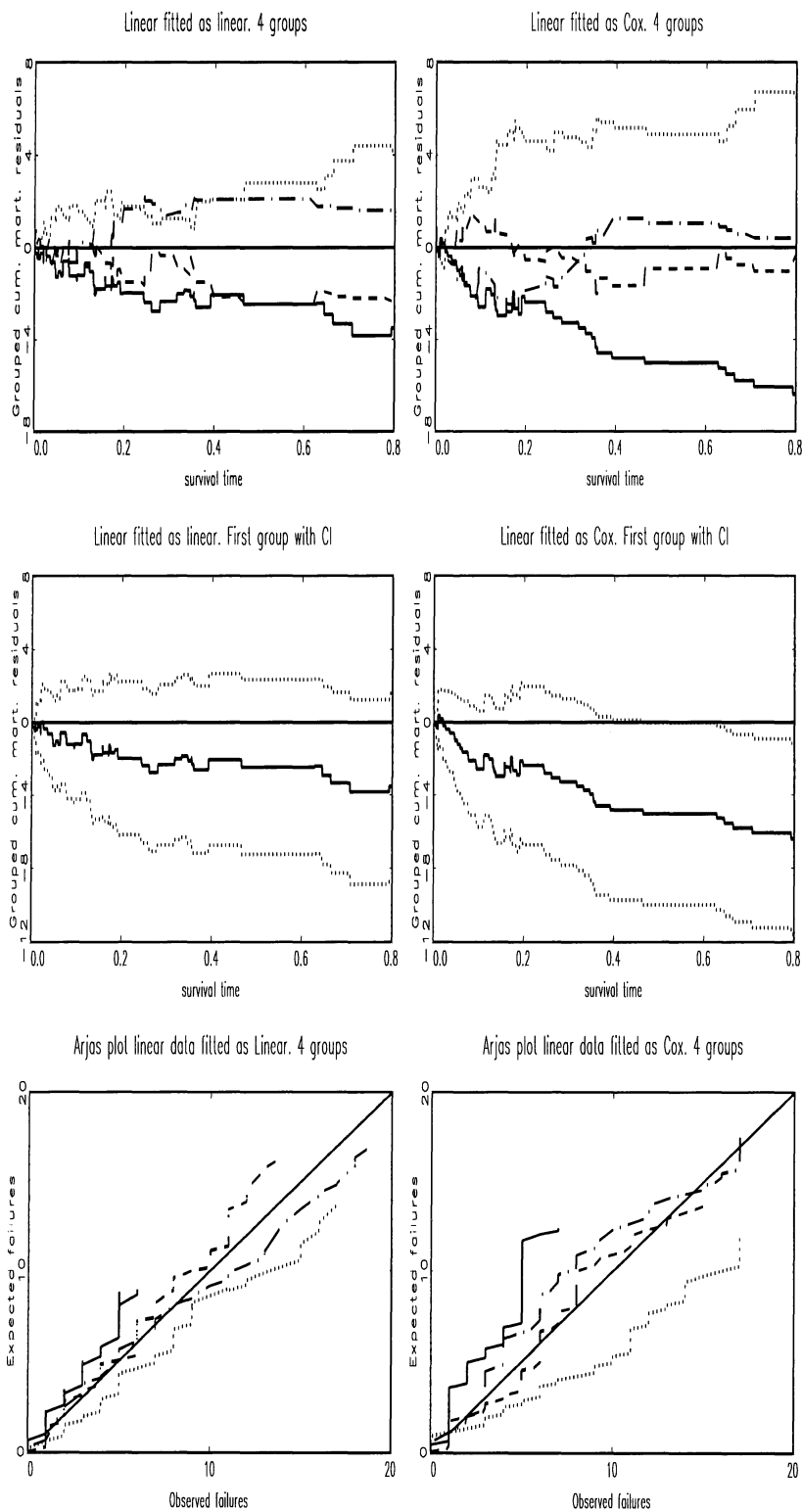


Figure 3: Model is linear with one vanishing effect and 2 standard lognormal covariates fitted as linear (left) and Cox (right). Upper curves: estimated grouped martingale residual (\mathbf{H}_t)-processes (All 4 categories). Lower curves: the corresponding Arjas-plot. Middle curves: estimated \mathbf{H}_t -processes for group I with confidence interval.

Legend. Solid: group I (lowest risk score); dotted: group II; stiped: group III; dotted/stiped: group IV (highest risk score)

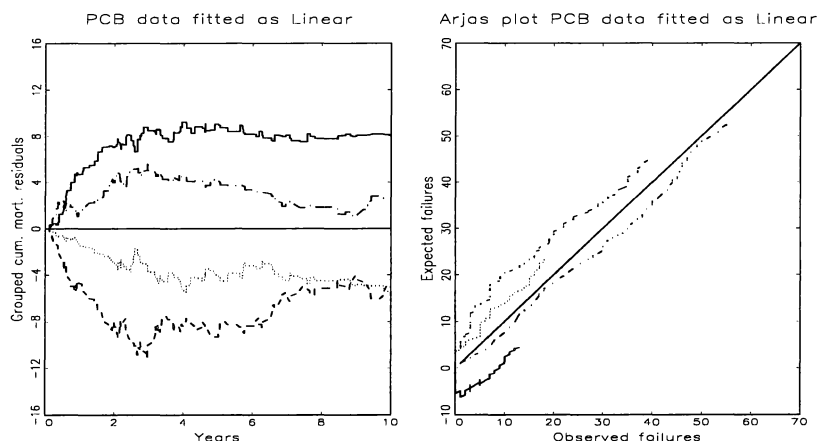


Figure 4: The PBC-data fitted as linear (model with only protime), and grouped by protime. Left: Grouped martingale residual process. Right: Arjas plot.

Legend. Solid: group I; dotted: group II; stipled: group III; dotted/stipled: group IV

the blood), bilirubin (liver bile pigment), edema (swelling caused by excess fluid in subcutaneous tissue) and prothrombin time (time until the blood sample starts coagulation, will be denoted protime). Edema is categorical (0 no edema and no diuretic therapy; 1 edema despite diuretic therapy; .5 either edema without diuretic therapy or edema resolved with diuretic therapy), while the other covariates are continuous.

We first fit a nonparametric linear model covariate-wise, and then discuss the fit of the final linear and proportional models.

5.1 Covariate-wise fit for nonparametric linear model

The fitting of the covariates protime and albumin turns out to be difficult. Both the original coding, and various covariate transformations have been tried out. All these models yield negative estimated intensities for the lowest risk group. As an illustration, we display in Figure 4 the grouped martingale residual processes and the Arjas plot when a linear model with only protime untransformed is fitted and grouped into quartile groups by protime.

The numbers of deaths during the first 2 years are 1, 0, 4 and 28 in the four protime groups respectively. Thus there are far too many failures for the individuals with highest protime values, and the initial rate is estimated almost exclusively from these high covariate values. As in Section 4.1, there is an initial underestimation of the rates for the lowest and the highest protime values, for group 1 we even have negative estimated intensities. These negative estimated intensities persist for the lowest risk scores when other covariates are included in the model. After about 2 years, the linear model fits quite well.

For the final linear model, the covariates age and bilirubin are as given originally; and edema dichotomized (.5 pooled together with 0). Further albumin is coded as zero for the highest half, and then linear; protime is zero for the lowest half, then linear. We have also included a first order interaction term between protime and albumin.

5.2 Overall fit

The final Cox model is as in Fleming and Harrington (1991) with albumin, bilirubin and protime logtransformed. The fit of the linear model and the Cox model is evaluated in Figure 5 by the martingale processes and the Arjas plots. The risk score is divided into quartile groups.

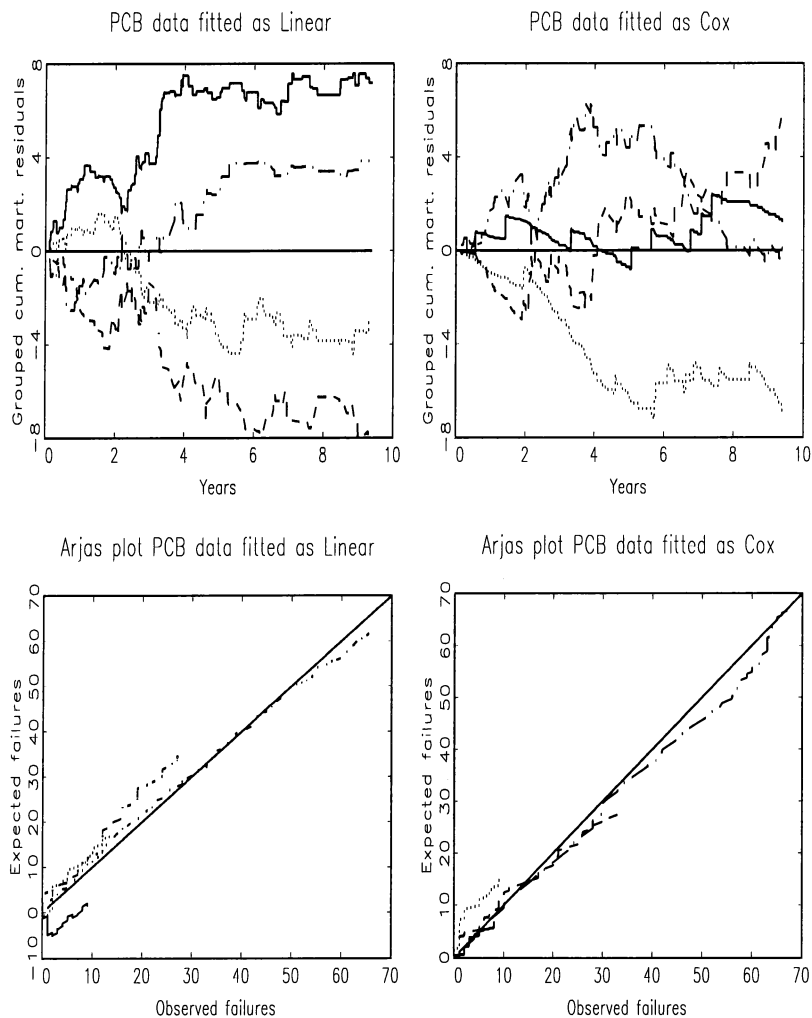


Figure 5: Linear (left) and Cox (right) models fitted to the PCB data. The grouped martingale residuals (upper) and the Arjas plots (lower) are shown. Legend. Solid: group I; dotted: group II; stiped: group III; dotted/stiped: group IV

Measured by the martingale residuals alone, the fit of both models is good (Figure 5). The goodness of fit tests take the values $T_C = 4.68$ (p-value = .197) and $T_L = 6.90$ (p-value = .075) respectively. Note, however, that the linear model suffers from the negative estimated intensities for the individuals with the lowest risk scores.

The martingale residual processes for the Cox model are in the same order of magnitude as the ones for the linear model. The asymptotic pointwise confidence intervals (not displayed) do not initially contain 0 for the two highest risk score groups, which indicate that the fit is not perfect. For the second group, the lack of fit persists over almost the whole timespan. We have not pursued this, but it may be due to the fact that for two of the covariates (edema and log(protime)), the assumed proportionality is questionable, see above.

Figure 6 displays the estimated integrated regression functions of the final Aalen model.

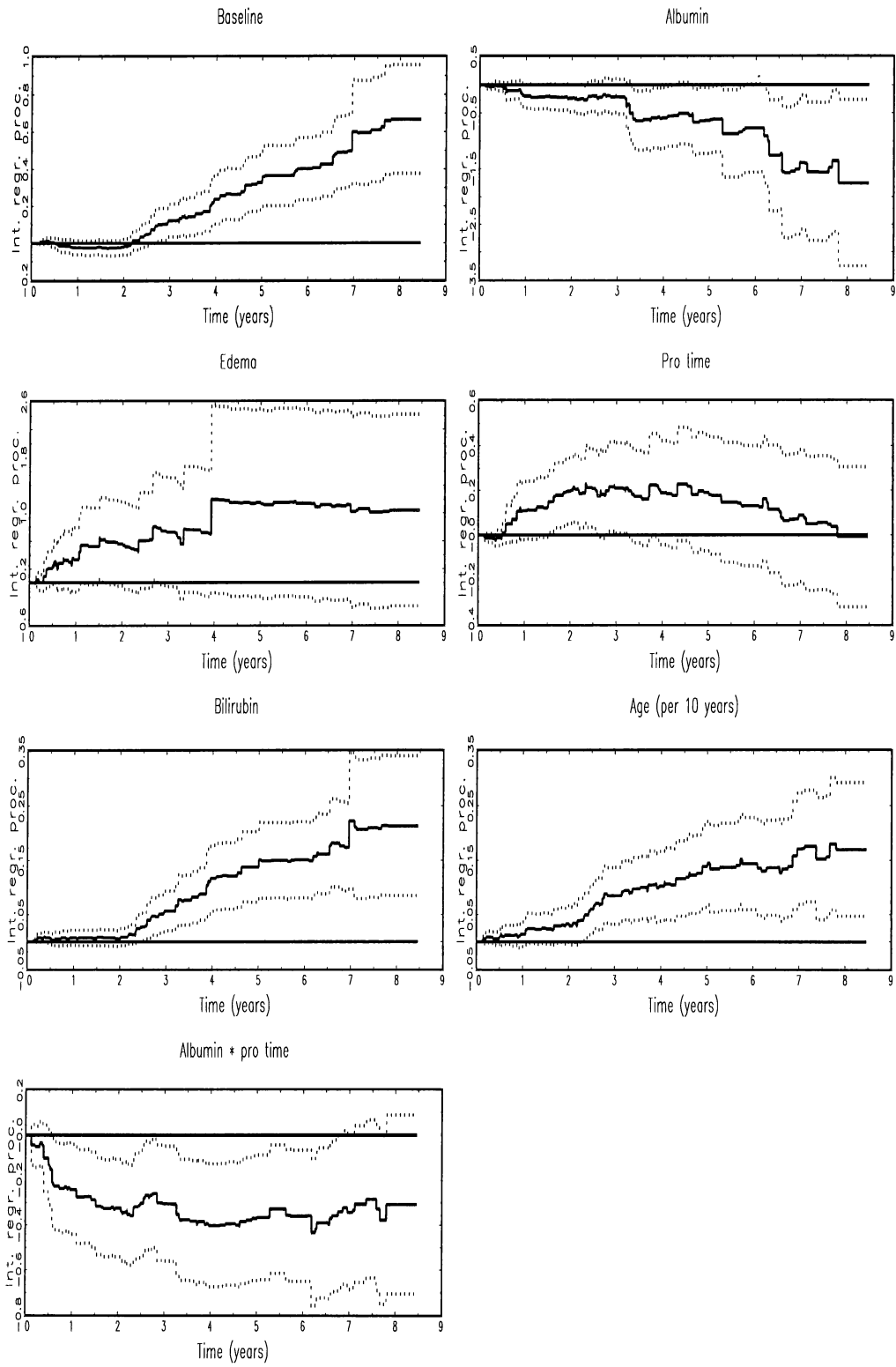


Figure 6: The cumulative regression functions with confidence interval for the final linear model

Note in particular that the effect of protime and the interaction term albumin by protime seem to be vanishing after about 2 years. There is a similar tendency for edema (though less clear), but note that only a very few individuals with edema despite diuretic therapy are left beyond 2 years. This may explain the lack of proportionality for protime and edema in the Cox model, noted by Fleming and Harrington (1991, page 191).

6 Discussion

The martingale residual processes and Arjas plots grouped after the risk score seem promising for checking the model fit for regression models in survival analysis. To group into equal size as we have done, may be suitable for a general model check without specific alternatives in mind. There may, however, be better ways to select the cutpoints. In particular for discriminating between semiparametric linear and proportional models, an alternative may be to estimate a linear baseline function and then choose two cutpoints such that the linear and proportional intensities are equal. As an alternative to the omnibus chisquared test proposed here one could consider weighted tests. For $g = 4$ groups an example is the weighted test $\sum_{i=1}^4 w_i H_i$ for the weights $(1, -1, -1, 1)$.

For moderate covariate effects, it may be difficult to discriminate between the models. The models can therefore be used as supplementary analysis for each other. If we are doing a Cox analysis, an alternative linear model can be used to detect nonproportionality and time dependent effects of the covariates. This point was also noted by Mau (1986).

The negative estimated intensities obtained when the data sets in Sections 4.1 and 5 are fitted by a nonparametric linear model is an unpleasant feature which reduces the applicability of the latter model. One another problem with the linear model, is that we do not have a deviance type measure for comparing the fit for different linear models. On the other hand the estimates and the model fit can be interpreted locally, so contrary to the Cox model the initial behaviour is not compensated by the behaviour at later times. When the Cox model is fitted, it can be difficult to detect model deviations.

The negative estimated intensities also creates problems in using alternative variance estimators to the optional variation process, see Section 3.3. In the spirit of Grønnesby (1996), we could use $\mathbf{Y}(t)d\hat{\mathbf{A}}(t)$ in the place of (12) into the equation (11) and hope for better results. However, this method yields negative estimated variation process for the martingale residuals for those individuals with negative estimated integrated hazards.

Appendix: Asymptotic distribution of the grouped martingale residual processes for the Cox model

The regularity conditions will not be stressed here, they are the same as for the asymptotic normality of the Cox estimator and are given in Andersen *et al.* (1993). We write:

$S^{(0)}(\boldsymbol{\beta}, u) = \sum_{i=1}^n Y_i(u) \exp(\mathbf{z}_i^T \boldsymbol{\beta})$, $\mathbf{S}^{(1)}(\boldsymbol{\beta}, u) = \sum_{i=1}^n Y_i(u) \mathbf{z}_i \exp(\mathbf{z}_i^T \boldsymbol{\beta})$ and analogously $S_J^{(0)}(\boldsymbol{\beta}, u)$ and $\mathbf{S}_J^{(1)}(\boldsymbol{\beta}, u)$ for the same entities summed over a subset J of $\{1, \dots, n\}$. Further define $\hat{\Lambda}_i(\boldsymbol{\beta}, t) = \int_0^t \exp(\mathbf{z}_i^T \boldsymbol{\beta}) Y_i(u) d\hat{H}_0(u)$, where $\hat{H}_0(u)$ is the Breslow-estimator (2). Similar notation is used when the true value $\boldsymbol{\beta}$ of the regression parameters is replaced by its estimator $\hat{\boldsymbol{\beta}}$.

The standardized martingale residuals aggregated over a set J may then be written

$$\frac{1}{\sqrt{n}} \hat{M}_J(t) = \frac{1}{\sqrt{n}} \sum_{i \in J} \hat{M}_i(t) = \frac{1}{\sqrt{n}} \sum_{i \in J} \{N_i(t) - \hat{\Lambda}_i(\boldsymbol{\beta}, t)\} + \frac{1}{\sqrt{n}} \sum_{i \in J} \{\hat{\Lambda}_i(\boldsymbol{\beta}, t) - \hat{\Lambda}_i(\hat{\boldsymbol{\beta}}, t)\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \left(\delta_{iJ} - \frac{S_J^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u) + \frac{1}{\sqrt{n}} \int_0^t \left(\frac{S_J^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} - \frac{S_J^{(0)}(\hat{\beta}, u)}{S^{(0)}(\hat{\beta}, u)} \right) \sum_{i \in J} dN_i(u). \quad (14)$$

Here $\delta_{iJ} = 1$ if $i \in J$, else $\delta_{iJ} = 0$.

By a first order Taylor expansion, the second term in (14) is asymptotically equivalent with

$$\sqrt{n}(\hat{\beta} - \beta)^T \int_0^t \frac{\mathbf{S}^{(1)}(\beta, u) S_J^{(0)}(\beta, u) - S^{(0)}(\beta, u) \mathbf{S}_J^{(1)}(\beta, u)}{S^{(0)}(\beta, u)^2} \frac{dN_i(u)}{n}.$$

We introduce $s^{(0)}$ and $\mathbf{s}^{(1)}$ for the uniform limits in probability of $1/n$ times $S^{(0)}$ and $\mathbf{S}^{(1)}$, $s_J^{(0)}$ and $\mathbf{s}_J^{(1)}$ are defined similarly. Then this is again asymptotically equivalent with

$$\sqrt{n}(\hat{\beta} - \beta)^T \int_0^t \frac{\mathbf{s}^{(1)}(\beta, u) s_J^{(0)}(\beta, u) - s^{(0)}(\beta, u) \mathbf{s}_J^{(1)}(\beta, u)}{s^{(0)}(\beta, u)^2} s^{(0)}(\beta, u) h_0(u) du.$$

But $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically equivalent with

$$\Sigma_{\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \left(\mathbf{z}_i - \frac{\mathbf{S}^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u)$$

where Σ_{β} is the asymptotic information matrix for the Cox model (see Andersen *et al.*, 1993, p 497), so it follows that $\frac{1}{\sqrt{n}} \hat{M}_J(t)$ has the same asymptotic distribution as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \left(\delta_{iJ} - \frac{S_J^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u) - \psi_J(t; \beta)^T \Sigma_{\beta}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\infty} \left(\mathbf{z}_i - \frac{\mathbf{S}^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u) \quad (15)$$

where

$$\psi_J(t; \beta) = \int_0^t \left(\frac{\mathbf{s}_J^{(1)}(\beta, u)}{s^{(0)}(\beta, u)} - \frac{s_J^{(0)}(\beta, u) \mathbf{s}^{(1)}(\beta, u)}{(s^{(0)}(\beta, u))^2} \right) s^{(0)}(\beta, u) h_0(u) du.$$

The data are partitioned into g groups. For any groups I, J :

$$\begin{aligned} & \left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \left(\delta_{iI} - \frac{S_I^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u), \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^t \left(\delta_{jJ} - \frac{S_J^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_j(u) \right\rangle (t) = \\ & \frac{1}{n} \int_0^t \frac{S_I^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \left(\delta_{IJ} - \frac{S_J^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) S^{(0)}(\beta, u) h_0(u) du \xrightarrow{\mathcal{P}} \\ & \int_0^t \frac{s_I^{(0)}(\beta, u)}{s^{(0)}(\beta, u)} \left(\delta_{IJ} - \frac{s_J^{(0)}(\beta, u)}{s^{(0)}(\beta, u)} \right) s^{(0)}(\beta, u) h_0(u) du = \phi_{IJ}(\beta, t). \end{aligned}$$

Here $\delta_{IJ} = 1$ if $I = J$, else $\delta_{IJ} = 0$.

Further:

$$\left\langle \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \left(\delta_{iI} - \frac{S_I^{(0)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_i(u), \frac{1}{\sqrt{n}} \sum_{j=1}^n \int_0^t \left(\mathbf{z}_j - \frac{\mathbf{S}^{(1)}(\beta, u)}{S^{(0)}(\beta, u)} \right) dM_j(u) \right\rangle (t) \xrightarrow{\mathcal{P}} \psi_I(t; \beta).$$

From the above results it follows that $n^{-1/2}(\hat{M}_{J_1}, \dots, \hat{M}_{J_g})^T$ converges to a mean zero multivariate Gaussian process $\mathbf{U} = (U_1, \dots, U_g)^T$. The covariance matrix between $\mathbf{U}(s)$ and $\mathbf{U}(t)$, denoted $\Sigma(s, t)$ has (i, j) -th entry:

$$\sigma_{ij}(s, t) = \text{Cov}(U_i(s), U_j(t)) = \phi_{J_i, J_j}(s \wedge t; \beta) - \psi_{J_i}(s; \beta)^T \Sigma_\beta^{-1} \psi_{J_j}(t; \beta). \quad (16)$$

For estimation of the covariance matrix, replace $s^{(k)}(\beta, u)$ by $S^{(k)}(\hat{\beta}, u)/n$; $s^{(k)}(\beta, u)h_0(u)du$ by $dN(u)/n$ and Σ_β by $n^{-1}\mathcal{I}(\hat{\beta})$, where $\mathcal{I}(\hat{\beta})$ is the observed information matrix at $\hat{\beta}$.

Acknowledgement We want to thank Odd O Aalen for valuable comments.

References

- Aalen O.O. (1980). A model for nonparametric regression analysis for counting processes. *Springer Lect Notes in Statist.* **2**, 1–25.
- Aalen O.O. (1989). A linear model for the analysis of life times. *Statist in Med* **8**, 907–925.
- Aalen O.O. (1993). Further results on the non-parametric linear regression model in survival analysis. *Statist in Med* **12**, 1569–1588.
- Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Verlag, New York.
- Arjas E. (1988). A graphical method for assessing goodness of fit in Cox's proportional hazards model. *JASA* **83**, 204–212.
- Barlow W.E. and Prentice R.L. (1988). Residuals for relative risk regression *Biometrika* **75**, 65–74.
- Breslow N.E. (1972). Discussion of the paper by D.R. Cox. *J Roy Statist Soc B* **34**, 216–217.
- Cleveland W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *JASA* **74**, 829–836.
- Cook R.D., Weisberg S. (1982). *Residuals and Influence in regression*. Chapman and Hall.
- Cox D.R., Oakes D. (1984). *Analysis of Survival Data*. Chapman and Hall.
- Fleming T.R., Harrington D.P. (1991). *Counting processes and survival analysis*. Wiley, New York.
- Grønnesby J.K. (1996). Testing covariate effects in Aalen's linear regression model. *Scand J Statist* **23**, xxx–yyy (in press).
- Hjort N.L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Ann Statist* **18**, 1221–1258.
- Hjort N.L. (1993). Variances of three estimators in Aalen's linear hazard rate regression model. *Manuscript*
- Hosmer D.W., Lemeshow S. (1989). *Applied logistic regression*. Wiley, New York.
- Huffer F.W., McKeague I.W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J Amer Statist Assoc* **86**, 114–129.
- Lin D.Y., Ying Z. (1994) Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–72.
- Lin D.Y., Wei L.J., Ying Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557–72
- Mau J. (1986). On a graphical method for the detection of time-dependent effects of covariates in survival analysis. *Appl Statist* **35**, 245–255.
- McKeague I.W., Sasieni P.D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–14.
- Schoenfeld D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67**, 145–53.
- Therneau T.M., Grambsch P.M. and Fleming T.R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147–60.
- Weisberg S. (1985). *Applied linear regression*. Wiley, New York.