Petr Volf

An application of nonprarametric Cox regression model in reliability analysis: a case study

Kybernetika, Vol. 40 (2004), No. 5, [639]--648

Persistent URL: http://dml.cz/dmlcz/135622

KYBERNETIKA — VOLUME 40 (2004), NUMBER 5, PAGES 639 - 648

AN APPLICATION OF NONPARAMETRIC COX REGRESSION MODEL IN RELIABILITY ANALYSIS: A CASE STUDY[1]

PETR VOLF

The contribution deals with an application of the nonparametric version of Cox regression model to the analysis and modeling of the failure rate of technical devices. The objective is to recall the method of statistical analysis of such a model, to adapt it to the real-case study, and in such a way to demonstrate the flexibility of the Cox model. The goodness-of-fit of the model is tested, too, with the aid of the graphical test procedure based on generalized residuals.

Keywords: hazard rate, counting process, Cox model, nonparametric regression, local likelihood, time-to-failure

AMS Subject Classification: 62N05, 60G55, 62G08

1. INTRODUCTION

In the present paper a generalized, nonparametric version of the Cox regression model of hazard rate is used for the modeling of the failure times distribution. There exists a well developed methodology of nonparametric estimation in generalized regression (i. e. in the exponential family of models, including the Cox one), described e. g. in Hastie and Tibshirani [6], further for instance in Gentleman and Crowley [5], O'Sullivan [10], Fan and Gijbels [4]. We use here a variant of the local likelihood maximization proposed in Volf [13]. Alternative way to formulation of a general regression model via regression splines has been proposed by C. Stone in a series of papers on "dimensionality reduction principle" (e. g. Stone, [11]). Kooperberg et al [8] have then adapted this idea for the case of nonparametric Cox model and proved consistency of such a spline approximation.

The organization of the paper is following: In Part 2 the scheme of counting process and Cox model are recalled briefly. Then, the rest of the study is devoted to the analysis of a real data known as the Reynolds Metals Company data (Part 3), namely to estimation of lifetimes of damaged electrolytic cells. The case has formerly

[1]This work has partially been supported by Grants 201/02/0049 and 402/01/0539 of the Grant Agency of the Czech Republic.

been statistically examined by several analysts, cf. contributions of Kalbfleisch and Struthers and of Thomas, both published in C. J. S. [7]. Nevertheless, it is so interesting from the point of statistical data analysis that it can be used as a bench-mark example for comparison of analytic techniques. Today interest in these data is due to recent development of the methodology of nonparametric identification of generalized regression models. Thus, in the paper of Arjas and Liu [3] the Bayesian approach and the Gibb's sampling procedure are used for the estimation of nonparametrized hazard rate. The novelty of our solution presented here consists in that we consider a more detailed and more flexible (also nonparametric) model. Part 4 presents two different methods of evaluation of losses caused by the shutdown. Finally, in Part 5 we check the fit of the model by a graphical test based on the properties of generalized residuals originally proposed by Arjas [2].

2. COUNTING PROCESS AND COX REGRESSION MODEL

Let us recall briefly the notion of counting process and Cox regression model (see e. g. Andersen et al, [1]). A multivariate counting process $N(t) = N_1(t), \ldots, N_n(t)$ is a set of right-continuous random step functions with $N_i(0) = 0$ and with steps $+1$ at the moment of (observed) event, in our case the failure of the ith device. The probability (the hazard) of the failure is modeled via the hazard function. The Cox regression model assumes that the hazard function is $h(t, \mathbf{x}) = h_0(t) \exp(f(\mathbf{x}))$, where $\mathbf{x}$ is a (K-dimensional) covariate and $h_0(t)$ is the baseline hazard function. The most frequently used semiparametric version has $f(\mathbf{x}) = \beta'\mathbf{x}$. We shall consider a case when $f(\mathbf{x})$ is a nonparametrized additive function $f(\mathbf{x}) = \sum_{j=1}^{K} f_j(x_j)$, the goal is to identify suitable functions $f_1, \ldots, f_K$ and function $h_0(t)$, or its cumulative version $H_0(t) = \int_0^t h_0(s)\, \mathrm{d}s$. Evident ambiguity (with respect to additional constants in f_j-s) can be overcome by a proper normalization of these functions.

Simultaneously, the model admits time-dependent covariate processes $\mathbf{X}_i(t)$, $i = 1, \ldots, n$. Then, the behaviour of each component $N_i(t)$ is governed directly by the intensity process

$$\lambda_i(t) = h(t, \mathbf{X}_i(t)) \cdot I_i(t) = h_0(t) \cdot \exp\{f(\mathbf{X}_i(t))\} \cdot I_i(t),$$

$i = 1, \ldots, n$, $t \in [0, \mathcal{T}]$, where $I_i(t)$ is an indicator process, $I_i(t) = 1$ if the ith object is in the risk set at moment t, $I_i(t) = 0$ otherwise. The inference is based on the Cox partial likelihood (cf. again Andersen et al, [1]). Its logarithm is:

$$\ell_n = \sum_{i=1}^{n} \int_0^{\mathcal{T}} \ln \frac{\exp f(\mathbf{X}_i(t))}{\sum_{j=1}^{n} \exp f(\mathbf{X}_j(t)) I_j(t)} \mathrm{d}N_i(t).$$

Notice that $\mathrm{d}N_i(t) = 1$ at points of events, $\mathrm{d}N_i(t) = 0$ otherwise, so that the integral transforms to a sum. When an estimate of function f is available, we can use the following generalized maximum likelihood (Breslow–Crowley) estimator of the cumulative baseline hazard function

$$\hat{H}_0(t) = \int_0^t \frac{\mathrm{d}\bar{N}(s)}{\sum_j \exp \hat{f}(\mathbf{X}_j(s)) I_j(s)}, \tag{1}$$

where $\bar{N}(t) = \sum_{i=1}^{n} N_i(t)$. $\hat{H}_0(t)$ is then a nondecreasing stepwise function, with steps at points of observed events (i.e. at points of counts of $\bar{N}(t)$). From the increments of $\hat{H}_0(t)$ an estimation of function $h_0(t)$ can be obtained, with the aid of a smoothing procedure. In the following part the model will be applied to a real data case. We shall also recall briefly some previous analyses.

3. A CASE STUDY

In 1967, a strike at a Quebec aluminium smelter resulted in the uncontrolled shutdown of electrolytic cells. The company claimed that the shutdown caused the shorter operating lives of cells operating at the time. The case led to a legal action and initialized a need of a deep statistical analysis of the data, in order to confirm expected higher failure rate after the intervention (shutdown) and to estimate statistically the losses caused by this (eventual) higher rate. The more details about the case as well as the complete data were published in "Case Studies in Data Analysis", a section of the Canadian Journal of Statistics, V. 10 [7]. The data are now available also on the web page *siprint.utia.cas.cz/public/income/volf/data_survival.* They could be divided to three parts. In the first one, there are data on 395 cells of standard types, of which 297 experienced the shutdown. There are 20 types of standard cells (denoted A1 – A20). The second data part refers to 104 cells of experimental design, their types are labeled as B, $C, \ldots,$ K. The survival of cells was measured in days, the highest observed time to failure was 2 541 days. The installation times differed from 2 287 to only 3 days before the shutdown. The survival times are known, noncensored. From all these 499 cells, 349 were in circuit at the moment of intervention. The third group, 73 experimental cells of types labeled from L do O, were installed and had failed before the intervention. There arises the question whether this group (no experiencing the intervention) is worth to be taken into account, not bringing any information about the influence of the shutdown. However, as soon as the model with a common baseline hazard is considered, even these data contribute to the estimation of the baseline characteristic.

3.1. Choice of the model

In the first part of their study, Kalbfleisch and Struthers [7] estimated and compared the age-specific hazard rates before and after intervention. The fact of experience of intervention was treated as a $\{0, 1\}$ covariate in Cox model, two-sample test showed substantial increase of aggregated hazard rate after the intervention. Other covariates have also been considered, namely the age of the cell at the moment of intervention and the time from intervention (provided the cell experienced it at all). Thus, the changes of the hazard rate in the course of individual time have been examined attentively, while the types of cells have not been considered.

Thomas [12] used the standard Cox model, considering the following covariates: date of installation, experience of intervention (0 or 1), sub-type of cell (the types of cells were aggregated to 6 subclasses). The pairwise interactions of these three covariates were considered, too. Thomas remarked that the statistical tests revealed a lack of fit of the model. It could be caused by a nonoptimal choice of covariates

or by a nonoptimal structure of the chosen model. It seems to be more appropriate to consider two hazard rates, one for non-damaged cells, and the second as a function of the time after intervention. These two hazard rates can be connected in a multiplicative way, so that one is regarded as the baseline hazard, the second as a function describing the covariate effect of the time after intervention. Following this idea, we shall work with the following model of the hazard rate of failure of ith cell, using the age of the cell as a reference time t:

$$\lambda_i(t) = h_0(t)\cdot\exp\{b(t-U_i)\cdot 1[t>U_i] + c(x_i)\}\,,\ t\in[0,T_i],\ \lambda_i(t)=0 \text{ otherwise,} \quad (2)$$

where T_i is the survival time of ith cell. U_i is now the age of ith cell at the moment of intervention and x_i is the type of cell i – we suppose that each type may have its specific survival. The values from 1 to 34 are assigned to types A1,...,A20, B, C,...,N,O. While the covariate x is categorized, the time is a continuous variable, functions h_0, b are assumed to be continuous and bounded.

3.2. The procedure of estimation

The procedure follows the version of local maximum likelihood method proposed in Volf [13]. Denote $f(t,u,x) = b(t-u)1[t>u] + c(x)$. As each cell encountered exactly one failure, at moment T_i, Cox's partial likelihood is now

$$\ell = \sum_{i=1}^{n} \ln \frac{\exp(f(T_i,U_i,x_i))}{S(i)},$$

where $S(i) = \sum_{j=1}^{n} \exp(f(T_i,U_j,x_j))I_j(T_i)$. The computations start from b, $c\equiv 0$. Let us imagine that we wish to estimate the value of function b at a fixed point s (>0). Therefore we regard (for the moment) function $b(\cdot)$ as a constant b_s in a chosen window $\mathcal{O}(s)$ around s. From the equation $\partial\ell\,/\,\partial b_s = 0$ we obtain

$$\sum_{i=1}^{n}\left\{1[T_i>U_i]\cdot 1[(T_i-U_i)\in\mathcal{O}(s)] - \frac{R(s,i)}{S(i)}\cdot\exp b_s\right\} = 0, \quad (3)$$

where $R(s,i) = \sum_{j=1}^{n} 1[(T_i-U_j)\in O(s)]\cdot 1[T_i>U_j]\cdot\exp c(x_j)\cdot I_j(T_i)$. An iteration step is then:

$$b_s^{(r+1)} = -\ln\left\{\sum_i \frac{R(s,i)}{S(i)}\Big/\sum 1[T_i>U_i]\,1[(T_i-U_i)\in\mathcal{O}(s)]\right\},$$

where the right side uses the estimates of functions b, c obtained from the preceding steps. The task of local estimation of function c is solved in the same way. As the variable x is categorized to $M=34$ classes, the task is equivalent to the solution in the framework of Cox model with M parameters $c_1,\dots,c_M$ and 0,1-valued covariates. We have

$$\frac{\partial\ell}{\partial c_m} = \sum_{i=1}^{n}\left\{1[x_i=m] - \frac{S(m,i)}{S(i)}\exp c_m\right\}, \quad (4)$$

$$\text{where } S(m,i) = \sum_{j=1}^{n} 1[x_j=m]\cdot\exp\{b(T_i-U_j)\cdot 1[T_i>U_j]\}\cdot I_j(T_i),$$

and the iteration step is

$$c_m^{(r+1)} = -\ln\left\{\sum_i \frac{S(m,i)}{S(i)} \Big/ \sum_i 1[x_i = m]\right\}.$$

Results: The progress of iteration was controlled and its convergence observed from the changes of estimated parameters c_m. Originally, function $b(s)$ was estimated at equidistant points, at each 10 (days), with the use of mowing window (neighborhood) changing its width in order to contain a prescribed number of points (so called k-nearest neighbors variant of kernel smoothing). The full domain of s was from 0 to 1 837, so that we obtained 183 values. Then the estimate was secondary smoothed, i. e. the values were averaged (in a weighted way) by a triangular kernel. The graphs of estimated functions b and c are displayed in Figures 1a, 1b. After we decided to stop the iterations (when changes of values of c_m were less than 0.1 %), we computed the estimate of cumulative baseline hazard function $H_0(t)$ in accordance with (1). From it, by a kernel smoothing of its increments, we obtained a graph of estimate of $h_0(t)$ in Figure 1c.

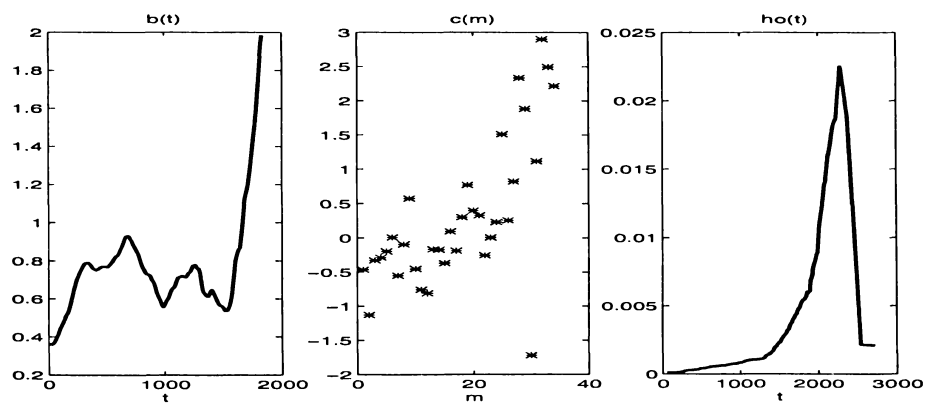

Fig. 1. Estimates of functions $b(s)$, $c(m)$ and baseline hazard rate $h_0(t)$.

4. EVALUATION OF THE COST OF INTERVENTION

We shall present here two different methods of such an evaluation.

Method 1. Let $T(m,U)$ be a random variable – the remaining time to failure of the cell type m, which has survived up to age U and which is supposed not to be involved in the intervention. The hazard rate of distribution of $T(m,U)$ is $h_0(s+U)\cdot \exp c_m$, $s \geq 0$, its estimate is available. The data contains n_m cells of type m which passed the intervention at ages U_i, their remaining survival times $S(m,i) = T_i - U_i$, $i = 1,\ldots,n_m$, have been observed. Then the total loss (in days)

of cells of type m is given by random variable $D_m = \sum_{i=1}^{n_m}(T(m,U_i) - S(m,i))$. The mean number of lost days is directly $E\,D_m = \sum_i(E\,T(m,U_i) - S(m,i))$.

The mean remaining lifetime is $E\,T(m,U) = -\int_0^\infty s\,\mathrm{d}P_{U,m}(s)$, where

$$\mathrm{d}P_{U,m}(s) = \mathrm{d}P_m(s+U)/\,P_m(U) \quad \text{and} \quad P_m(t) = P_0(t)^{\exp c(m)},$$

$P_m(t)$ is the survival function (for nondamaged cells of type m), $P_0(t) = \exp(-H_0(t))$ is the baseline survival function. The estimate of the mean is given by the sum

$$\hat{E}\,T(m,U) = -\sum_{j=1}^{n}(T_{(j)} - U)\,1[T_{(j)} > U]\cdot\Delta\,\hat{P}_m(T_{(j)})\,\Big/\,\hat{P}_m(U),$$

where $T_{(j)}$ are ordered all survival times, $T_{(1)} \le T_{(2)} \le \ldots \le T_{(n)}$,

$$\Delta\,\hat{P}_m(T_{(j)}) = \hat{P}_m(T_{(j)}) - \hat{P}_m(T_{(j-1)}), \quad \hat{P}_m(t) = \exp\left(-\hat{H}_0(t)\cdot\exp\hat{c}(m)\right)$$

and $\hat{P}_m(T_{(0)}) = 1$. The method 1 has certain disadvantages: It yields for each cell just the estimate of the mean value of its "would-be" remaining survival. It is true that, once the (estimate of) model is available, we can generate randomly the data corresponding to this model. They provide a sample representation of the underlying distribution. Hence, we can also for each cell obtain an estimate of probability distribution of remaining survival and of corresponding loss, Nevertheless, we actually take each cell of the same type as possessing the same distribution of remaining lifetime (depending on its age at moment of intervention), though our data contain also an information on different "frailty" of individual cells (i. e. their proneness to earlier or later failure). This information is utilized in the second method of evaluation the difference between actual and hypothetical remaining survival time:

Method 2. Let us first recall the connection of general survival times and standard exponential distribution, Let T_i be a set of independent identically distributed random variables – waiting times – possessing a common cumulative hazard rate $L(t)$, then $\sum_{i=1}^{k} L(T_i)$ is the waiting time to the kth event of standard Poisson process, or, equivalently, it is a random variable distributed according to the gamma $(1,k)$ law (each $L(T_i)$ has standard exponential distribution and they are mutually independent). This connection between the actual process and the standard Poisson process is used for the testing the fit of the model, because the accuracy of this transformation depends strongly on the accuracy of the model of hazard rate (cf. Arjas [2]).

It has also another important consequence. Assume that a cell of type m had experienced the intervention at age U and then it survived another $S = T - U$ days. The cumulative hazard of this remaining lifetime was $L_{1,m}(S) = \int_U^T h_0(s)\exp(b(s - U) + c(m))\,\mathrm{d}s$. Provided the model is correct, $L_{1,m}(S)$ is a standard exponential random variable. The realization of S shows how quickly the 'hazard clock' of the cell has been running after U.

Table 1. Estimates of number of lost days.

	Method 1:		Method 2:	
Cell Type	Remain.Lifetimes	Differences	Remain.Lifetimes	Differences
1	703.021	323.021	502.1910	122.1910
2	6186.977	1176.977	7214.6456	2204.6456
3	8018.959	2720.959	8403.7884	3105.7884
4	8691.149	2169.149	9625.9951	3103.9951
5	9457.718	3022.718	10077.0737	3642.0737
6	11735.956	3895.956	11530.0891	3690.0891
7	20500.358	4856.358	20920.3909	5276.3909
8	12074.879	4390.879	11402.1867	3718.1867
9	3501.335	876.335	3795.4695	1170.4695
10	3244.579	1368.579	2862.9982	986.9982
11	10865.187	2232.187	11583.7112	2950.7112
12	13970.679	2770.679	14346.7773	3146.7773
13	12199.764	3366.764	11973.1333	3140.1333
14	7221.806	1820.806	7461.6275	2060.6275
15	12802.812	3357.812	12653.0895	3208.0895
16	15184.571	4319.571	14294.2969	3429.2969
17	29215.036	7062.036	29166.3350	7013.3350
18	16568.215	3726.215	16690.5910	3948.5910
19	3193.496	528.496	3621.4483	966.4483
20	15309.948	3232.948	15793.7712	3716.7712
21	198.770	−835.230	1034.0000	0.0000
22	3649.869	2142.869	2233.7557	726.7557
23	1507.566	391.566	1613.0126	497.0126
24	250.948	250.948	0.0000	0.0000
25	890.102	264.102	909.0553	283.0553
26	5237.939	2283.939	4338.3263	1384.3263
27	2712.441	690.441	2914.5540	892.5540
28	1869.895	413.895	2100.9605	644.9605
29	1534.350	176.350	1949.4194	591.4194
30	3442.019	1273.019	2294.1161	125.1161
T O T A L:	241840.344	64270.344	243316.8094	65746.8094

If the intervention did not occurred, the cumulative hazard on the interval (U, t) would be $L_{2,m}(t-U) = \int_U^t h_0(s) \exp c(m)\, ds = (H_0(t) - H_0(U)) \cdot \exp c(m)$. So that $R = L_{2,m}^{-1}(L_{1,m}(S))$ is now the remaining survival time after U of the same cell but in "another world" in which the cell did not pass the intervention. Again, provided the model is correct. Natural estimates are:

$$\hat{L}_{1,m}(S) = \sum_{i=1}^{n} 1[U < T_i \le U + S] \exp(\hat{b}(T_i - U) + \hat{c}(m)) \Delta \hat{H}_0(T_i),$$

$$\hat{L}_{2,m}(s) = \exp \hat{c}(m) \cdot \sum 1[U < T_i \le U + s]\, \Delta \hat{H}_0(T_i) \quad \text{and}$$

$$\hat{L}_{2,m}^{-1}(z) = \inf \left\{ s : \hat{L}_{2,m}(s) \ge z \right\},$$

eventually the inverse function can be computed with the help of interpolation. So that the estimate of $R-S$ yields another estimate of losses caused by the intervention. Notice that the evaluation of R from equation $L_{2m}(R) = L_{1m}(S)$ does not require the knowledge (estimate) of function c (while the former method of evaluation of expected remaining lifetimes did not use the function b).

The estimation of numbers of lost days is summarized in Table 1. Both methods yield similar results, though we have seen that they differ in their substance (subjectively, we tend to prefer the approach 2). The computations used the estimates of cumulative baseline hazard function H_0, of function c and secondary smoothed estimate of function b. The first two columns contain the results of method 1, namely the expected remaining lifetimes summarized for type m, $\sum_i \mathsf{E}\,T(m;U)$, and the differences from really achieved remaining survival times. The last two columns display the values of estimated remaining lifetimes, R_i, again summarized for types of cells, and summarized differences $R_i - S_i$. Both R_i and S_i are computed following the second method. Our results differ from the estimate (in much rougher, semiparametric model) of Kalbfleisch and Struthers [7]). Their estimate of the mean number of lost days was 82 653.5.

5. TESTING THE GOODNESS-OF-FIT

In all variants of tests based on generalized residuals the sample of examined objects is divided into two or more strata and in each stratum S the counting process of ordered observed failure times is examined. It concerns also to graphical method (Arjas [2]) used here, as well as to numerical methods proposed later in Marzec and Marzec [9] or Volf ([14], for the case of additive Aalen model).

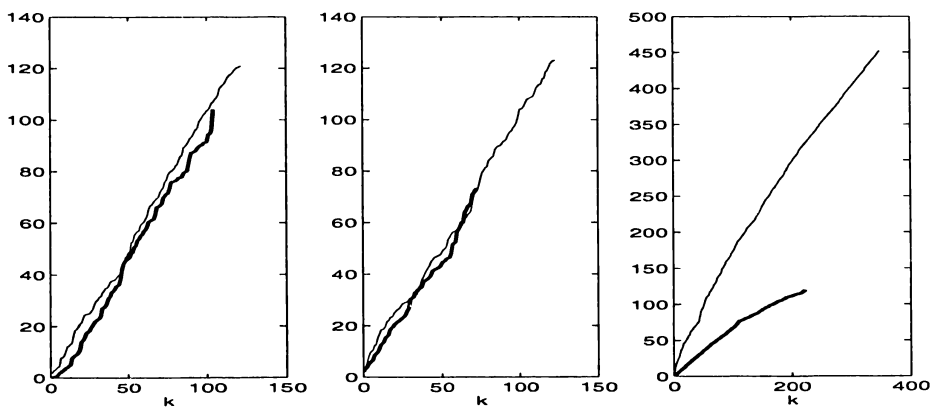

Fig. 2. Graphical goodness-of-fit tests. Plots of $\hat{T}_{kS}$ for: a) thick – cell types 21 – 30, thin – types 1 – 4; b) thin – types 11 – 20, thick – types 31 – 34; c) thick – cells without intervention, thin – cells which passed intervention.

Let failure times be $0 \leq T_{1,S} \leq T_{2,S} \leq \ldots$. Their transformation to the times

$\mathcal{T}_{k,S}$ of events of a standard Poisson process is given (provided the model holds) by

$$\mathcal{T}_{k,S} = \sum_{i \in S} \int_0^{T_{k,S}} \lambda_i(t)\, \mathrm{d}t.$$

Their estimates are obtained from the estimates of components of the model (2), i. e.

$$\hat{\mathcal{T}}_{k,S} = \sum_{i \in S} \sum_j 1[T_j \leq \min(T_{k,S}, T_i)] \cdot \exp\left\{\hat{b}(T_j - U_i)\, 1[T_j > U_i] + \hat{c}(x_i)\right\} \cdot \Delta \hat{H}_0(T_j).$$

These $\hat{\mathcal{T}}_{k,S}$ are plotted graphically, the "ideal" value of $\mathcal{T}_{k,S}$ should be k. If the plot of $\hat{\mathcal{T}}_{k,S}$ differs from k significantly, it is an indication that the model does not correspond to the data. As the test uses estimated response functions, and, in general, there does not exist a relevant theory of large sample properties of local likelihood estimates, it is better to use the graphical version of the test. In order to check the fit of (estimated) model (2), we stratified the data along to the types of cells and we performed the same kind of test for different selections of such subsamples. Even the worst results of tests did not contradict to the model. For instance, Figure 2a shows the plots for cell types 1–4 (121 cells, high survival, mostly without intervention) and cell types 21–30 (104 cells, lower survival, mostly with intervention). Then, in Figure 2b there are the plots for cells of types 11–20 (121 cells with rather high survival, in spite of the intervention) and cells 31–34 (73 experimental cells with low survival, without intervention). The graphs suggest that the model fits well for all types of cells.

On the contrary, Figure 2c demonstrates that the stratification cannot be arbitrary, that it should be independent of measured survival time. Figure displays the plots for cells which have passed the intervention and for cells which have not. The picture shows that for our data the actual hazard of damaged cells was lower than it was assumed by the model (and was higher for nondamaged cells). The reason was rather simple and natural. There was a high positive correlation between the survival of a cell and the event that this cell passed the intervention. In other words, the cells which had higher survival (caused only by their individual frailty, in the framework of the probabilistic model regarded as random fluctuations) were more likely to survive until the moment of intervention.

6. CONCLUSION

The nonparametric estimation has its clear advantages (consisting mainly in its universality), but its weak sides are well known, too. One of them consists in a rather slow global consistency of nonparametric estimates, while an estimate of parametrized model is able to reflect the main features of model very quickly. The nonparametric estimate depends more strongly on the data and their local nonregularities. Nevertheless, the experience with the method presented here is encouraging, it has been tested successfully by a number of simulated as well as of real-data examples.

In the framework of Cox regression model, the procedure of estimation yields also the asymptotic confidence regions for regression parameters and for cumulative baseline hazard function. It follows from the asymptotic normality of estimates (cf. Andersen et al, [1]). In our case, the application of these asymptotic results is rather limited, because we use a nonparametrized function b(s). This could be overcome with the help of a bootstrapped construction. However, even if we have confidence bounds for characteristics of the model, it is not clear how to derive confidence intervals for the mean number of lost days. That is why the confidence intervals were not constructed.

(Received October 1, 2003.)

REFERENCES

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding: Statistical Models Based on Counting Processes. Springer, New York 1993.
[2] E. Arjas: A graphical method for assesing goodnes of fit in Cox's proportional hazard model. J. Amer. Statist. Assoc. *83* (1988), 204–212.
[3] E. Arjas and L. Liu: Assessing the losses caused by an industrial intervention – a hiearchical Bayesian approach. J. Roy. Statist. Soc. Ser. C *44* (1995), 357–368.
[4] J. Fan and I. Gijbels: Local Polynomial Modelling and Its Applications. Chapman and Hall, London 1996.
[5] R. Gentleman and J. Crowley: Local full likelihood estimation for the proportional hazard model. Biometrics *47* (1991), 1283–1296.
[6] T. Hastie and R. Tibshirani: Generalized Additive Models. Chapman and Hall, London 1990.
[7] J. D. Kalbfleisch and C. A. Struthers: An analysis of the Reynolds Metals Company data. In: Case Studies in Data Analysis, Canad. J. Statist. *10* (1982), 237–259.
[8] C. Kooperberg, C. J. Stone, and Y. K. Truong: The L_2 rate of convergence for hazard regression. Scand. J. Statist. *22* (1995), 143–157.
[9] L. Marzec and P. Marzec: Generalized martingale-residual processes for goodness-of-fit inference in Cox's type regression model. Ann. Statist. *25* (1997), 683–714.
[10] F. O'Sullivan: Nonparametric estimation in the Cox model. Ann. Statist. *21* (1993), 124–145.
[11] C. J. Stone: The use of polynomial splines and their tensor products in multivariate function estimation; with discussion. Ann. Statist. *22* (1994), 118–184.
[12] D. C. Thomas: Case analysis using Cox's model. In: Case Studies in Data Analysis. Canad. J. Statist. *10* (1982), 237–259.
[13] P. Volf: A large sample study of nonparametric proportion hazard regression model. Kybernetika *29* (1993), 404–415.
[14] P. Volf: Analysis of generalized residuals in hazard regression models. Kybernetika *32* (1996), 501–510.

Petr Volf, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.
e-mail: volf@utia.cas.cz