Записки научных
семинаров ПОМИ
Том 339, 2006 г.

M. Nikulin, Hong-Dar Isaac Wu

FLEXIBLE REGRESSION MODELS FOR CARCINOGENESIS STUDIES

Abstract. Dynamic or flexible regression models are used more and more often in carcinogenesis studies to relate lifetime distribution to the time-depending explanatory variables. In addition to the classical regression models such as Cox model, AFT model, Linear transformation model, Frailty model, etc. In the paper exposed the so-called flexible regression models which well adapted to study the cross-effects of survival functions, which are sometimes observed in clinical trials. A classical example is the well-known data concerning effects of chemotherapy (CH) and chemotherapy plus radiotherapy (CH+R) on the survival times of gastric cancers patients, see Stablein and Koutrouvelis (1985), Kleinbaum (1996), Klei and Moeschberger (1997), Wu, Hsieh and Chen (2002), Bagdonavicius, Hafdi and Nikulin (2004), etc. In this paper we give examples to illustrate possible applications of the Hsieh model (2001) and the SCE model , proposed by Bagdonavicius and Nikulin (2005), adapted to treat survival data with one crossing point. We do the comparison of both models.

1. Introduction

Survival regression models relate lifetime distribution to the explanatory variables, possibly time-dependent. These models are used for estimation of the effect of *covariates* or *stresses* on survival and for estimation of the survival function under given covariate values.

The most popular and successful model is the *proportional hazards* (PH) model (called also the Cox model), introduced by Sir David Cox (1972). The popularity of this model is based on the fact that there exist simple semi-parametric estimation procedures which can be used when the form of survival distribution function is not specified, see Cox (1975). It is evident that the Cox model is not applicable in all the situations that one can meet in the practice, in particular when the 'proportional hazards' assumption is apparently violated. For example, it is not possible to explain correctly the influence of considered covariates (stresses) on duration of the life through the Cox model in some cases offered in Klein and Moeschberger (1997), Hsieh (2001), Wu, Hsieh, Chen (2002), Bagdonavicius and Nikulin (2002). More sophisticated models are needed in such situations.

In carcinogenesis studies, a specified survival model relates the hazard rate (or survival function) to a set of explanatory variables in some particular way. The explanatory variables may be modelled by stochastic processes, deterministic time functions or constants (possibly different for different individuals). Preliminarily, let $x(\cdot)$ be a vector of covariates, $x(\cdot) = (x_1(\cdot), ..., x_m(\cdot))^T : [0, \infty) \to \mathbf{R}^m$, which is a multidimensional and deterministic time function. We denote by E the set of all possible covariates. If $x(\cdot)$ is constant in time, $x(t) \equiv x$, then we shall write x instead of $x(\cdot)$, and E_1 is the set of all constant covariates ($E_1 \subset E$).

The distribution of survival under covariates will be defined in terms of hazard-rate function. Denoted by T the failure time, then the *survival* and the *hazard rate functions* given $x(\cdot)$ are

$$S_{x(\cdot)}(t) = \mathbf{P}\{T \geq t \mid x(u), 0 \leq u \leq t\}, \quad x(\cdot) \in E,$$

$$\lambda_{x(\cdot)}(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbf{P}\{T \in [t, t+h) \mid T \geq t, x(u), 0 \leq u \leq t\} = -\frac{S'_{x(\cdot)}(t)}{S_{x(\cdot)}(t)}.$$

The *cumulative hazard* under $x(\cdot)$ is defined as

$$\Lambda_{x(\cdot)}(t) = \int_0^t \lambda_{x(\cdot)}(u) du = -\ln\{S_{x(\cdot)}(t)\}, \quad x(\cdot) \in E.$$

In addition to the Cox's PH model, *flexible* or *dynamic* regression models are well adapted for many interesting and practical situations. People also use the term *accelerated life models* for some of these situations. Following Hsieh (2001), Wu, Hsieh, Chen (2002), and Bagdonavicius and Nikulin (2002), we consider herein one particular case in carcinogenesis studies that we observe 'cross-effects' of the Kaplan–Meier estimates according to different covariates.

Before our elucidation on the Cox PH model in Section 2, it is worth noting that conventional survival models (including the PH model) do not design so-called "dynamic" modeling automatically in the model formulation. To some extent, *dynamics* means that some covariate(s) has different "effect" at different times. This may induce a consideration of varying-coefficient Cox model (Martinussen and Scheike, 2006). On the other hand, "dynamics" may originate from an intrinsic mechanism; which is modelled through an adequate description on the "data history".

In epidemiologic studies, however, it is not always possible to capture all covariates that have dynamic influence; and, it is not possible that one can foresee which covariate(s) possess a dynamic effect, or a *cumulative* (or *summing*) effect. In such cases, modelling the hazard rates (at time t) through a set of covariates as well as through the *cumulative hazards just befor time* t $(t-)$ is a natural consideration. In the sequel, we also discuss several models that have a "dynamic" meaning.

2. The Cox or the proportional hazards model

The *proportional hazards* (PH) or *Cox model* holds on E, if the hazard rate has the form

$$\lambda_{x(\cdot)}(t) = r\{x(t)\}\,\lambda_0(t), \quad x(\cdot) \in E, \tag{1}$$

where $\lambda_0(\cdot)$ is a *baseline hazard rate* function, and $r(\cdot)$ is a positive function on E. The role of function r is very important: it explains the *summing effect* of all $x_i(\cdot)$ on distribution of T. If r and λ_0 are unknown, it is the *non-parametric model.* In most applications the function r is parametrized in the form

$$r(x) = \exp\{\beta^T x\},$$

where $\beta = (\beta_1, \cdots, \beta_m)^T$ is the vector of *regression parameters.* Under this parametrization we obtain the classical *semi-parametric* Cox model on E with time-dependent covariables:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}\lambda_0(t), \quad x(\cdot) \in E. \tag{2}$$

If r is parametrized and λ_0 is taken from some parametric family of hazard rates, for example, from Weibull family, then we have a parametric model.

The PH model implies that the *hazard ratio* $HR(t, x_1, x_2)$ of hazard rates under different fixed constant covariates x_1 and x_2 is constant over time:

$$HR(t, x_1, x_2) = \frac{\lambda_{x_2}(t)}{\lambda_{x_1}(t)} = \frac{r\{x_2\}}{r\{x_1\}} = \text{const}, \quad x_1, x_2 \in E_1.$$

Usually the Cox model is considered as semiparametric: the finite-dimensional parameter β and the baseline hazard function λ_0 are supposed to be completely unknown.

The PH model is mostly applied for analysis of survival data but the statisticians working in reliability are very cautious to use it, especially for analysis of accelerated life testing data. It can be explained by the fact that the PH model has one very unnatural property: the conditional probability to fail in a time interval $(t, t+s)$ given that a unit is functioning at the moment t depends only on the values of the stress (or covariable) $x(\cdot)$ in that interval but does not depend on the values of the stress until the moment t:

$$\mathbf{P}(T \leq t+s \mid T > t) = 1 - e^{-\int_t^{t+s} e^{\beta^T x(u)} \lambda_0(u) du},$$

here T is failure time, λ_0 is the baseline hazard function which does not depend on stress. For this reason we can say that PH model has the *absence of memory property.*

The common sense says that if a unit functioned in high stress conditions, it used a large amount of it's resource and is "older" than a unit which functioned in mild stress conditions the same time t, so the conditional probabilities of failure after the moment t should be different for those units.

Nevertheless, in survival analysis the PH model usually works quite well, because the values of covariates under which estimation of survival is needed are in the range of covariate values used in experiments. So the use of a not very exact but simple model often is preferable to the use of a more adequate but complicated model. It is similar with application of the linear regression models in the classical regression analysis: the mean of the dependent variable is rarely a linear function of the independent variables but the linear approximation works reasonably well in some range of independent variable values.

At the end we note that based on the partial likelihood estimation (Cox, 1972, 1975), the PH analysis has the merits of easy implementation, semiparametric efficient, nice interpretation in hazard ratio, and readily available analysis packages in various statistical software (such as SAS, S-plus, and SPSS, etc.)

3. AFT Model

Let us note that under covariate $x(\cdot)$ the probability $S_{x(\cdot)}(t)$ characterizes for any fixed t the summing effect of covariate values in the interval $[0, t]$ on survival. The equality $\Lambda_{x(\cdot)}(t) = -\ln S_{x(\cdot)}(t)$ implies that the cumulative hazard also characterizes this summing effect. So it can be supposed that the hazard rate at any moment t is a function of the covariate value $x(t)$ and the value of the cumulative hazard $\Lambda_{x(\cdot)}(t)$.

The *generalized Sedyakin's* (GS) model on E namely supposes it (Sedyakin (1966)):

$$\lambda_{x(\cdot)}(t) = g\left(x(t), \Lambda_{x(\cdot)}(t)\right), \quad x(\cdot) \in E. \tag{3}$$

This model with g completely unknown is too general to do statistical inference. But if we choose some regression model for constant covariates, the form of the function g can be made more concrete.

Suppose that under different constant covariates $x \in E_1$ the survival functions differ only in scale:

$$S_x(t) = S_0\left(r(x)t\right), \quad x \in E_1, \tag{4}$$

for some positive r on E_1. If the GS model holds on a set $E, E \supset E_1$, of covariates, then (4) holds on E_1 if and only if the function g has the form $g(x,s) = r(x)q(s)$ for some positive q.

We obtain the following model:

$$\lambda_{x(\cdot)}(t) = r\{x(t)\}\, q\{\Lambda_{x(\cdot)}(t)\}, \quad x(\cdot) \in E. \tag{5}$$

Using the relation between the survival and the cumulative hazard functions, it is easy to obtain the so-called AFT (*accelerated failure time*) model on E according to which the survival function has the form

$$S_{x(\cdot)}(t) = S_0\left(\int_0^t r(x(u))du\right), \quad x(\cdot) \in E, \tag{6}$$

where the function S_0 does not depend on $x(\cdot)$. The function r changes locally with the time-scale, and is often parametrized in the following form:

$$r(x) = e^{-\beta^T x},$$

where $\beta = (\beta_1, \cdots, \beta_m)^T$ is a vector of unknown regression parameters.

Under the parametrized AFT model the survival function is

$$S_{x(\cdot)}(t) = S_0\left(\int_0^t e^{-\beta^T x(u)}du\right), \quad x(\cdot) \in E, \tag{7}$$

and the hazard rate is

$$\lambda_{x(\cdot)}(t) = e^{-\beta^T x(t)}\, \lambda_0\left(\int_0^t e^{-\beta^T x(u)}du\right), \quad x(\cdot) \in E, \tag{8}$$

and for constant covariates

$$S_x(t) = S_0\left(e^{-\beta^T x}\, t\right), \quad x \in E_1.$$

So the AFT model on E_1 can also be written as a *loglinear* model or *linear trasformation* (LT) model, since the logarithm of the failure time T_x under constant covariate x can be written as

$$\ln\{T_x\} = \beta^T x + \varepsilon, \quad x \in E_1, \tag{9}$$

where the survival function of the random variable ε does not depend on x and is $S(t) = S_0(\ln t)$.

In the case of lognormal failure-time distribution the distribution of ε is normal and we have the standard linear regression model. The equality (8) implies that if the survival function under any constant covariate belongs to parametric families such as Weibull, log-logistic, lognormal, then the survival function under any other constant covariate also belongs to that family.

Different from the PH model, the AFT model is mostly applied in survival analysis as a parametric model: the function S_0 (or the distribution of ε) is taken from some parametric class of distributions and the parameters to estimate are the parameters of this class and the regression parameters β.

In the case of semiparametric estimation the function S_0 is supposed to be completely unknown and the regression parameters as the function S_0 are the parameters to estimate in the model (7). The semiparametric AFT model is much less used in survival analysis than the Cox's PH model because of complicated estimation procedures: modified variants of likelihood functions are not differentiable and even not continuous functions, the limit covariance matrices of the regression parameters depend on the derivatives of the probability density functions.

The parametric AFT model is used in failure time regression analysis and accererated life testing. Under special experiment plans even nonparametric estimation procedures are used. In such a case not only the function S_0 but also the function r in the model (6) are assumed to be completely unknown.

The AFT model is a good choice when the lifetime distribution class is supposed to be known. Nevertherless, it is as restrictive as the PH model. The assumption that the survival distributions under different covariate values differ only in scale is a rather strong assumption. So more sophisticated models are also needed. More about the properties of the AFT model one can find in Meeker and Escobar (1998), Hsieh (2001), Bagdonavičius and Nikulin (2002), Dabrowska (2005, 2006). Concerning different biomedical applications of the AFT model, one can see in the recent book of Martinussen and Scheike (2006).

4. GPH, LT and Frailty models

The AFT and PH models are rather restrictive. Under the PH model, lifetime distributions with constant covariates are from the narrow class of distributions: the *ratio of hazard rates* (or simply *hazard ratio*) with respect to any two different constant covariates is constant over time.

Under the AFT model the covariate changes (locally, if the covariate is not constant) only the scale.

Generalized proportional hazards (GPH) models on E, proposed by Bagdonavičius and Nikulin (1999), allow the ratios of the hazard rates under constant covariables to be not only constant but also increase or decrease. They include AFT and PH models as particular cases.

As was discussed in the previous section, the survival function $S_{x(\cdot)}(t)$ (or, equivalently, the cumulative hazard function $\Lambda_{x(\cdot)}(t)$) characterizes the summing effect of covariate values in the interval $[0, t]$ on survival. So suppose that the hazard rate at any moment t is proportional not only to a function of the covariate applied at this moment and to a baseline rate, but also to a function of the probability of survival until t (or, equivalently, to the cumulative hazard at t):

$$\lambda_{x(\cdot)}(t) = r\{x(t)\}\, q\{\Lambda_{x(\cdot)}(t)\}\, \lambda_0(t), \quad x(\cdot) \in E. \tag{10}$$

We call the model (10) the generalized proportional hazards (GPH) model, see Bagdonavičius and Nikulin (1999), Dabrowska (2005, 2006), Martinussen and Scheike (2006). Particular cases of the GPH model are the PH model ($q(u) \equiv 1$) and the AFT model ($\lambda_0(t) \equiv \lambda_0 = \text{const}$).

Under the GPH model on E the survival functions $S_{x(\cdot)}$ have the form

$$S_{x(\cdot)}(t) = G\left\{\int_0^t r(x(\tau))d\Lambda_0(t)\right\}, \quad x(\cdot) \in E, \tag{11}$$

where

$$\Lambda_0(t) = \int_0^t \lambda_0(u)du, \quad G = H^{-1}, \quad H(u) = \int_0^{-\ln u} \frac{dv}{q(v)}.$$

We denote by H^{-1} the function inverse to G.

Models of different levels of generality can be obtained by completely specifying q, parametrizing q, or considering q as unknown.

Completely specifying q gives rather strict models which are alternatives to the PH model and the field of their application is relatively narrow (see Bagdonavičius and Nikulin (1994)). Under constant covariates such models are the *linear transformation* (LT) models, proposed by Dabrowska and Doksum (1988). Indeed, if q is specified and r is parametrized by $r(x) = e^{\beta^T x}$ then under constant covariables the survival functions have the form

$$S_{x(\cdot)}(t) = G\left\{e^{\beta^T x}\Lambda_0(t)\right\}, \quad x \in E_1,$$

with G specified. This implies that the random variable T_x can be transformed by the function $h(t) = \ln\{H(S_0(t))\}$ to the random variable of the form

$$h(T_x) = -\beta^T x + \varepsilon, \quad x \in E_1 \tag{12}$$

where ε is a random error with the parameter-free distribution function

$$Q(u) = 1 - G(e^u).$$

It is the LT model of Dabrowska and Doksum (1988), Dabrowska (2005).

Examples of the LT models include:

1) PH model (G is the Weibull survival function, ε has the extreme value distribution).

2) Proportional odds model (G is the log-logistic survival function, ε has the logistic distribution):

$$\frac{1}{S_x(t)} - 1 = r(x)\left(\frac{1}{S_0(t)} - 1\right), \quad x \in E_1,$$

see Bennett (1983), Murphy and van der Vaart (1997), Yang and Prentice (1999). More about semiparametric estimation for LT models one can see in Dabrowska (2005,2006), Martinussen and Scheke (2006), Chen, Jin and Ying (2002), Bagdonavicius and Nikulin (2002), etc.

3) generalized probit model (G is a lognormal survival function, has the normal distribution):

$$\Phi^{-1}\left(S_x(t)\right) = \log\left(r(x)\right) + \Phi^{-1}\left(S_0(t)\right), \quad x \in E_1,$$

where Φ is the standard normal cumulative distribution function.

The last two models are alternatives to the PH model. They are widely used for analysis of dichotomous data when the probability of “success” in dependence of some factors is analyzed. If application of the PH model is dubious then better is to use a (not very) wider GPH model which is obtained from the general GPH model not by complete specification of the function q but taking a simple parametric model for it.

Let us consider relations between the GPH models and the *frailty models* (Hougaard (1986)) with covariates.

The hazard rate can be influenced not only by the observable covariate $x(\cdot)$ but also by a non-observable positive random covariate Z, called the *frailty variable*. Suppose that the hazard rate given the frailty variable value is

$$\lambda_{x(\cdot)}(t|Z=z) = z\, r(x(t))\, \lambda_0(t), \quad x(\cdot) \in E.$$

Then

$$S_{x(\cdot)}(t) = \mathbf{E}\, exp\{-Z \int_0^t r(x(\tau))\, d\Lambda_0(\tau)\} = G\{\int_0^t r(x(\tau)) d\Lambda_0(\tau)\},$$

where $G(s) = \mathbf{E}e^{-sZ}$.

So the GPH model can be defined by specification of the frailty variable distribution.

a. The GPH models with monotone hazard ratios

The following parametrizations of r and q give submodels of the GPH model with monotone ratios of hazard rates under constant covariates. Using only one parameter and power or exponential functions for function q parametrization several important models were obtained by Bagdonazvicius and Nikulin (1999).

Suppose that $q(0) = 1$ (if it is not so, we can include $q(0)$ in λ_0, which is considered as unknown). Taking a power function $q(u) = (1+u)^{-\gamma+1}$ and $r(x) = e^{\beta^T x}$ we obtain the first GPH (GPH1) model on E:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}(1 + \Lambda_{x(\cdot)}(t))^{-\gamma+1} \lambda_0(t), \quad x(\cdot) \in E. \tag{13}$$

It coincides with the PH model when $\gamma = 1$. The supports of the survival functions $S_{x(\cdot)}$ are $[0, \infty)$ when $\gamma \geq 0$ and $[0, sp_{x(\cdot)})$ with finite right ends $sp_{x(\cdot)}$, $sp_{x(\cdot)} < \infty$, when $\gamma < 0$. Finite supports are very possible in accelerated life testing: failures of units at different accelerated stresses are concentrated in intervals with different finite right limits.

Suppose that at the point $t = 0$ the hazard ratio $HR(t, x_1, x_2)$ of the hazard rates under constant covariates x_1 and x_2 is greater then 1:

$$HR(0, x_1, x_2) = \frac{r(x_2)}{r(x_1)} = c_0 > 1, \quad x_1, x_2 \in E_1.$$

The hazard ratio $HR(t, x_1, x_2)$ has the following properties:

a) if $\gamma > 1$, then the ratio of the hazard rates decreases from the value $c_0 > 1$ to the value $c_\infty = c_0^{\frac{1}{\gamma}} \in (1, c_0)$.

b) if $\gamma = 1$ (PH model), the ratio of the hazard rates is constant.

c) if $0 \leq \gamma < 1$, then the ratio of the hazard rates increases from the value $c_0 > 1$ to the value $c_\infty = c_0^{\frac{1}{\gamma}} \in (c_0, \infty)$.

d) if $\gamma < 0$, then the ratio of the hazard rates increases from the value $c_0 > 1$ to ∞, end the infinity is attained at the point $sp_{x_2} = \Lambda_0^{-1}\{-1/((\gamma)r(x_2))\}$. The hazard rates go away one from another quickly when t increases.

The GPH1 model is a generalization of the *positive stable frailty model with explanatory variables*: the GPH model with $\gamma = 1/\alpha > 0$ is obtained taking the frailty variable Z which follows the positive stable distribution with the density

$$p_Z(z) = -\frac{1}{\pi z}\exp\{-\alpha z + 1\}\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\sin(\pi\alpha k)\frac{\Gamma(\alpha k + 1)}{z^{\alpha k}}, \quad z > 0,$$

where α is a stable index, $0 < \alpha < 1$. More about one can see in Bagdonavicius and Nikulin (2002), Greenwood and Nikulin (1996).

a.1. The second GPH model

Under the GPH1 model the support of the survival functions is infinite when $\gamma \geq 0$ and finite when $\gamma < 0$. The limit is $\gamma = 1$. So it is interesting to take a model with the following parametrization: $q(u) = (1 + \gamma u)^{-1}$. We obtain the second GPH (GPH2) model on E:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}(1 + \gamma\Lambda_{x(\cdot)}(t))^{-1}\lambda_0(t), \quad \gamma \geq 0, \quad x(\cdot) \in E. \qquad (14)$$

It also coincides with the PH model when $\gamma = 0$. The supports of the survival functions $S_{x(\cdot)}$ are $[0, \infty)$.

The hazard ratio

$$R(t, x_1, x_2) = \lambda_{x_2}(t)/\lambda_{x_1}(t), \quad x_1, x_2 \in E_1,$$

has the following properties:

a) if $\gamma > 0$, then the ratio of the hazard rates decreases from $c_0 > 1$ to the value $\sqrt{c_0} \in (1, c_0)$, i.e. the hazard rates approach one another when t increases.

b) if $\gamma = 0$ (PH model), the ratio of the hazard rates is constant.

The GPH2 model equivalent to the *inverse gaussian frailty model with explanatory variables*: the GPH model with $\gamma = (4\sigma\theta)^{1/2} > 0$ is obtained taking the frailty variable Z which follows the inverse gaussian distribution with the density

$$p_Z(z) = \left(\frac{\sigma}{\pi}\right)^{1/2} e^{\sqrt{4\sigma\theta}} z^{-3/2} e^{-\theta z - \frac{\sigma}{z}}, \quad z > 0.$$

More about one can see, for example, in Voinov and Nikulin (1993).

a.2. The third GPH model

Taking the exponential function $q(u) = e^{-\gamma u}$ and $r(x) = e^{\beta^T x}$ we obtain the third GPH (GPH3) model:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t) - \gamma \Lambda_{x(\cdot)}(t)} \lambda_0(t), \quad x(\cdot) \in E. \tag{15}$$

It coincides with the PH model when $\gamma = 0$. The supports of the survival functions $S_{x(\cdot)}$ are $[0, \infty)$ when $\gamma \geq 0$ and $[0, sp_{x(\cdot)})$ with finite right ends when $\gamma < 0$.

Suppose that $HR(0, x_1, x_2) = r(x_2)/r(x_1) = c_0 > 1$.

The hazard ratio $HR(t, x_1, x_2), x_1, x_2 \in E_1$, has the following properties:

a) if $\gamma > 0$, then the ratio of the hazard rates decreases from the value $c > 0$ to 1, i.e. the hazard rates approach one another and meet at infinity.

b) if $\gamma = 0$ (PH model), the ratio of the hazard rates is constant.

c) if $\gamma < 0$, then the ratio of the hazard rates increases from the value $c_0 > 1$ to ∞, end the infinity is attained at the point $sp_{x_2} = \Lambda_0^{-1}\{-1/(\gamma r(x_2))\}$. The hazard rates go away one from another quickly when t increases.

The GPH3 model is a generalization of the *gamma frailty model with explanatory variables*: the GPH model with $\gamma = 1/k > 0$ is obtained taking the frailty variable Z which follows the gamma distribution with the density

$$p_Z(z) = \frac{z^{k-1}}{\theta^k \Gamma(k)} e^{-z/\theta}, \quad z > 0.$$

All the three GPH models are considered as semiparametric: finite-dimensional parameters β and γ and unknown baseline function Λ_0 are the unknown parameters.

5. Cross-effect models of survival functions

When analysing survival data from clinical trials, cross-effects of survival functions are sometimes observed. A classical example is the well-known data concerning effects of chemotherapy (CH) and chemotherapy plus radiotherapy (CH+R) on the survival times of gastric cancer patients, see Stablein and Koutrouvelis (1985), Kleinbaum (1996), Klein and Moeschberger (1997), Hsieh (2001), Bagdonavičius et al. (2002), Wu, Hsieh and Chen (2002), Bagdonavičius, Hafdi and Nikulin (2004). At the beginning of treatment mortality of CH+R patients is greater but at a certain moment the survival functions of CH+R and CH patients intersect and later mortality of CH patients is greater, i.e. if patients survive CH+R therapy during a certain period then later this treatment is more beneficial than CH therapy. This dataset is to be further explained and implemented in Section 6. Doses of CH and R therapy can be different so *regression data with not necessarily dichotomous covariates* may be gathered.

Let us consider models proposed for analysis of data with cross-effects of survival functions with possibly time-dependent covariates.

a. Hsieh model

Hsieh (2001) was one of the first who considered the cross-effects models. According to the idea of Hsieh, one possible way to obtain a cross-effect (CE) of hazard rates is to take a power function of Λ_0. Namely Hsieh proposed the following model with cross effects (CE) of the survival functions:

$$\Lambda_x(t) = e^{\beta^T x(t)}\{\Lambda_0(t)\}^{e^{\gamma^T x(t)}}, \quad x(\cdot) \in E. \tag{16}$$

The parameters γ and β are m-dimensional for the Hsieh model. It is a generalization of the PH model taking the power $e^{\gamma^T x(t)}$ of $\Lambda_0(t)$ instead of the power 1. Actually it is derived from a *linear transformation model with heteroscedastic errors.* It is easy to show that the Hsieh model (HM) implies that the hazard ratio $HR(t)$ between strata x_2 and x_1:

$$HR(t, x_1, x_2) = \frac{\lambda_{x_2}(t)}{\lambda_{x_1}(t)}, \quad x_1, x_2 \in E_1$$

is increasing from 0 to ∞ or decreasing from ∞ to 0. So we have a cross-effect of hazard rates and of survival functions (Wu, Hsieh and Chen (2002), Wu (2004, 2007)).

Estimating equations for the parameters β and γ in the Hsieh model can be obtained from taking derivatives of the "partial likelihood" with respect to β and γ separately. (See Johansen (1983) and Section 8 of this article.) This results in the following estimating-equation processes:

$$\sum_{i=1}^{n}\int_{0}^{\tau}\left\{x_i(u)-\frac{\sum_j Y_j(u)x_j(u)e^{(\beta+\gamma)^T x_j(u)}\{\Lambda_0(u)\}^{e^{\gamma^T x_j}-1}}{\sum_j Y_j(u)e^{(\beta+\gamma)^T x_j(u)}\{\Lambda_0(u)\}^{e^{\gamma^T x_j}-1}}\right\}dN_i(u)=0$$

and

$$\sum_{i=1}^{n}\int_{0}^{\tau}\left\{v_i(u)-\frac{\sum_j Y_j(u)v_j(u)e^{(\beta+\gamma)^T x_j(u)}\{\Lambda_0(u)\}^{e^{\gamma^T x_j}-1}}{\sum_j Y_j(u)e^{(\beta+\gamma)^T x_j(u)}\{\Lambda_0(u)\}^{e^{\gamma^T x_j}-1}}\right\}dN_i(u)=0,$$

where $v_i(t)=x_i(t)\exp(\gamma^T x_i(t))\ln\{\Lambda_0(t)\}$ and τ is the maximal observation time. The definition of *at-risk indicator process* $Y_i(t)$ and *counting process* $N_i(t)$ will be explained in Section 8. For the baseline cumulative hazard $\Lambda_0(t)$, Hsieh (2001) derived the following estimating equation

$$\Lambda_0(t)=\sum_{i=1}^{n}\int_{0}^{t}\frac{dN_i(u)}{\sum_j Y_j(u)e^{(\beta+\gamma)^T x_j(u)}\{\Lambda_0(u)\}^{e^{\gamma^T x_j}-1}}$$

and used a *histogram sieve* to approximate $\Lambda_0(\cdot)$.

b. Simple cross-effects model

Here we consider another model with cross-effect of survivals. The SCE model, proposed by Bagdonavičius and Nikulin (2002), of survival functions can be obtained from the first GPH model considered in the previous section replacing the scalar parameter γ by $e^{\gamma^T x(t)}$ in the formula (13). According to this SCE model of Bagdonavicius and Nikulin we have

$$\lambda_x(t)=e^{\beta^T x}\{1+\Lambda_x(t)\}^{1-e^{\gamma^T x}}\,\lambda_0(t)$$
$$=e^{\beta^T x}\{1+e^{(\beta+\gamma)^T x}\Lambda_0(t)\}^{e^{-\gamma^T x}-1}\,\lambda_0(t),\quad x\in E_1. \qquad (17)$$

The regression parameters γ and β are m-dimensional for this SCE model. Suppose that at the point $t=0$ the hazard ratio of the hazard rates

$$HR(t,x_1,x_2)=\frac{\lambda_{x_2}(t)}{\lambda_{x_1}(t)},\quad x_1,x_2\in E_1,$$

under constant covariates x_1 and x_2 is greater then 1:

$$HR(0,x_1,x_2)=e^{\beta^T(x_2-x_1)}=c_0>1\quad\text{and}\quad \gamma^T(x_1-x_2)<0.$$

In this case the hazard ratio $HR(t, x_1, x_2)$ decreases from the value $c_0 > 1$ to 0, i.e. the hazard rates intersect once. The survival functions S_{x_1} and S_{x_2} also intersect once in the interval $(0, \infty)$. (See Bagdonavičius and Nikulin (2002, 2005) for more details.)

Note that the difference between the Hsieh model and the SCE model is the following. In the case of the SCE model the ratios of the hazard rates and even the ratios of the cumulative hazards go to ∞ (or 0) as $t \to 0$. In the case of the SCE model these ratios are defined and finite at $t = 0$. This property of the SCE model is more natural and helps avoid complications when seeking efficient estimators.

c. Multiple cross-effects model

Another interesting model with cross-effects of survival was proposed recently by Bagdonavičius and Nikulin (2005). According to this so-called *Multiple Cross-Effects* (MCE) model we have

$$\lambda_x(t) = e^{\beta^T x} \left(1 + \gamma^T x \Lambda_0(t) + \delta^T x \Lambda_0^2(t)\right) \lambda_0(t), \quad x \in E_1. \tag{18}$$

The regression parameters β, γ and δ are m-dimensional for this MCE model.

In case one concerns the situation of "homogeneity effect" (or "no lifetime regression"), we can see that it takes place if $\gamma = 0$ and $\beta = 0$ for Hsieh model, $\gamma = 0$ and $\beta = 0$ for SCE model, and $\beta = \delta = \gamma = 0$ for MCE model.

Here we give a brief summary on the properties of the models considered:

Hsieh model: the hazard rates and the survival function do not intersect or intersect once in the interval $(0, \infty)$.

SCE model: the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival function do not intersect or intersect once in the interval $(0, \infty)$.

MCE model: the ratios of the hazard rates increase, decrease or are constant, the hazard rates and the survival function do not intersect, intersect once or twice in the interval $(0, \infty)$.

Note also that in the case of the GPH model the ratio of the hazard rates is increasing, decreasing or constant in time, the hazard functions and the survival functions do not intersect in the interval $(0, \infty)$.

The PH is a particular case of all four models mentioned here.

Other CE models can be obtained using the same procedure for the second and the third GPH models.

At the end of this section we note that in the case of the generalization of the positive stable frailty model given by Aalen (1994):

$$\lambda_x(t) = r(x)(1 + r(x)\mu^{-1}\delta\Lambda_0(t))^{\mu}\lambda_0(t), \quad x \in E_1,$$

with $\mu > 1$ the hazard rates intersect but the survival functions do not intersect.

6. Examples of Cross-Effect Models

In this section we give examples to illustrate possible applications of the Hsieh model and the SCE model; both of them describe survival data with one crossing point. To avoid copious exposition, only the example of gastric cancer data (stated in Section 5) will be implemented in details. The gastric cancer data consists of 90 patients. They were randomized into two groups: one group (45 patients) received chemotherapy (CH) and the other received chemo- plus radio-therapy (CH+R). The Kaplan-Meier (KM) survival estimates exhibit a crossing point around 2.7 years. Before this point, the CH-group has superior survival probability than the CH+R group. After 2.7 years, however, the relative effect reverses. This kind of cross-effects can be nicely captured by the Hsieh model and the SCE model. In Figure 1, KM survival estimates and the estimates based on the Hsieh model are plotted. From the pattern of KM estimates we can easily see the limitation of using PH analysis. In Figure 2, the survival functions estimated from the SCE model are compared with the KM estimates. By contrasting these two figures, the estimates of survivor functions based on the SCE model has a better fit than the Hsieh model. This is due to the fact that Hsieh (2001) used the piecewise constant approximation, in which each 'piece' of time segment contains $O(n^{\frac{2}{3}})$ observations. Hsieh's approach is basically a parametric approximation to a semi-parametric problem. On the other hand, the SCE model and the associated estimates are semiparametric efficient (Bagdonavičius, Hafdi, and Nikulin, 2004); and, based on the modified score equations, the estimates of survival are rather "smooth". Moreover, the SCE model has a 'dynamic meaning' in its model formulation.

Although the gastric cancer data is a two-sample example, we should note that both of the Hsieh and SCE models can be implemented with a regression setting. These two models naturally render a score-type test for equality between the two groups (CH and CH+R). Compared with a class of weighted log-rank tests, the score tests based on the Hsieh and

SCE models give significant difference between the two treatment groups (Bagdonavičius, Hafdi, and Nikulin, 2004; Wu, 2007.)

An important feature of the current two models is that: the 'hazard ratio' (denoted as 'HR') according to two covariate values z_i and z_j is diverse in time. Since $HR(t) = \lambda(t; x_i)/\lambda(t; x_j)$, a simple plot of $HR(t)$ versus t under the fitted Hsieh or SCE models reveals their time-dependent property. That also means that the two models have heterogeneity effect in the space of "covariate" and "time" (Wu, 2006).

Finally, we should mention that there are many actual datasets in oncology and clinical studies (which concern the lifetime distributions of some groups) possess the cross-effect property. See Wu, Hsieh, and Chen (2002) for the analysis of a set of bile duct cancer data, Piantodosi (1997) for a lung cancer dataset, and Klein and Moeschberger(1997) for more examples with cross-effects.

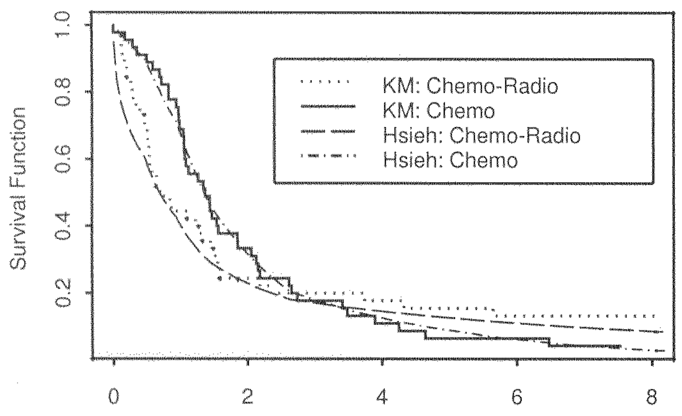

Fig. 1. MK and Hsieh estimates of survivals (in Years) of gastric cancer patients.

7. Some extentions for the PH and the AFT models

An interesting alternative of the PH hazards model is the *additive hazards* (AH) model:

$$\lambda_{x(\cdot)}(t) = \lambda_0(t) + \beta^T x(t), \quad x(\cdot) \in E, \tag{19}$$

where β is the vector of regression parameters. If the AH model holds then the difference of hazards rates under constant covariates does not

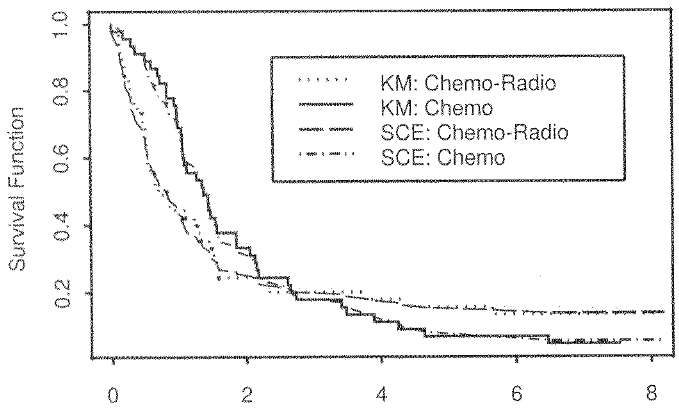

Fig. 2. MK and SCE estimates of survivals (in Years) of gastric cancer patients.

depend on t. As the PH model, this model has the absence of memory property: the hazard rate at the moment t does not depend on on the values of the covariate before the moment t. Usually the AH model is used in the semiparametric form: the parameters β and the baseline hazard rate λ_0 are supposed to be unknown.

Both the PH and AH models are included in the so-called *additive-multiplicative hazards* (AMH) model:

$$\lambda_{x(\cdot)}(t) = e^{\beta^T x(t)}\lambda_0(t) + \gamma^T x(t), \quad x(\cdot) \in E, \tag{20}$$

proposed by Lin and Ying (1996). Note that the *Lin-and-Ying model* also has the absence of memory property.

Analogously as in the case of the PH model the AH model can be generalized by the *generalized additive hazards* (GAH) model:

$$\lambda_{x(\cdot)}(t) = q\{A_{x(\cdot)}(t)\}(\lambda_0(t) + \beta^T x(t)), \quad x(\cdot) \in E, \tag{21}$$

where the function q is parametrized as in the case of GPH models.

Both the GPH and the GAH models, which are the extentions of the PH and the AFT models, can be included into the *generalized additive-multiplicative hazards* (GAMH) model:

$$\lambda_{x(\cdot)}(t) = q\{A_{x(\cdot)}(t)\}\left(e^{\beta^T x(t)}\lambda_0(t) + \delta^T x(t)\right). \quad x(\cdot) \in E, \tag{22}$$

studied by Bagdonavicius and Nikulin (1998, 2002). In both GAH and GAMH models the function q is parametrized as in the GPH models:

$q(u) = (1+u)^{-\gamma+1}$, $(1+\gamma u)^{-1}$, $e^{-\gamma u}$, and the GAH1, GAH2, GAH3 or GAMH1, GAMH2 and GAMH3 models are obtained.

8. Computations in parametric and semiparametric estimation

The literature on parametric and non-parametric estimation for the above considered models is enormous, see, for example, Wu, Hsieh and Chen (2002), Wu (2004, 2006), Dabrowska (2005, 2006), Bagdonavicius and Nikulin (2002, 2005), Martinussen and Scheike (2006), etc.

Methods of estimation depend on experimental plans, censoring mechanism, covariate types, etc. Here we only give two general approaches, for parametric and semiparametric cases, which work well for all models.

For *parametric* models, the maximum likelihood (ML) estimation procedure gives the best estimators. For simplicity let us consider only the "right-censored" data (which is typical in survival analysis):

$$(X_1, \delta_1, x_1(\cdot)), \cdots, (X_n, \delta_n, x_n(\cdot))),$$

where $X_i = min(T_i, C_i)$, $\delta_i = \mathbf{1}_{\{T_i \le C_i\}}$ for $i = 1, \cdots, n,$; T_i and C_i are the failure and censoring times, $\mathbf{1}_A$ is the indicator of an event A, and $x_i(\cdot)$ is the covariate corresponding to the ith subject. Equivalently, right censored data can be presented in the following form:

$$(N_1(t), Y_1(t), x_1(t), t \ge 0), \cdots, (N_n(t), Y_n(t), x_n(\cdot), t \ge 0),$$

where $N_i(t) = \mathbf{1}_{\{X_i \le t, \delta_i = 1\}}$ is the counting process and $Y_i(t) = \mathbf{1}_{\{X_i \ge t\}}$ is the "at-risk" process for subject i respectively. In this case, for any $t > 0$, $N(t) = \sum\limits_{i=1}^{n} N_i(t)$ and $Y(t) = \sum\limits_{i=1}^{n} Y_i(t)$ are the number of observed failures in the interval $[0, t]$ and the number of at-risk subjects just prior to time t. Further, the *compensators* of the counting processes N_i with respect to the *history* of the observed processes are $\int Y_i \lambda_i du$.

Suppose that the survival distributions of all n subbjects given $x_i(\cdot)$ are *absolutely continuous* with survival functions $S_i(t, \theta)$ and hazard rates $\lambda_i(t, \theta)$, indexed by $\theta \in \Theta \subset \mathbf{R}^s$; and that the distributions of C_i. and $x_i(\cdot)$ do not depend on θ. The likelihood function for θ is $L(\theta) = \prod\limits_{i=1}^{n} \lambda_i^{\delta_i}(X_i, \theta)\, S_i(X_i, \theta)$. Or, more elaborately,

$$L(\theta) = \prod_{i=1}^{n} \left(\int_0^\infty \lambda_i(u, \theta)\, dN_i(u) \right)^{\delta_i} \exp\left\{ - \int_0^\infty Y_i(u) \lambda_i(u, \theta)\, du \right\}. \quad (23)$$

The ML estimator $\hat{\theta}$ of the parameter θ maximizes the likelihood function. It satisfies the equation $U(\hat{\theta}) = 0$, where U is the score function:

$$U(\theta) = \frac{\partial}{\partial\theta} \ln L(\theta) = \sum_{i=1}^{n} \int_0^{\infty} \frac{\partial}{\partial\theta} \log \lambda_i(u,\theta)\{dN_i(u) - Y_i(u)\lambda_i(u,\theta)du. \tag{24}$$

The form of the hazard rates λ_i for the PH, AFT, GPH, AH ,AMH, GAH, GAMH are given by the formulas (2), (7), (13), (14), (15), (19), (20), (21), (22). The parameter θ contains the regression parameter β, the complementary parameter γ (for some models) and the baseline hazard λ_0 (which is taken from some parametric family).

For *semiparametric* estimation, we expose shortly a general approach proposed by Bagdonavičius and Nikulin (1998, 2002) based on the 'modified partial likelihood' when λ_0 is supposed to be unknown. (See also Dabrowska (2005, 2006), Martinussen and Scheike (2006).) The martingale property of the difference

$$N_i(t) - \int_0^t Y_i(u)\lambda_i(u,\theta)du \tag{25}$$

implies a "pseudo estimator" (which depends on θ) of the baseline cumulative hazard Λ_0. Indeed, all the above considered models can be classified into two groups depending on the form of $\lambda_i(t,\theta)dt$. It is of the form

$$g(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta)d\Lambda_0(t)$$

for PH, GPH, CE models, $d\Lambda_0(f_i(t,\theta))$ for AFT model, and

$$g_1(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta)d\Lambda_0(t) + g_2(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta)dt$$

for AH, AMH, GAH, GAMH models. Estimation for the PH and AFT models with time-dependent regression coefficients and time-dependent covariates is analogous to the estimation for the PH and AFT models with constant regression coefficients and properly chosen time-dependent "covariates".

For the first group the martingale property of the difference (25) implies the recurrently defined "estimator":

$$\tilde{\Lambda}_0(t,\theta) = \int_0^t \frac{dN(u)}{\sum_{j=1}^{n} Y_j(u)g(x_j(v), \tilde{\Lambda}_0(v,\theta), 0 \le v < u, \theta)}.$$

For the second group

$$\tilde{\Lambda}_0(t,\theta)=\sum_{i=1}^{n}\int_0^t \frac{dN_i(h_i(u,\theta))}{\sum\limits_{l=1}^{n} Y_l(h_l(u,\theta))},$$

where $h_i(u,\theta)$ is the function inverse to $f_i(u,\theta)$ with respect to the first argument.

Note that for PH, GPH1, GPH2, GPH3 models

$$g(x(s),\Lambda_0(s),0\le s\le t,\theta)=e^{\beta^T x(t)},$$

$$e^{\beta^T x(t)}(1+\gamma\int_0^t e^{\beta^T x(u)}d\Lambda_0(u))^{\frac{1}{\gamma}-1},$$

$$e^{\beta^T x(t)}(1+2\gamma\int_0^t e^{\beta^T x(u)}d\Lambda_0(u))^{-\frac{1}{2}},\quad e^{\beta^T x(t)}(1+\gamma\int_0^t e^{\beta^T x(u)}d\Lambda_0(u))^{-1},$$

respectively. For the CE model

$$g(x(s),\Lambda_0(s),0\le s\le t,\theta)=e^{\beta^T x(t)}\{1+\Lambda_{x(\cdot)}(t)\}^{1-e^{\gamma^T x(t)}},$$

where the function $\Lambda_{x(\cdot)}$ is defined by the equation

$$\int_0^t e^{\beta^T x(u)}\{1+\Lambda_{x(\cdot)}(u)\}^{1-e^{\gamma^T x(u)}}d\Lambda_0(u)=\Lambda_{x(\cdot)}(t).$$

If x is constant in time then for the CE model

$$g(x,\Lambda_0(s),0\le s\le t,\theta)=e^{\beta^T x}\{1+e^{(\beta+\gamma)^T x}\Lambda_0(t)\}^{e^{-\gamma^T x}-1}.$$

For the AFT model

$$f_i(t,\theta)=\int_0^t e^{-\beta^T x(u)}du.$$

For the AH, AMH, GAH and GAMH models

$$g_1(x_i(s),\Lambda_0(s),0\le s\le t,\theta)=1,\quad e^{\beta^T x(t)},\quad x^T,\quad x_1^T$$

and

$$g_2(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta) = \beta^T x(t), \quad \beta^T x(t), \quad 0, \quad \beta_2^T x(t),$$

respectively. For the GAMH1 model (formulas are analogous for the GAMH2, GAMH3, GAH1, GAH2, GAH3 models):

$$g_1(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta) = e^{\beta^T x(t)}\, g(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta),$$

$$g_2(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta) = \delta^T x(t)\, g(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta),$$

where

$$g(x_i(s), \Lambda_0(s), 0 \le s \le t, \theta)$$
$$= \left(1 + \gamma(\int_0^t e^{\beta^T x(u)} d\Lambda_0(u) + \delta^T \int_0^t x(u)du)\right)^{\frac{1}{\gamma}-1}.$$

For the PH, GPH and CE models the weight $\frac{\partial}{\partial\theta}\log\lambda_i(u,\theta)$ in (24) is a function of $x_i(\cdot)(v), \Lambda_0(v), 0 \le v \le u$ and θ. So the modified score function is obtained from replacing Λ_0 by a consistent estimator $\tilde\Lambda_0$ in the parametric score function (24).

In the case of the AFT, AH, AMH models the weight depends not only on Λ_0 but also on λ_0 and/or λ_0'. But the more important thing is that $\lambda_i(u)du$ does not depend on λ_0 and λ_0'. So construction of the modified likelihood function can be done by different ways. One way is to replace Λ_0 by $\tilde\Lambda_0$, λ_0 and λ_0' by nonparametric kernel estimators which are easily obtained from the estimator $\tilde\Lambda_0$.

Computation of the modified likelihood estimators is simple for the PH, GPH and CE models. It is due to the remarkable fact that these estimators can be obtained by writing the partial likelihood function as:

$$L_P(\theta) = \prod_{i=1}^n \left[\int_0^\infty \frac{g\{x_i(v), \Lambda_0(v), 0 \le v \le u, \theta\}}{\sum_{j=1}^n Y_j(u) g\{x_j(v), \Lambda_0(v), 0 \le v \le u, \theta\}}\, dN_i(u)\right]^{\delta_i}, \quad (26)$$

and suppose at first that Λ_0 is known. If Λ_0 in the score function is substituted by $\tilde\Lambda_0$, then exactly the same modified score function is obtained as if coming from the full likelihood. So computing the estimator $\hat\theta$ in the score equation is not needed.

In conclusion, it is better to maximize the modified partial likelihood function, which is simply the partial likelihood function (26) except for replacing Λ_0 by $\tilde{\Lambda}_0$. The general quasi-Newton optimization algorithm (such as that given in R language) works very well for seeking the value of θ which maximizes the target function. Note that the modified maximum likelihood estimators are the values of θ minimizing the distance of the modified score function from zero. Computational methods for such estimators are also given in Lin and Geyer (1992), Wu, Hsieh and Chen (2002), Wu (2007), Dabrowska (2005, 2006), Martinussen and Scheike (2006).

References

1. O. Aalen, *A model for nonparametric regression analysis of counting processes.* — Lect. Notes Statists. **2** (eds. W. Klonecki, A. Kozek and J. Rosinski), New York, Springer Verlag (1980), 1–25.
2. V. Bagdonavičius and M. Nikulin, *Stochastic models of accelerated life.* Advanced Topics in Stochastic Modelling (eds. J. Gutienez, M. Valderrama), World Scient., Singapore (1994).
3. V. Bagdonavičius and M. Nikulin, *Additive and Multiplicative Semiparametric Models in Accelerated Life Testing and Survival Analysis.* — Queen's Papers in Pure and Applied Mathematics, **108**, Kingston, Ontarioo, Canada (1998).
4. V. Bagdonavičius and M. Nikulin, *Generalized proportional hazards model based on modified partial likelihood.* — Lifetime Data Analysis **5** (1999), 329–350.
5. V. Bagdonavičius, M. Hafdi, and M. Nikulin, *The Generalized Proportional Hazards Model and its Application for Statistical Analysis of the Hsieh Model.* — In: Proceedings of The Second Euro-Japanese Workshop on Stochastic Risk Modelling for Finance, Insurance, Production and Reliability, September 18–20, Chamonix, France (eds. T. Dohi, N. Limnios, S. Osaki) (2002), p. 42–53.
6. V. Bagdonavičius, M. Hafdi, K. El Himdi, and M. Nikulin, *Analyse du modèle des hazards proportionnels généralisé. Application sur les donnés du cancer des poumons.* Preprint 0201, I.F.R. "Santé Publique" (2002).
7. V. Bagdonavičius, M. Hafdi, K. El Himdi, and M. Nikulin, *Analysis of Survival Data with Cross-Effects of Survival Functions. Applications for Chemo and Radiotherapy Data.* Preprint 0202, I.F.R. "Santé Publique" (2002).
8. V. Bagdonavičius and M. Nikulin, *Accelerated Life Models: Modeling and Statistical Analysis.* Boca Raton: Chapman and Hall/CRC (2002).
9. V. Bagdonavičius and M. Nikulin, *Statistical analysis of survival and reliability data with multiple crossings of survival functions.* — Comptes Rendus de l'Academie des Sciences de Paris, Ser. I **t.340** (2005), 377–382.
10. V. Bagdonavičius and M. Nikulin, *Analysis of survival data with non-proportional hazards and crossings of survival functions.* — In: Quantitative Methods in Cancer and Humain Risk Assesment (eds. L. Edler, Ch. Kitsos), J. Wiley and Sons, N.Y. (2005), 193–209.
11. S. Bennett, *Analysis of survival data by the proportional odds model.* — Statistics in Medicine **2** (1983), 273–277.

12. D. R. Cox, *Regression models and life tables.* — J. R. Statist. Soc., B **34** (1972), 187–220.
13. D. R. Cox, *Partial likelihood.* — Biometrika **62** (1975), 269–276.
14. D. R. Cox and D. Oakes, *Analysis of Suvival Data.* — Methuen (Chapman and Hall), New York (1984).
15. D. M. Dabrowska and K. A. Doksum, *Partial likelihood in Transformations Models with Censored Data.* — Scand. J. Statist. **15** (1988), 1–23.
16. D. Dabrowska, *Quantile regression in transformation models.* — Sankhya **67** (2005), 153–187.
17. D. Dabrowska, *Information bounds and efficient estimation in a class of censored transformation models.* — Webpage: arXiv:math.ST/0608088 v1 3 Aug 2006 (2006).
18. D. Dabrowska, *Estimation in a class of semi-parametric transformation models.* — In: Second Eric L. Lehmann Symposium – Optimality (ed. J. Rojo), Institute of Math. Statistics, Lecture Notes and Monograph Series **49** (2006), 166–216.
19. P. Greenwood and M. Nikulin, *A Guide to Chi-Squared Testing.* John Wiley and Sons, New York (1996).
20. P. Hougaard, *Survival models for heterogeneous populations derived from stable distributions.* — Biometrika **73**, 3 (1986), 387–396.
21. F. Hsieh, *On heteroscedastic hazards regression models: theory and application.* — Journal of the Royal Statistical Society, Series B **63** (2001), 63–79.
22. S. Johansen, *An extension of Cox's regression model.* — Internat. Statist. Rev. **51** (1983), 165–174.
23. K. Chen, Z. Jin, and Z. Ying, *Semiparametric analysis of transformation models with censored data.* — Biometrika **89**, 3 (2002), 639–668.
24. D. Kleinbaum, *Survival Analysis: A Self-Learning text.* Springer-Verlag, New York (1996).
25. J. P. Klein and M. L. Moeschberger, *Survival Analysis.* Springer, New York (1997).
26. J. F. Lawless, *Statistical Models and Methods for Lifetime Data.* J. Wiley, New York (1982).
27. D. Y. Lin and C. J. Geyer, *Computational methods for semiparametric linear regression with censored data.* — J. Comput. Graph. Statist. **1** (1992) 77–90.
28. D. Y. Lin and Z. Ying, *Semiparametric analysis of the general additive-multiplicative hazard models for counting processes.* — The Annals of Statistics **23**, 5 (1996), 1712–1734.
29. T. Martinussen and T. Scheike, *Dynamic regression models for survival analysis.* Springer, New York (2006).
30. I. W. McKeague and P. D. Sasieni, *A partly parametric additive risk model.* — Biometrika **81**, #3 (1994), 501–514.
31. S. A. Murphy and A. W. van der Vaart, *MLE in the proportional odds model.* — Journal of the American Statistical Association **92** (1997), 968–976.
32. S. Piantadosi, *Clinical Trials.* J. Wiley and Sons, New York (1997).
33. N. M. Sedyakin, *On one physical principle in reliability theory. — Techn. Cybernetics.* **3** (1966), 80–87.
34. D. M. Stablein and I. A. Koutrouvelis, *A two sample test sensitive to crossing hazards in uncensored and singly censored data.* — Biometrics **41** (1985), 643–652.

35. H-D. I. Wu, *Effect of Ignoring Heterogeneity in Hazards Regression.* — In: Parametric and Semipametric Models with Applicatiions to Reliability, Survival Analysis, and Quality of Life (eds. M. Nikulin, N. Balakrishnan, M. Mesbah, N. Limnios). Birkhauser: Boston (2004), 239–250.
36. H-D. I. Wu, *Statistical Inference for two-sample and regression models with heterogeneity effects: a collected-sample perspective.* — In: Probability, Statistics and Modelling in Public Health (eds. M. Nikulin, D. Commenges, C. Huber), Springer, New York (2006), 452–465.
37. H-D. I. Wu, *A Partial score test for difference among heterogeneous populations.* — Journal of Statistical Planning and Inference **137** (2007), 527–537.
38. H-D. I. Wu and F. Hsieh, *Heterogeneity and varying effect in hazards regression* — Preprint. Schoool of Public Health, China Medical University, Taichung, Taiwan (2002), 19 p.
39. H-D. I. Wu, F. Hsieh, and C.-H. Chen, *Validation of a heteroscedastic hazards regression model.* — Lifetime Data Analysis **8** (2002), 21–34.
40. S. Yang and R. L. Prentice, *Semiparametric inference in the Proportional odds regression model.* — Journal of the American Statistical Association **94** (1999), 125–136.

Université Victor Segalen,
Bordeaux, France
Biostatistics Center and
School of Public Health,
China Medical University, Taiwan

Поступило 1 декабря 2006 г.