PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 41, Fasc. 1 (2021), pp. 77–87
Published online 8.4.2021
doi:10.37190/0208-4147.41.1.6

SCALED FISHER CONSISTENCY OF THE PARTIAL LIKELIHOOD ESTIMATOR IN THE COX MODEL WITH ARBITRARY FRAILTY

BY

TADEUSZ BEDNARSKI (WROCŁAW) AND PIOTR B. NOWAK (WROCŁAW)

Abstract. It is argued that inference based on the Cox regression model and the partial likelihood estimator is possible for various extensions of the model, which in particular include an arbitrary frailty variable. We demonstrate that the estimator in such a general setup is Fisher consistent up to a scaling factor under symmetry type distributional assumptions on explanatory variables. A simulation experiment shows exemplary behaviour of the estimator and also of a test of fit based on the Anderson–Darling statistic for different Cox model extensions.

2020 Mathematics Subject Classification: Primary 62N01; Secondary 62F12.

Key words and phrases: Cox model, frailty models, Fisher consistency.

1. INTRODUCTION

Modeling with survival regression models, common in time-to-event economic data analysis, is nearly always susceptible to omission of influential explanatory variables. In some cases this may cause inferential perturbations that are out of researcher's control. Robust estimation, as proposed for instance by Bednarski [4] and Sasieni [12] for the Cox model [7], was aimed at making estimation resistant to occasional outliers but it did not cope with oversimplified modeling like variable omission.

A common remedy for the estimation problem was then to extend the model by including a frailty variable, which allowed heterogeneity in longevity endowment. Voupel [14] proposed to use a gamma distributed frailty to improve biased estimation for the life tables. There are numerous extensions of the gamma distributed frailty: Murphy [10] shows consistency of the partial likelihood estimator for cumulative baseline and the variance of frailty under very general conditions; Aalen [1] suggested a compound Poisson frailty model; Henderson [8] tried to quantify the bias which may occur in estimated covariate effects and fitted marginal distributions when frailty effects are present in survival data. Aalen et al. [2] give a review of inference for unobserved heterogeneity in survival modeling and give their treatment for counting processes. Wienke [15] gives a comprehensive review of frailty modeling in survival data analysis.

Below we demonstrate that the partial likelihood estimator for the Cox regression model is Fisher consistent up to a scaling factor for a variety of Cox model extensions, which in particular include an arbitrary unobserved frailty variable. Scaled consistency complicates inference for the Cox model in the sense that the relative risk corresponding to two different regressor values is known up to unknown power. One should bear in mind that this happens only when the Cox model with a single unspecified baseline hazard is not good enough to describe the conditional time distribution and when the population is composed of different hazard-homogeneous, not necessarily identifiable strata with a common linear structure. In the most extreme case we can think of every population unit having time distribution depending on a different cumulated hazard. Of course modeling the frailty distribution along with cumulated hazard is a remedy for inconsistent estimation of regression parameters but its limited choice relies on analytical convenience rather than experimental knowledge. Therefore taking a very general Cox model extension and using the partial likelihood at the price of only up-to-scale consistent estimation makes sense.

In our approach to scaled Fisher consistent estimation for the extended Cox models we adapt the general ideas of Ruud [11] who studied a regression binary dependent variable model and showed sufficient conditions for scale consistent estimation of regression parameters. Another important account of such studies is due to Stoker [13] who considered a general regression model where the conditional expectation of the dependent variable y given a vector X of explanatory variables can be written as $E(y \,|\, X) = M(\alpha + \beta^\top X)$ and the function M is misspecified or unknown. Recently, Bednarski and Skolimowska-Kulig [5], [6] have shown that the maximum likelihood estimator for the regression parameters in the classical exponential regression model is also scaled Fisher consistent for the extended model.

2. SCALED FISHER CONSISTENCY FOR EXTENDED COX MODEL

This section introduces notation and explains merits of scale consistent estimation of regression parameters in the Cox model with frailty.

Let T be a random variable denoting the survival time and let $X = (X_1, \ldots, X_p)^\top$ be a vector of covariates with cumulative distribution function H. The simplest version of the Cox model assumes that the conditional survival function of T given $X = x$ has the form

$$P(T > t \,|\, x) = 1 - F(t \,|\, x) = \exp(-\Lambda(t) \exp(\beta^\top x)),$$

where $\beta \in \mathbb{R}^p$ denotes unknown regression parameters and $\Lambda(t) = \int_0^t \lambda(s)\, ds$ is the baseline cumulative hazard function whereas λ is the baseline hazard function.

Equivalently, the conditional hazard function for the survival time T is $\lambda(t \mid x) = \lambda(t)\exp(\beta^\top x)$. It is well known that for the random sample from this model,

$$(T_1,(X_{11},\dots,X_{1p})),(T_2,(X_{21},\dots,X_{2p}))\dots,(T_n,(X_{n1},\dots,X_{np})),$$

and its empirical distribution function denoted by $F_n(t,x)$, the partial likelihood estimator for the Cox model solves the equation

$$\int\left[y - \frac{\int_{t\geqslant w} x\exp(\beta^\top x)\,dF_n(t,x)}{\int_{t\geqslant w}\exp(\beta^\top x)\,dF_n(t,x)}\right]dF_n(w,y) = 0. \tag{2.1}$$

If F_n is replaced by the joint distribution of (T,X) from the Cox model with regression parameter β_0 (the true parameter value), then $\beta=\beta_0$ is the only solution (see Bednarski [4]). The case of right censored time observations is also covered by our considerations, but to simplify the forthcoming formulas we will just briefly comment on it when appropriate.

The conditional survival function given the covariates $X = x$ and the frailty $A = a$ for the Cox model has the form

$$P(T > t \mid x,a) = 1 - F_{\beta_0}(t \mid x,a) = \exp(-\Lambda(t)a\exp(\beta_0^\top x)), \tag{2.2}$$

where β_0 is the true parameter vector and A is a positive (with probability one) random variable with cumulative distribution function G. Equivalently, the hazard function for the survival time T is $\lambda(t \mid x,a) = a\lambda(t)\exp(\beta_0^\top x)$. We further assume, unless otherwise stated, that X and A are independent.

By replacing the empirical distribution function F_n in equation (2.1) by the distribution of the Cox model with frailty $F_{\beta_0}(t,x,a) = F_{\beta_0}(t \mid x,a)H(x)G(a)$ (the true distribution of (T,X,A)) we can define scaled Fisher consistency. Formally, it means that the equation

$$\int\left[y - \frac{\int xe^{\beta^\top x}\exp(-\Lambda(w)ae^{\beta_0^\top x})\,dH(x)\,dG(a)}{\int e^{\beta^\top x}\exp(-\Lambda(w)ae^{\beta_0^\top x})\,dH(x)\,dG(a)}\right]dF_{\beta_0}(w,y,b) = 0 \tag{2.3}$$

is satisfied for $\beta = \alpha\beta_0$, where $\alpha > 0$ is some scaling factor.

One may assume that the explanatory variables in (2.3) are centered at zero. Indeed, denote by μ_0 the expectation of X. Then, after a simple calculation, the left hand side of (2.3) takes the form

$$-\int \frac{\int xe^{\beta^\top x}\exp(-\Lambda_0(w)ae^{\beta_0^\top x})\,dH_0(x)\,dG(a)}{\int e^{\beta^\top x}\exp(-\Lambda_0(w)ae^{\beta_0^\top x})\,dH_0(x)\,dG(a)}\,d\tilde{F}_{\beta_0}(w \mid y,b)\,dH_0(y)\,dG(b),$$

where $H_0(x) = H(x+\mu_0)$, $\Lambda_0(w) = e^{\beta_0^\top \mu_0}\Lambda(w)$, $\lambda_0(w) = e^{\beta_0^\top \mu_0}\lambda(w)$ and $d\tilde{F}_{\beta_0}(w \mid y,b) = be^{\beta_0^\top y}\lambda_0(w)\exp(-\Lambda_0(w)be^{\beta_0^\top y})dw$.

Fisher consistency is a primary step in establishing the asymptotic distribution of solutions to (2.3) when empirical distributions replace the theoretical one. This is routinely done in robustness in the case of M-estimation (see Marona et al. [9]) and in Bednarski [4] for a robust version of the Cox estimation.

By standard asymptotic arguments, scaled Fisher consistency, as formulated above for the Cox model, implies that solutions to equation (2.3), if the true model distributions are replaced by empirical distribution functions, converge to scaled regression parameters as the sample size increases, and they are asymptotically normal.

The convenience of inference for the Cox model, when it is correct, is that for any two values of explanatory vectors x_1, x_2 we can estimate the ratio of hazards (of instantaneous risks) as $\exp(\beta_0^\top x_1)/\exp(\beta_0^\top x_2)$. In the Cox model with frailty, scaled Fisher consistency leads to $(\exp(\beta_0^\top x_1)/\exp(\beta_0^\top x_2))^\alpha$ where α is not known. Therefore if the ratio with unknown power is greater than 1 we can only deduce that the time distribution for an individual with $X = x_1$ is stochastically greater than it is for the same individual with $X = x_2$. But, considering the generality of the model, this is already quite a lot!

3. ANALYTIC RESULTS

For further considerations we assume that $EX = 0$, the explanatory variables are nontrivial (none of them has mass 1 at 0) and all the integrals involved exist and are finite. In what follows, for brevity we introduce the notation $\psi(w, \alpha, y, a) = e^{\alpha y} \exp(-\Lambda(w) a e^y)$.

We define a function $f_\beta : [0, \infty) \to \mathbb{R}$ by

$$f_\beta(\alpha) = \int \left[\frac{\int (\beta^\top x) \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a)}{\int \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a)} \right] dF_\beta(w),$$

where F_β is the marginal time distribution. Note that the scalar product of β_0 and the left hand side of (2.3) for $\beta = \alpha\beta_0$ is equal to $f_{\beta_0}(\alpha)$.

LEMMA 3.1. *For any β the function f_β has the following properties:*

(a) *f_β is continuous and strictly increasing on* $[0, \infty)$.
(b) $f_\beta(0) < 0$.
(c) $\lim_{\alpha \to \infty} f_\beta(\alpha) > 0$.

Proof. (a) The continuity follows from the properties of exponential functions and their integrals. Now, if we set

$$I_1 = \int \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a),$$
$$I_2 = \int (\beta^\top x) \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a),$$
$$I_3 = \int (\beta^\top x)^2 \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a),$$

then

$$\frac{d}{d\alpha} f_\beta(\alpha) = \int \left[\frac{I_1 \cdot I_3 - I_2^2}{I_1^2} \right] dF_\beta(w).$$

Therefore, applying the Cauchy–Schwarz inequality jointly with the fact that H is not concentrated in one point we get $I_1 \cdot I_3 - I_2^2 > 0$ and the assertion follows.

(b) It is enough to demonstrate that the inner integral of $f_\beta(0)$,

$$\int (\beta^\top x) \exp(-\Lambda(w) a e^{\beta^\top x})\, dH(x)\, dG(a),$$

is negative for every w for which $\Lambda(w) > 0$. By our assumption that $\int (\beta^\top x)\, dH(x) = 0$ and the fact that the function $\exp(-\Lambda(w) a e^{\beta^\top x})$ is strictly decreasing with respect to $\beta^\top x$ we conclude that for every $a > 0$,

$$\int (\beta^\top x) \exp(-\Lambda(w) a e^{\beta^\top x})\, dH(x) < 0.$$

Hence if $\Lambda(w) > 0$ we get

$$\int (\beta^\top x) \exp(-\Lambda(w) a e^{\beta^\top x})\, dH(x)\, dG(a) < 0,$$

which proves (b).

(c) Since

$$\lim_{\alpha \to \infty} e^{\alpha \beta^\top x} = \begin{cases} \infty & \text{if } \beta^\top x > 0, \\ 1 & \text{if } \beta^\top x = 0, \\ 0 & \text{if } \beta^\top x < 0, \end{cases}$$

the limit $\lim_{\alpha \to \infty} f_\beta(\alpha)$ equals the limit of

$$\frac{\int_{\beta^\top x > 0} (\beta^\top x) \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a)}{\int_{\beta^\top x > 0} \psi(w, \alpha, \beta^\top x, a)\, dH(x)\, dG(a)}$$

as $\alpha \to \infty$. Since the derivative of the above ratio with respect α is positive, we conclude that the ratio is increasing in α. This completes the proof. ∎

An immediate consequence is

LEMMA 3.2. *There exists* $\alpha_0 > 0$ *such that* $f_{\beta_0}(\alpha_0) = 0$.

Recall that a p-dimensional random vector X has *spherically symmetric distribution* if for every orthogonal $p \times p$ matrix Γ (i.e. $\Gamma\Gamma^\top = \Gamma^\top\Gamma = I$) the random vector ΓX is distributed as X. In order to prove the main result of this paper we need an auxiliary lemma on special properties of conditional expectations with spherically symmetric distributions.

LEMMA 3.3. *Let* X *be a* p*-dimensional random vector which has spherically symmetric distribution. Then for any* $\beta \in \mathbb{R}^p$ *and any* $c \in \mathbb{R}$

$$E[X \mid \beta^\top X = c] = c \frac{\beta}{\|\beta\|^2}.$$

Proof. Obviously, it is enough to prove the assertion for unit vectors β. So, take such a β, choose vectors $v_1, \ldots, v_{p-1}$ that jointly with β form an orthonormal basis of $\mathbb{R}^p$ and let Γ be a matrix with rows $\beta^\top, v_1^\top, \ldots, v_{p-1}^\top$. Then by spherical symmetry

$$\begin{aligned} E(\Gamma X \mid \beta^\top X = c) &= E(\Gamma X \mid \beta^\top \Gamma^\top \Gamma X = c) = E(X \mid (\Gamma\beta)^\top X = c) \\ &= E(X \mid X_1 = c) = (c, 0, \ldots, 0)^\top, \end{aligned}$$

since, again by spherical symmetry, $E(X_i \mid X_1 = c) = 0$ for $i = 2, \ldots, p$. Hence

$$E(X \mid \beta^\top X = c) = \Gamma^\top E(\Gamma X \mid \beta^\top X = c) = \Gamma^\top (c, 0, \ldots, 0)^\top = c\beta,$$

which completes the proof. ■

Now we can formulate the following theorem which gives sufficient conditions for scaled Fisher consistency when the partial likelihood estimator is used for the Cox model with frailty.

THEOREM 3.1. *Suppose the vector $X = (X_1, \ldots, X_p)^\top$ of explanatory variables has spherically symmetric distribution. Then for any positive frailty A independent of X, the partial likelihood estimator is Fisher consistent up to scale for the Cox model with frailty, or equivalently, the equation* (2.3) *is satisfied for $\beta = \alpha\beta_0$, where $\alpha > 0$ is some scaling factor.*

Proof. From Lemma 3.2 it follows that there exists $\alpha_0 > 0$ such that $f_{\beta_0}(\alpha_0) = 0$. Using the notation above and Lemma 3.3 we can write the inner integral from the numerator in (2.3) in the form

$$\begin{aligned} &\int x\psi(w, \alpha_0, \beta_0^\top x, a)\, dH(x) = E(X\psi(w, \alpha_0, \beta_0^\top X, a)) \\ &\quad = E(E(X\psi(w, \alpha_0, \beta_0^\top X, a) | \beta_0^\top X)) = E\bigg(\frac{\beta_0^\top X}{\|\beta_0\|^2}\psi(w, \alpha_0, \beta_0^\top X, a)\bigg)\beta_0 \\ &\quad = \frac{\beta_0}{\|\beta_0\|^2}\int (\beta_0^\top x)\psi(w, \alpha_0, \beta_0^\top x, a)\, dH(x). \end{aligned}$$

Hence

$$\int \frac{\int x\psi(w, \alpha_0, \beta_0^\top x, a)\, dH(x)\, dG(a)}{\int \psi(w, \alpha_0, \beta_0^\top x, a)\, dH(x)\, dG(a)}\, dF_{\beta_0}(w) = \frac{\beta_0}{\|\beta_0\|^2} f_{\beta_0}(\alpha_0) = 0.$$

This completes the proof. ■

Assume now that the vector X of explanatory variables has the form $X = TY$, where T is a nonsingular $p \times p$ matrix and Y is a spherically symmetric random vector. Using Lemma 3.3 one can easily prove that

$$E[X \mid \beta^\top X = c] = TE[Y \mid (T^\top\beta)^\top Y = c] = c\frac{TT^\top\beta}{\|T^\top\beta\|^2} = c\frac{\Sigma_X\beta}{\beta^\top\Sigma_X\beta},$$

where Σ_X is the covariance matrix of X. From the above we can see that the vector X which is a linear transformation of a spherically symmetric vector has again the property that its conditional expectation $E[X \mid \beta^T X = c]$ is linear in c. Hence we immediately get the following corollary.

COROLLARY 3.1. *Suppose the vector X of explanatory variables is a linear transformation of a spherically symmetric vector. Then for any positive frailty A independent of X, the partial likelihood estimator is Fisher consistent up to scale for the Cox model with frailty.*

REMARK 3.1 (Other Cox model extensions). Lemma 3.1 is formulated for the Cox model with frailty independent of X but in fact it holds for a much larger class of Cox model extensions. The frailty variable in the expression $\Lambda(t)a$ can be replaced by $\Lambda(t, a)$ satisfying reasonable regularity conditions. The proof of Theorem 3.1 also holds for Cox model extensions where instead of the frailty variable we have $\Lambda(t, a)$.

REMARK 3.2 (Presence of a censoring variable). One of the common features of time-to-event data in clinical and economic studies is the presence of censored data. The partial likelihood estimator is of course Fisher consistent under censoring. Below we explain why, under right censored survival times and frailty, scaled Fisher consistency holds for the partial likelihood estimator.

Let then, for a censoring variable C and frailty A, $F(t, c, x, a)$ denote the joint distribution of (T, C, X, A) under the Cox model. By independence of T and C given X and independence of X, C and A (we need to assume the independence of frailty and censoring here),

$$dF_{\beta_0}(t, c, x, a) = dF_{\beta_0}(t \mid x, a)dC(c)dH(x)dG(a).$$

Thus, scaled Fisher consistency means that the equation $L(\beta, \beta_0) = 0$ is satisfied for $\beta = \alpha\beta_0$, $\alpha > 0$, where

$$L(\beta, \beta_0) = \int \left[\frac{\int I_{t \wedge c \geqslant w} x \exp(\beta^\top x)\, dF_{\beta_0}(t, c, x, a)}{\int I_{t \wedge c \geqslant w} \exp(\beta^\top x)\, dF_{\beta_0}(t, c, x, a)} \right] I_{w \leqslant \bar{c}}\, dF_{\beta_0}(w, \bar{c}, y, b).$$

It is easy to show that

$$L(\alpha\beta_0, \beta_0) = \int \left[\frac{\int x\psi(w, \alpha, \beta_0^\top x, a)\, dH(x)\, dG(a)}{\int \psi(w, \alpha, \beta_0^\top x, a)\, dH(x)\, dG(a)} \right] [1 - C(w)]\, dF_{\beta_0}(w, y, b).$$

The form of the last equations indicates that Lemma 3.1 and Theorem 3.1 remain applicable.

4. SIMULATIONS

A Monte Carlo experiment was conducted to investigate properties of the partial likelihood estimation when data are generated from the Cox model with frailty (2.2). The R programming language was used and a general purpose *coxph* procedure from the “survival” package was applied. Random covariates with a given correlation matrix were generated using the package “multiRNG”. Two cumulated baseline intensities, $\Lambda(t) = t^{1/2}$ and $\Lambda(t) = t^2$, and a three-dimensional regressor $X = (X_1, X_2, X_3)^\top$ with correlation matrix

$$R = \begin{pmatrix} 1 & 0.8 & 0.8 \\ 0.8 & 1 & 0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}$$

was used in three cases:

(a) X has multivariate normal distribution.
(b) X has multivariate t distribution with 5 degrees of freedom.
(c) X has multivariate Laplace distribution.

For the true parameter value $\beta_0 = (1, -0.5, 0.5)^\top$ two types of frailty variables were used: the first one was squared standard normal plus 1 and the other was three times Bernoulli plus 1 with success probability 0.5. The sample size was 500 and the estimation was repeated 1000 times. All estimates were scaled to one, and empirical means, standard deviations and estimated scales were computed. The results are summarized in Table 1.

The simulation results indicate a good asymptotic behavior of the estimators under elliptical covariates.

5. DIAGNOSTIC FOR THE COX MODEL

In this section we give a preliminary justification for the behaviour of a diagnostic procedure meant to detect a nontrivial frailty variable within the Cox model. We begin with the following observation. If the model is correct (i.e. the frailty variable is degenerate at the point mass $A = 1$), then the random variable $\exp(-\Lambda(T)e^{\beta_0^\top X})$ follows $U(0, 1)$, a uniform distribution over the interval $(0, 1)$.

Hence, the procedure of checking the validity of the model is based on comparing $r_i = \exp(-\hat{\Lambda}(T_i)e^{\hat{\beta}_0^\top X_i})$, $i = 1, \dots, n$, against uniform distribution $U(0, 1)$, where $\hat{\beta}_0$ is the maximum partial likelihood estimator of β_0 and $\hat{\Lambda}$ is the Breslow estimator of cumulative baseline hazard. For the numerical example the Cox model with $\beta_0 = (1.5, -1, 1)^\top$, $\Lambda(t) = t^{1/2}$, normally distributed regressors and frailty based on shifted χ^2 distribution was taken.

The Anderson–Darling test was used for goodness of fit for uniformity of r_i, $i = 1, \dots, n$. The histograms of p-values are presented in Figure 1, respectively for sample size $n = 50$ and $n = 500$. Moreover, the powers of the test are 0.001 and 0.78, respectively. Apparently the procedure is much less sensitive for small sample sizes. The same procedure for the Cox models without frailty gives a p-value of approximately 0.81 with $\mathrm{sd} = 0.16$ $(n = 50)$ and 0.75 with $\mathrm{sd} = 0.21$ $(n = 500)$.

TABLE 1. Results of the simulation experiment. The first and the second vector in each cell refers to mean values and standard deviations of the normalized vector estimates of true parameter values. The third one is composed of the means of ratios of components of estimates and the true parameters.

	$\Lambda(t) = t^{1/2}$	$\Lambda(t) = t^2$
parameter scaled parameter	(1, −0.5, 0.5) (0.8165, −0.4082, 0.4082)	(1, −0.5, 0.5) (0.8165, −0.4082, 0.4082)
The frailty variable: shifted χ^2		
Normal	(0.8187, −0.4054, 0.4068) (0.0891, 0.0877, 0.0895) (0.8875, 0.8789, 0.8820)	(0.8188, −0.4073, 0.4046) (0.0917, 0.0894, 0.0918) (0.8884, 0.8840, 0.8780)
Student	(0.8185, −0.4069, 0.4056) (0.1537, 0.0986, 0.1015) (0.8436, 0.8387, 0.8362)	(0.8188, −0.4051, 0.4069) (0.0757, 0.0719, 0.0694) (0.8955, 0.8861, 0.8900)
Laplace	(0.8178, −0.4065, 0.4074) (0.0934, 0.0908, 0.0882) (0.8880, 0.8827, 0.8848)	(0.8165, −0.4098, 0.4067) (0.0904, 0.0885, 0.0869) (0.8893, 0.8927, 0.8858)
The frailty variable: shifted Bernoulli		
Normal	(0.8178, −0.4065, 0.4074) (0.0927, 0.0884, 0.0867) (0.7971, 0.7924, 0.7943)	(0.8127, −0.4084, 0.4156) (0.0945, 0.0895, 0.0874) (0.7896, 0.7936, 0.8076)
Student	(0.8171, −0.4073, 0.4080) (0.1349, 0.0920, 0.0914) (0.7619, 0.7596, 0.7609)	(0.8178, −0.4071, 0.4069) (0.0776, 0.0730, 0.0723) (0.8069, 0.8033, 0.8029)
Laplace	(0.8142, −0.4087, 0.4123) (0.0915, 0.0863, 0.0879) (0.7913, 0.7943, 0.8013)	(0.8159, −0.4087, 0.4089) (0.0926, 0.0878, 0.0885) (0.7930, 0.7944, 0.7948)

This simple check of the model assumptions is closely related to the test of exponentiality based on Cox–Snell residuals. A detailed discussion of this procedure can be found in Baltazar-Aban and Peña [3].

The frailty variable A is supposed to describe proportional changes of the cumulated hazard $\Lambda(t)$ for individuals within the population. Since

$$P(T > t \,|\, x, a) = \exp(-\Lambda(t) a \exp(\beta_0^\top x)) = \exp(-\Lambda(t) \exp(\beta_0^\top x + \ln(a))$$

it can as well be interpreted as a missing (independent) covariate. In practical data analysis it would be difficult to specify in any reasonable way the distributional form of the missing covariate.

In order to demonstrate the effect of variable's omission the following Cox model was taken: $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^\top = (1.5, -1, 1, 1.5)^\top$, normally distributed covariates and $\Lambda(t) = t^{1/2}$ with fourth covariate unobserved

$$(T_1, (X_{11}, X_{12}, X_{13})), (T_2, (X_{21}, X_{22}, X_{23})), \ldots, (T_n, (X_{n1}, X_{n2}, X_{n3})).$$

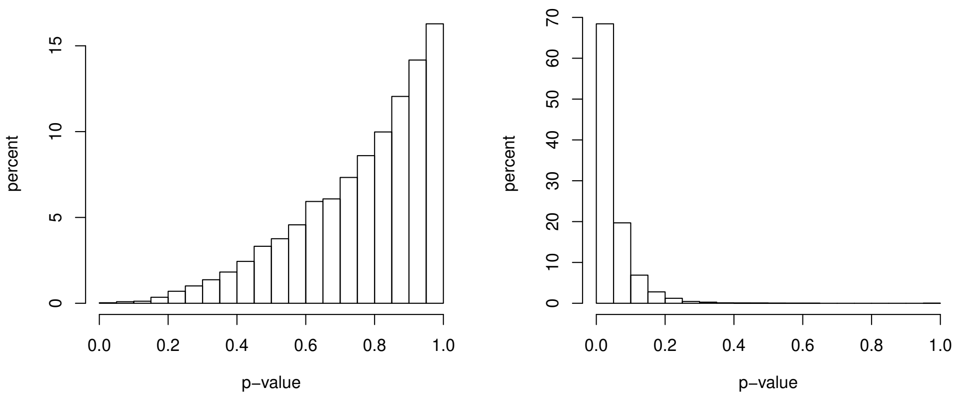

FIGURE 1. Histograms of p-values for the Cox model with frailty based on shifted χ^2 distribution for $n = 50$ and $n = 500$, respectively. The number of runs is 1000.

In view of the above we can only estimate $(\beta_{01}, \beta_{02}, \beta_{03})$ up to some scaling factor. The simulation experiment for 1000 runs and $n = 500$ shows that the mean empirical estimates scaled to one are $(0.7285, -0.4845, 0.4844)$ with standard deviations $(0.0675, 0.0591, 0.0597)$, while the true scaled parameter is $(0.7276, -0.4851, 0.4851)$. The scaling factor α is approximately 0.5. In this case the Anderson–Darling test based on $\{r_i\}$ detects departure from the Cox model with power equal to 0.7.

6. FINAL CONCLUSIONS

The paper discusses the problem of consistent estimation in the Cox model with arbitrary frailty. It is shown that despite the violation of the proportionality assumptions, the classical procedure of estimation leads to the consistent estimation of regression parameters up to a scaling factor if in particular covariates are spherically symmetric. More precisely, a sufficient condition for the consistency is that the conditional expectation of the explanatory vector X given $\beta^\top X = c$ is linear in c. Ruud [11] similarly pointed out that this condition is sufficient for consistent estimation up to scale in a regression binary dependent variable model.

This property holds for the family of spherically symmetric distributions and their linear combinations, that is, elliptical distributions. Special cases are multivariate normal distribution, t distribution, logistic distribution or Laplace distribution.

Although elliptical distributions are an important class of distributions, they do not contain discrete distributions. The authors suspect that the proposed sufficient condition can be to some extent relaxed as indicated by numerous simulation experiments conducted during this study.

Acknowledgments. We thank the referee for very constructive criticism.

REFERENCES

[1] O. O. Aalen, *Modelling heterogeneity in survival analysis by the compound Poisson distribution*, Ann. Appl. Probab. 4 (1992), 951–972.

[2] O. O. Aalen, O. Borgan, and H. K. Gjessing, *Survival and Event History Analysis: A Process Point of View*, Springer, New York, 2008.

[3] I. Baltazar-Aban and E. A. Peña, *Properties of hazard-based residuals and implications in model diagnostics*, J. Amer. Statist. Assoc. 90 (1995), 185–197.

[4] T. Bednarski, *Robust estimation in Cox's regression model*, Scand. J. Statist. 20 (1993), 213–225.

[5] T. Bednarski, and M. Skolimowska-Kulig, *Scaled consistent estimation of regression parameters in frailty models*, Acta Univ. Lodziensis Folia Oeconomica 5(338) (2018), 133–142.

[6] T. Bednarski and N. Skolimowska-Kulig, *On scale Fisher consistency of maximum likelihood estimator for the exponential regression model under arbitrary frailty*, Statist. Probab. Lett. 150 (2019), 9–12.

[7] D. R. Cox, *Regression models and life-tables (with discussion)*, J. Roy. Statist. Soc. B 34 (1972), 187–220.

[8] R. Henderson and P. Oman, *Effect of frailty on marginal regression estimates in survival analysis*, J. Roy. Statist. Soc. B 61 (1999), 367–379.

[9] R. A. Marona, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*, Wiley, 2006.

[10] S. A. Murphy, *Consistency in a proportional hazard model incorporating a random effect*, Ann. Statist. 22 (1994), 712–734.

[11] P. Ruud, *Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models*, Econometrica 51 (1983), 225–228.

[12] P. D. Sasieni, *Maximum weighted partial estimates for Cox model*, J. Amer. Statist. Assoc. 88 (1993), 144–152.

[13] T. Stoker, *Consistent estimation of scaled coefficients*, Econometrica 54 (1986), 1461–1481.

[14] J. W. Vaupel, K. G. Manton, and E. Stallard, *The impact of heterogeneity in individual frailty on the dynamics of mortality*, Demography 16 (1979), 439–454.

[15] A. Wienke, *Frailty Models in Survival Analysis*, Chapman and Hall, 2011.

Tadeusz Bednarski, Piotr B. Nowak
Institute of Economic Sciences
University of Wrocław
Faculty of Law, Administration and Economics
Uniwersytecka 22/26
50-145 Wrocław, Poland
E-mail: tadeusz.bednarski@uwr.edu.pl
piotr.nowak2@uwr.edu.pl

Received 1.7.2019;
revised version 14.2.2020