STATISTICS IN TRANSITION new series, June 2022
Vol. 23, No. 2, pp. 185–196, DOI 10.2478/stattrans-2022-0023
Received – 29.11.2021; accepted – 16.03.2022

Scaled Fisher consistency for the partial likelihood estimation in various extensions of the Cox model

Tadeusz Bednarski, **Piotr B. Nowak,**[1]
Magdalena Skolimowska-Kulig[2]

ABSTRACT

The Cox proportional hazards model has become the most widely used procedure in survival analysis. The theoretical basis of the original model has been developed in various extensions. In the recent years, vital research has been undertaken involving the incorporation of random effects to survival models. In this setting, the random effect is a variable (*frailty*) which embraces a variation among individuals or groups of individuals which cannot be explained by observable covariates. The right choice of the frailty distribution is essential for an accurate description of the dependence structure present in the data. In this paper, we aim to investigate the accuracy of inference based on the primer Cox model in the existence of unobserved heterogeneity, that is, when the data generating mechanism is more complex than presumed and described by the kind of an extension of the Cox model with undefined frailty. We show that the conventional partial likelihood estimator under the considered extension is Fisher-consistent up to a scaling factor, provided symmetry-type distributional assumptions on covariates. We also present the results of simulation experiments that reveal an exemplary behaviour of the estimators.

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

1. Introduction

Statistical analysis of time-to-event data through survival regression models has become common practise in a variety of disciplines including mainly demography and medicine but also economics, actuarial science, reliability research and others. The regression framework allows for the inclusion of relevant factors, like gender, socio-economic status, or received treatment, which explain variation among the individuals or items being studied. However, such an analysis is nearly always susceptible to the omission of influential covariates and leaves unexplained variation. In some cases, the unobserved heterogeneity may cause inferential perturbations that are beyond the control of the researcher. One way of accounting for this estimation problem is to extend the model by including an unobserved random effect - a frailty variable, which allowed heterogeneity in longevity endowment. The notion of frailty was introduced and applied to the population data by Vaupel et al. (1979). The term *frailty* indicates that some individuals are frailer than others, that is, the event under consideration is more likely to happen for them. In its classical and mostly applied form, a

[1]Institute of Economic Sciences, University of Wrocław, Poland. E-mail: piotr.nowak2@uwr.edu.pl. ORCID:https://orcid.org/0000-0002-7404-2946.

[2]Institute of Economic Sciences, University of Wrocław, Poland. E-mail: magdalena.skolimowska-kulig@uwr.edu.pl. ORCID:https://orcid.org/0000-0002-4748-7624.

frailty model assumes proportional hazards and includes an unobservable random variable acting multiplicatively on the baseline hazard function. In recent years a number of papers and textbooks have appeared discussing extensions of common survival models to a wide variety of frailty models that are suitable to handle more complex survival data. A comprehensive review of frailty modelling in survival data analysis can be found in Hougaard (2000) and Wienke (2011). Kalbfleisch and Prentice (2002) give detailed theoretical treatments using the counting process theory. More applied presentations are given by Klein and Moeschberger (2003), and Therneau and Grambsch (2000). Aalen et al. (2008) provide an insight into the theoretical and applied structure of frailty models used in survival and event history analysis on the counting processes basis. Henderson and Oman (1999) investigate the consequences of ignoring frailty in analysis and fitting misspecified Cox proportional hazards models to the marginal distributions. The usual approach to statistical inference with unobserved frailty assumes a parametric family of distributions for frailty, usually gamma but also inverse Gaussian, positive stable, compound Poisson, or more general the power variance function family. For particular types of parametric frailty models the maximisation of the marginal likelihood leads to estimates of the parameters in the model, but for semiparametric frailty models more complex estimation techniques are needed (see Duchateau and Janssen, 2008). Certainly, modelling the frailty distribution is a remedy for biased estimation of regression parameters, but its limited choice relies mainly on their mathematical tractability.

Our objective is to investigate whether the traditionally used partial likelihood estimation method can be worthwhile when the model is misspecified, more precisely, the existing heterogeneity is neglected. In our considerations we apply the approach taken by Bednarski and Skolimowska-Kulig (2018, 2019) and Bednarski and Nowak (2021). They focused on the fundamental requirement needed in sound statistical inference about parameters, the Fisher consistency of estimators. They studied the behaviour of the standard estimators, like maximum likelihood for the exponential model or partial likelihood for the Cox model under extended models, with no assumption about distributional structure of frailty. Then, of course, the Fisher consistency condition need not be true, but it is shown that the commonly used procedures for the estimation of regression parameters in certain hazard-based survival models generate consistent up-to-scale estimators for extensions of these models. In the article we demonstrate that the partial likelihood estimator for the Cox regression model is Fisher consistent up to a scaling factor under an extended model with unobserved generalised frailty. The limitation we make is the distribution of covariates, which is assumed to be elliptically symmetric. In our approach to the scaled Fisher consistent estimation, we adapt the general ideas of Ruud (1983) and Stoker (1986) who studied regression coefficient estimators in particular regression models with assumed misspecification, and showed their up-to-scale consistency.

2. Maximisation criterion for the Cox estimator

In this part we remind the criterion that yields the regression parameter estimator in the Cox proportional hazards model (Cox, 1972). We assume that the survival time T, given the covariate vector X, has the conditional distribution function

$$F(t|x) = 1 - \exp\left(-\Lambda(t)e^{\beta' x}\right),$$

where $\Lambda(t) = \int_0^t \lambda(u)du$, λ is the baseline hazard function and β is the parameter vector. Suppose that we observe a sample $(T_i \wedge C_i, X_i)$, $i = 1, 2, \ldots, n$, where the censoring variable C is independent of the survival time T given the covariate vector X. To formulate the partial likelihood let us denote by $t_{(1)} < t_{(2)} < \ldots < t_{(m)}$ the unique failure times. The partial likelihood for the Cox model can be written as

$$L(\beta) = \prod_{i=1}^{m} \frac{e^{\beta' X_{(i)}}}{\sum_{j \in R_i} e^{\beta' X_j}},$$

where the risk set $R_i = \{k : T_k \wedge C_k \geq t_{(i)}\}$ consists of subjects that have not failed or been censored by time $t_{(i)}$ and $X_{(i)}$ is the covariate vector for the subject that fails at $t_{(i)}$.

The Cox estimator maximises $L(\beta)$, or equivalently

$$\left(\prod_{i=1}^{m} \frac{e^{\beta' X_{(i)}}}{\frac{1}{n}\sum_{j \in R_i} e^{\beta' X_{(j)}}}\right)^{1/n}.$$

Thus, it is given as

$$\arg\max_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left(\beta' X_i - \ln \frac{1}{n} \sum_{T_j \geq T_i} e^{\beta' X_j}\right) \mathbb{I}_{\{T_i \leq C_i\}}.$$

If $F_n(t, c, x)$ denotes the empirical distribution function of the sample (T_i, C_i, X_i), $i = 1, 2, \ldots, n$, and sums are replaced by empirical integrals, then the above expression can be stated as

$$\arg\max_{\beta} \int \left(\beta' y - \ln \int_{t \wedge b \geq w} e^{\beta' x} dF_n(t, b, x)\right) \mathbb{I}_{\{w \leq c\}} dF_n(w, c, y).$$

Since F_n converge uniformly in probability to a true distribution F we can expect that under sufficiently stringent conditions, the maximising $\hat{\beta}_n$ converge in probability to

$$\arg\max_{\beta} \int \left(\beta' y - \ln \int_{t \wedge b \geq w} e^{\beta' x} dF(t, b, x)\right) \mathbb{I}_{\{w \leq c\}} dF(w, c, y),$$

if the latter solution exists.

3. Fisher consistency and scaled Fisher consistency

Since the right time censoring present in the Cox regression model plays no essential role in the argumentation presented in the paper we skip it in order to make the notation more concise. If the underlying cumulative distribution F comes from the Cox model, that is there is a parameter vector β_0 and a nonnegative baseline hazard λ yielding $\Lambda(t) = \int_0^t \lambda(s)ds$ such that

$$F(t|x) = 1 - \exp\left(-\Lambda(t)e^{\beta_0' x}\right),$$

then

$$\beta_0 = \arg\max_{\beta} \int \left(\beta' y - \ln \int_{t\geq w} e^{\beta' x} dF(t,x)\right) dF(w,y) \tag{1}$$

for every parameter vector β_0. The last property means that the Cox estimator is Fisher consistent at the model. The consistency is independent of the baseline hazard λ and it holds under censoring independent of survival time T given the covariate values X.

In general, all statistical estimators based on random samples are defined by explicit or implicit functionals of the corresponding empirical distribution functions. If values of such a functional coincide with true parameters when the empirical distribution is replaced by the true model distribution then we say that the functional is Fisher consistent. Without Fisher consistency the estimator cannot even be consistent asymptotically. Therefore, when studying estimation proposals we would put its Fisher consistency property in the first place. In practice, Fisher-consistent functionals (estimators) associated with a given parametric family are used even if we think the family imperfectly describes a real distribution and the discrepancy is deeper than the one resulting from occasional influential errors. It may therefore be justified in some instances to study what happens when the estimator-functional is used under reasonable nonparametric extensions of the original model. We make it precise in the case of the partial likelihood estimator. Define then a Cox model with generalised frailty and regression parameter β_0 via cumulative distribution function of time, conditional on covariates $X = x$ and frailty $A = a$ as

$$F(t|x,a) = 1 - \exp\left(-\Lambda(t,a)e^{\beta_0' x}\right), \tag{2}$$

where the cumulated hazard $\Lambda(t,a) = \int_0^t \lambda(t,a)dt$ takes now into account possibly complex individual changes in time to event distribution. To simplify forthcoming notations we will use the same letter F for model distributions from the Cox model with extended frailty as in the case of the strict Cox model. The scaled Fisher consistency of the Cox estimator means here that for each parameter value β_0 there exists $c > 0$, possibly depending on β_0, such that

$$c\beta_0 = \arg\max_{\beta} \int \left(\beta' y - \ln \int_{t\geq w} e^{\beta' x} dF(t,x,a)\right) dF(w,y,b) \tag{3}$$

for $F(t,x,a) = F(t|x,a)G(x)H(a)$, where G and H are the distributions of covariates and frailty respectively, and the random variables X and A are assumed independent.

4. Results

Consider the extended Cox model with generalised frailty independent on the covariates, given by (2) and the problem of maximisation of

$$\int \left(\beta' y - \ln \int_{t\geq w} e^{\beta' x} dF(t|x,a)dG(x)dH(a)\right) dF(w|y,b)dG(y)dH(b) \tag{4}$$

with respect to β. Since the above expression can be written as

$$\int \left(-\ln \int_{t \geq w} e^{\beta' x} dF_0(t|x,a) dG_0(x) dH(a) \right) dF_0(w|y,b) dG_0(y) dH(b),$$

where $F_0(t|x,a) = 1 - \exp\left(-\Lambda_0(t,a) e^{\beta_0' x}\right)$, $\Lambda_0(t,a) = \Lambda(t,a) e^{\beta_0' EX}$ and G_0 is the distribution of centred covariates $X - EX$, the maximisation can be reduced to covariates with expectation zero.

Lemma 1. *Let β_0 be the true parameter value and the covariate vector with zero mean be such that for every vector β the conditional expectation $E((\beta - proj_{\beta_0}\beta)X|\beta_0' X)$ is almost surely zero. Then β maximising (4) equals $c\beta_0$ for some real c.*

Proof. For F with centred covariates the maximisation of (4) is equivalent to the minimisation of

$$\int \left(\ln \int_{t \geq w} e^{\beta' x} dF(t|x,a) dG(x) dH(a) \right) dF(w|y,b) dG(y) dH(b),$$

which in turn can be stated as

$$\int \left(\ln \int_{t \geq w} e^{\beta' x} dF(t|x,a) dG_1(x|\beta_0' x) dG_2(\beta_0' x) dH(a) \right) dF(w|y,b) dG(y) dH(b), \quad (5)$$

where β_0 denotes the true parameter value, G_1 is the conditional distribution of X given $\beta_0' X$ while G_2 is the distribution of $\beta_0' X$.
Notice that for any nonzero parameter vector β we can write $\beta = c\beta_0 + \beta_1$, where $c\beta_0 = proj_{\beta_0}\beta$ is the orthogonal projection of β on β_0. Then (5) becomes

$$\begin{aligned}
&\int \left(\ln \int_{t \geq w} e^{c\beta_0' x} dF(t|\beta_0' x, a) \int e^{\beta_1' x} dG_1(x|\beta_0' x) dG_2(\beta_0' x) dH(a) \right) \\
&\qquad dF(w|y,b) dG(y) dH(b) = \\
&\int \left(\ln \int_{t \geq w} e^{c\beta_0' x} dF(t|\beta_0' x, a) E\left(e^{\beta_1' X} | \beta_0' X = \beta_0' x\right) dG_2(\beta_0' x) dH(a) \right) \\
&\qquad dF(w|y,b) dG(y) dH(b) \geq \\
&\int \left(\ln \int_{t \geq w} e^{c\beta_0' x} dF(t|\beta_0' x, a) e^{E(\beta_1' X | \beta_0' X = \beta_0' x)} dG_2(\beta_0' x) dH(a) \right) \\
&\qquad dF(w|y,b) dG(y) dH(b). \quad (6)
\end{aligned}$$

If for every vector β the conditional expectation $E(\beta_1' X|\beta_0' X)$ is almost surely zero then the right side of the inequality (6) equals

$$\int \left(\ln \int_{t \geq w} e^{c\beta_0' x} dF(t|x,a) dG(x) dH(a) \right) dF(w|y,b) dG(y) dH(b)$$

and we can conclude that the minimising value of β, if it exists, must be equal to $c\beta_0$. □

The following theorem is an immediate consequence of the above lemma and the scaled

Fisher consistency condition (3).

Theorem 1. *If β_0 is the true parameter value and $E(X|\beta_0'X) \in lin(\beta_0)$ almost surely then the partial likelihood estimator is scaled Fisher consistent under the Cox model with generalised frailty.*

Proof. Notice that for X such that $E(X|\beta_0'X) \in lin(\beta_0)$ almost surely and for any $\beta = c\beta_0 + \beta_1$, where $c\beta_0$ is the orthogonal projection of β on $lin(\beta_0)$, we have $E(\beta_1'X|\beta_0'X) = 0$ almost surely. □

There are important cases when the condition assumed in the above theorem holds. One of them is when X is spherically symmetric, that is if for every orthogonal matrix Γ the random vector ΓX is distributed as X.

Corollary 1. *If X has a spherically symmetric distribution then the partial likelihood estimator is scaled Fisher consistent under the Cox model with generalised frailty.*

The following conclusion results directly from the preceding considerations.

Corollary 2. *If β_0 is the true parameter value, M a nonsingular matrix, $X = MZ$ and $E(Z|\gamma_0'Z) \in lin(\gamma_0)$ almost surely for $\gamma_0 = \beta_0'M$ then the partial likelihood estimator is scaled Fisher consistent under the Cox model with generalised frailty.*

It is also quite straightforward to show that if $X = MZ$, where M is a nonsingular matrix and Z is spherically symmetric, then again we have the scaled Fisher consistency of the partial likelihood estimator.

Corollary 3. *If $X = MZ$ and Z has a spherically symmetric distribution then the partial likelihood estimator is scaled Fisher consistent under the Cox model with generalised frailty.*

Another special case covered by Theorem 1 is considered below.

Corollary 4. *Let the random vector $X = (X_1, \ldots, X_k)'$ be exchangeable, i.e. $(X_{\pi(1)}, \ldots, X_{\pi(p)})'$ and X have the same distribution for any permutation π of the set $\{1, 2, \ldots, n\}$. If $\beta_0' = (b, b, \ldots, b)$, $b \neq 0$, then the partial likelihood estimator is scaled Fisher consistent under the Cox model with generalised frailty.*

Proof. The exchangeability of X implies that $E(X_1|\beta_0'X) = \cdots = E(X_k|\beta_0'X)$. On the other hand $E(X_1 + \cdots + X_k|\beta_0'X) = X_1 + \cdots + X_k$ and finally $E(X_1|\beta_0'X) = (X_1 + \cdots + X_k)/k$. Therefore $E(X|\beta_0'X) \in lin(\beta_0)$ and the Fisher consistency holds. □

5. Simulation studies

In this section we present the results of simulation studies conducted to investigate how the violation of the symmetry assumption on the regressors distribution or the omission of the covariates may affect properties of the partial likelihood estimation of the regression parameters. The experiments also show exemplary up-to-scale estimation under the Cox model with various choices of generalised frailty. All simulations were run with the `R` programming language.

Example 1. *Estimation under non-symmetric covariates*

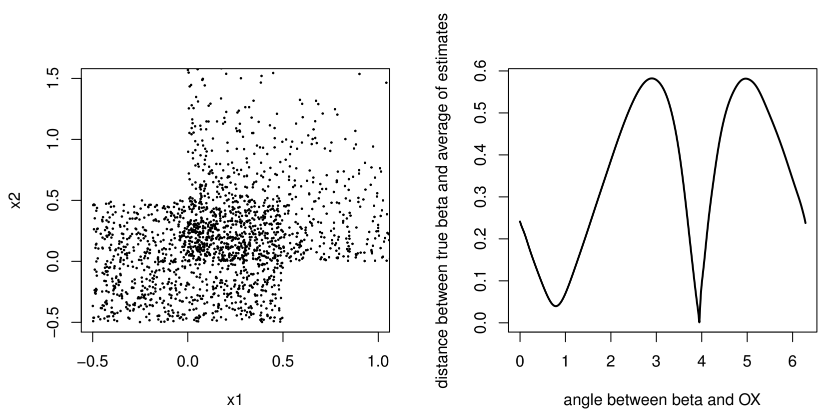
The sample of size 500 was generated from model (2) with $\Lambda(t,a) = ta$ and frailty given as a mixture of two gamma distributions. The covariates are from a two-dimensional uniform distribution mixed with a two-dimensional independent chi-square $\chi^2(2)$ (see the left panel of Figure 1). The curve shown in the right-hand panel of Figure 1 represents the distance $d(\alpha)$ between the true beta $\beta_0' = (\cos(\alpha), \sin(\alpha))$ and the normalized averaged of estimates for $\alpha \in [0, 2\pi]$. The values on x axis are angles between the horizontal axis and true betas.

From the above description it follows that the density of the covariate vector $X = (X_1, X_2)'$ has the form

$$f(x_1,x_2) = \frac{1}{2} \cdot \mathbb{I}_{(-0.5,0.5)^2}(x_1,x_2) + \frac{1}{2} \exp(-(x_1+x_2)/2) \cdot \mathbb{I}_{(0,\infty)^2}(x_1,x_2).$$

Obviously X is not elliptically symmetric, however, since $f(x_1,x_2) = f(x_2,x_1)$, it is exchangeable. Thus, for $\beta_0' = (\sqrt{2}/2, \sqrt{2}/2)$ and $\beta_0' = (-\sqrt{2}/2, -\sqrt{2}/2)$ by Corollary 4 the partial likelihood estimation is scaled Fisher consistent. Hence, we can see that function $d(\alpha)$ attains minimum for $\alpha = \pi/4$ and $\alpha = 5/4\pi$.

Figure 1: Typical covariates and distance between true parameter and average estimates.



Example 2. *Up-to-scale estimation for various types of generalised frailty*

This example provides Monte Carlo simulations for different choices of generalised frailty $\Lambda(t,a)$. Five forms of generalised frailty for A distributed as shifted binomial distribution $binom(1,0.5)+1$ are considered (see Table 1). Observe that all functions $\Lambda(t,a)$ have the following properties: for $a \in \{1,2\}$ $\Lambda(0,a) = 0$ and $\Lambda(t,a) > 0$ for every positive t, $\lim_{t\to\infty} \Lambda(t,a) = \infty$ and they are continuously differentiable and strictly increasing on $t \in (0,\infty)$. The conditional distribution of the survival time T given $[X = x, A = a]$ was generated using the formula $\Lambda^{-1}(-\ln(U)\exp(-\beta_0' x), a)$, where U follows the uniform distribution on the interval $[0,1]$. The true parameter value was taken as $\beta_0' = (1, 0.5, 0.2)$

and the regressors $X = (X_1, X_2, X_3)'$ were used: either with standard normal or exponential distributions and $Cor(X_i, X_j) = 0.7$ for $i \neq j$. Estimations were repeated 5000 times for a sample size of 500. For each combination of the covariates and the generalised frailty two vectors are given as a result. The first one refers to scales - the means of ratios of estimates and the true parameters. Under normally distributed regressors the scaled Fisher consistency holds, so we expect the scales to be the same. The second vector in each cell consists of standard deviations of estimates. Simulations indicate good asymptotic behaviour of the estimators in the model with normal regressors as the differences in scales are very slight. Other choices of elliptically symmetric regressors, not presented in this example, lead to similar results. In the case of non-symmetric covariates the estimation brings worse results.

Table 1: Results of simulation experiment for different choices of generalised frailty. The first vector in each cell refers to the means of ratios of components of estimates and the true parameters. The second one refers to the standard deviations of the vector estimates of true parameter values.

Generalised frailty	Normal regressors	Nonormal regressors
$\Lambda(t,a) = a\sqrt{t}$	(0.9361, 0.9398, 0.9324) (0.0791, 0.0756, 0.0729)	(0.9260, 0.9357, 0.9434) (0.0865, 0.0739, 0.0728)
$\Lambda(t,a) = a\sqrt{t} + t$	(0.6228, 0.6228, 0.6122) (0.0770, 0.0700, 0.0676)	(0.2747, 0.3125, 0.3886) (0.0864, 0.0784, 0.0761)
$\Lambda(t,a) = t^2 + at - t$	(0.8855, 0.8817, 0.8839) (0.0781, 0.0726, 0.0704)	(0.6628, 0.6889, 0.7401) (0.0835, 0.0784, 0.0731)
$\Lambda(t,a) = t^{3a/2-1}$	(0.7588, 0.7589, 0.7572) (0.0775, 0.0733, 0.0708)	(0.4612, 0.5104, 0.5937) (0.0924, 0.0844, 0.0792)
$\Lambda(t,a) = t^{3a/2-1} + t^2$	(0.8560, 0.8520, 0.8498) (0.0804, 0.0726, 0.0695)	(0.4799, 0.5271, 0.6110) (0.0949, 0.0844, 0.0817)

Example 3. *The effect of variable's omission in the Cox model*

The above considerations show a wide range of distributional possibilities for the explanatory variables for which the estimation of regression parameters in the Cox model is scale Fisher consistent under the extended model with generalised frailty. As a particular case of $\Lambda(t,a)$ assume that $\Lambda(t,a) = a\Lambda(t)$. The frailty variable A has a special interpretation in survival analysis for the Cox model, where it is supposed to describe proportional changes of cumulated hazard $\Lambda(t)$ for individual units within the population. Since

$$P(T > t|x,a) = \exp\left(-\Lambda(t)a\exp(\beta_0'x)\right) = \exp\left(-\Lambda(t)\exp(\beta_0'x + \ln(a)\right)$$

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.

A Monte Carlo experiment was conducted to investigate properties of the partial likelihood estimation when variables are omitted and data are generated from the true Cox model. In order to demonstrate the effect of variable's omission the following Cox model was taken: $\beta_0' = (-1, -1, 0.5, 1)$, $\Lambda(t) = t^2$ and $X = (X_1, X_2, X_3, X_4)'$ where: (X_1, X_2) has the distribution described in Example 1, (X_3, X_4) has the two dimensional normal distribution with correlation of $1/\sqrt{2}$ and vectors (X_1, X_2) and (X_3, X_4) are independent.

The sample size of $n = 500$ was taken and the estimation was repeated 5000 times. Simulation results, given by the means of ratios of estimates and parameters, and by standard deviations of estimates, are presented in Table 2. The up-to-scale consistent estimation of the corresponding regression coefficients is revealed for covariates vector (X_1, X_2) and (X_3, X_4). For the estimation based on the entire set of regressors the estimation is consistent with the scale of one. For other regressor vectors it can be observed that departure from the elliptically symmetric distribution implies the lack of scaled consistency in estimation.

Table 2: Results of simulation experiment. The first and the second vector in each cell refers to mean scales and standard deviations of estimates of true parameter values corresponding to the subset of regressors.

Subsets of X	Scales \Sd	Subsets of X	Scales\Sd
(X_1, X_2)	(0.4299, 0.4228) (0.0774, 0.0816)	(X_1, X_2, X_3)	(0.6637, 0.6611, 1.9944) (0.0856, 0.0844, 0.0328)
(X_1, X_3)	(0.7613, 1.9430) (0.0895, 0.0330)	(X_1, X_2, X_4)	(0.9226, 0.9330, 1.1621) (0.0824, 0.0840, 0.0257)
(X_1, X_4)	(1.0396, 1.1299) (0.0841, 0.0247)	(X_1, X_3, X_4)	(1.1106, 0.9455, 0.9476) (0.0795, 0.0336, 0.0303)
(X_3, X_4)	(0.8820, 0.8854) (0.0345, 0.0301)	(X_1, X_2, X_3, X_4)	(1.0060, 1.0001, 1.0007, 1.0020) (0.0775, 0.0765, 0.0327, 0.0295)

Example 4. *Mayo Clinic Primary Biliary Cirrhosis Data*

The example is based on the data from the Mayo Clinic trial in PBC, available in the package `survival` of `R` program. In this example we consider four explanatory variables:

- age
- edema (0 for no edema, 0.5 for moderate and 1 for severe edema)
- bili (serum bilirunbin mg/dl)
- protime (standardised blood clotting time)

Before fitting the Cox model we logarithmically transformed variables bili and protime. The assumption of elliptical symmetry was checked by three tests implemented in the package `ellipticalsymmetry` in `R` program. We chose test MPQ by Manzottii et al. (2002) (Test 1), test by Schott (2002) (Test 2) and test by Huffer and Park (2007) (Test 3). The results are summarized in Table 3. Let $\hat{\beta}_0$ denote the estimate of the coefficients in the Cox model for

Table 3: Fitting the Cox model for PBC data with dropped regressors.

Subsets of regressors	P-values of tests for elliptical symmetry			Scales
	Test 1	Test 2	Test 3	
age, edema	<2.2e-16	<2.2e-16	<2.2e-16	(0.7585, 1.7932)
age, ln(bili)	0.1412	0.3712	0.0024	(1.0506, 1.1204)
age, ln(protime)	0.0078	0.0954	0.0008	(0.9026, 1.5432)
edema, ln(bili)	<2.2e-16	<2.2e-16	<2.2e-16	(1.1993, 0.9885)
edema, ln(protime)	<2.2e-16	0.0118	<2.2e-16	(1.6157, 1.2628)
ln(bili), ln(protime)	0.6853	0.0600	0.6400	(1.0273, 1.1492)
age, edema, ln(bili)	<2.2e-16	<2.2e-16	<2.2e-16	(1.0126, 1.1380, 1.0403)
age, edema, ln(protime)	<2.2e-16	0.0002	<2.2e-16	(0.7483, 1.5422, 1.2809)
edema, ln(bili), ln(protime)	<2.2e-16	0.0004	<2.2e-16	(1.0657, 0.9487, 0.9526)
age, ln(bili), ln(protime)	0.0015	0.0933	4.8e-05	(1.0335, 1.0550, 1.2253)

regressors: age, edema, ln(bili), ln(protime) and let $\hat{\beta}$ denote the estimate of the coefficients in the Cox model based on the subset of this regressors. For each subset of regressors we computed scales as ratios of the coefficients for corresponding variables in $\hat{\beta}_0$ and $\hat{\beta}$. All tests detect correctly the lack of elliptical symmetry when regressors include discrete variable edema. The lack of elliptical symmetry of explanatory variables may imply the scales not being the same after omitting regressors. For the regressors (age, ln(bili)) and (ln(bili), ln(protime)) the differences in the scales are slight. They seem to be elliptical symmetric (see p-values in Table 3).

6. Concluding discussion

An important property of the Cox model is that the baseline hazard is an unspecified function and makes the Cox model of a semiparametric type. A key reason for the popularity of the Cox model is that even though the baseline hazard is not specified, reasonably good estimates of regression coefficients, hazard ratios of interest, and adjusted survival curves can be obtained for a wide variety of data situations. Frailty models arise naturally from the Cox model with unobserved covariates, which form the frailty parameter and handle right-censoring and left-truncation, which is crucial in time-to-event analysis. Frailty gives way to explain additional time variability that could not be grasped by the original Cox model. The usual investigation of the partial likelihood estimator for the Cox regression model involved an interest in the consistency of the partial likelihood estimator under the Cox model with frailty, which presumes the time distribution dependent on a single baseline hazard λ, multiplied by a positive random variable A called frailty. Fisher consistency of the Cox estimator was studied under the independence of frailty A from the covariates X and under analytically convenient frailty distributions. Nevertheless, attributing to population individuals the same baseline up to a proportionality factor (frailty) and making consistent estimation dependent on purely analytic properties of frailty distribution seemed far from satisfactory. In fact studies show that the Fisher consistency does not hold under arbitrary frailty. What we could naturally hope for then, would be the so-called scaled Fisher consistency - regression parameters could be estimated consistently up to an unknown scaling factor. In the paper we demonstrate that this is attainable, the classical partial likelihood procedure leads to the estimator satisfying this condition up to a scaling factor under the extended Cox model with generalised frailty and an elliptically symmetric distribution of the covariates. The simulation studies indicate the lack of this property in the case of violating the assumption. The Cox model with generalised frailty is of great importance in various analyses of time-to-event data. However, it should be noted that the omission of an influential variable or misspecification of the frailty distribution may lead to severe estimation errors. In this light, considering estimation consistent up to scale may result in meaningful comparisons of impact of covariates on hazards.

References

Aalen, O. O., Borgan, O., Gjessing, H. K., (2008). *Survival and Event History Analysis. A Process Point of View*, Springer.

Bednarski, T., (1993). Robust estimation in Cox regression model. *Scandinavian Journal of Statistics*, 20, pp. 213–225.

Bednarski, T., Nowak, P. B., (2021). Scaled Fisher consistency of the partial likelihood estimator in the Cox model with arbitrary frailty. *Probability and Mathematical Statistics*, 41(1), pp. 77–87.

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

Bednarski, T., Skolimowska-Kulig, M., (2019). On scale Fisher consistency of maximum likelihood estimator for the exponential regression model under arbitrary frailty. *Statistics and Probability Letters*, 150, pp. 9–12.

Cox, D. R., (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society B*, 34, pp. 187–220.

Duchateau, L., Janssen, P., (2008). *The Frailty Model*, Springer-Verlag.

Henderson, R., Oman, P., (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society B*, 61, pp. 367–379.

Hougaard, P., (2000). *Analysis of Multivariate Survival Data*, Springer Verlag.

Huffer, F. W., Park, C., (2007). A test for elliptical symmetry. *Journal of Multivariate Analysis* 98 (2), 256–281.

Kalbfleisch, J. D., Prentice, R. L., (2002). *The Statistical Analysis of Failure Time Data*, Hoboken, N.J.: Wiley.

Klein, J. P., Moeschberger, M. L., (2003). *Survival Analysis. Techniques for Censored and Truncated Data*, New York: Springer-Verlag.

Manzotti, A., Pérez, F. J., Quiroz, A. J., (2002). A Statistic for Testing the Null Hypothesis of Elliptical Symmetry. *Journal of Multivariate Analysis*, 81 (2), 274–285.

Ruud, P., (1983). Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models. *Econometrica*, 51(1), pp. 225–228.

Schott, J. R., (2002). Testing for elliptical symmetry in covariance-matrix-based analyses. *Statistics & Probability Letters*, 60 (4), pp. 395–404.

Stoker, T., (1986). Consistent estimation of scaled coefficients. *Econometrica*, 54(6), pp. 1461–1481.

Therneau, T. M., Grambsch, P. M., (2000). *Modeling Survival Data: Extending the Cox Model*, New York: Springer-Verlag.

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

Wienke, A., (2010). *Frailty Models in Survival Analysis*, Chapman & Hall/CRC.