Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
5(338) 2018

DOI: http://dx.doi.org/10.18778/0208-6018.338.08

Tadeusz Bednarski
University of Wrocław, Faculty of Law, Administration and Economics, Institute of Economic Sciences, tadeusz.bednarski@uwr.edu.pl

Magdalena Skolimowska-Kulig
University of Wrocław, Faculty of Law, Administration and Economics, Institute of Economic Sciences, magdalena.skolimowska-kulig@uwr.edu.pl

Scaled Consistent Estimation of Regression Parameters in Frailty Models

Abstract: A computationally attractive method of estimation of parameters for a class of frailty regression models is discussed. The method uses maximum likelihood estimation for the classical exponential regression model. Scaled Fisher consistency is shown to hold and a simulation study indicating good asymptotic properties of the method, as well as real data case analysis, are presented.

Keywords: frailty models, maximum likelihood estimation, Fisher consistency

JEL: C13, C18

1. Introduction

Modeling with survival regression models 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 procedures, as they were for instance proposed in Bednarski (1993) and Sasieni (1993) for the Cox model, Cox (1972), were aimed at making estimation resistant to occasional outliers and they could not cope with oversimplified modeling. A common remedy to the estimation problem was to include a frailty variable to allow heterogeneity in longevity endowment. Vaupel, Manton and Stallard (1979) proposed to use a gamma distributed frailty to improve biased estimation for the life tables. There are numerous applications and extensions on using the gamma distributed frailty. Murphy (1994) shows, under very general conditions, consistency of the partial likelihood estimator for cumulative baseline and the variance of frailty. Aalen (1992) suggested a compound Poisson frailty model. Henderson and Oman (1999) 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, Borgan and Gjessing (2008) give a good review of inference for unobserved heterogeneity in survival modeling and their treatment for counting processes.

The usual approach to statistical inference with unobserved frailties assumes a parametric family of distributions for frailties and then uses maximum likelihood for the marginal densities. The approach taken here assumes no distributional structure on frailty but a natural parametric family for the cumulated hazard. It is shown that the maximum likelihood estimator for the regression parameters in the linear exponential model is scaled Fisher consistent for the extended model. The estimates then allow for the measuring of a relative risk between units or between strata equipped with different regressors.

There are general results on consistent scaled estimation (Ruud, 1983; Stocker, 1986) and some of them could possibly be adapted to survival analysis. Our objective here however, is to study a particular method which seems very simple in applications. A simulation study compares efficiency of the method with partial likelihood estimation.

2. The method

Two statistical models are defined below. The first one, the exponential regression model, will give the estimation method, while the second one, with a frailty variable, will presume a distributional structure of the data generating mechanism. Since the properties of the estimation method are studied via Fisher consistency condition, it will be necessary to state and consistently use densities for the two models. To make the presentation more adapted to event history studies we add the corresponding hazard functions.

In the sequel, the following notation for the considered models is used. If T is the time variable, $X = (X_1, \ldots, X_k)$ is a vector of explanatory variables and $\beta = (\beta_1, \ldots, \beta_k)$ is the corresponding vector of regression parameters, then

$$e^{\left[\log(\lambda_0),\beta\right]\left[1,x\right]'} e^{-te^{\left[\log(\lambda_0),\beta\right]\left[1,x\right]'}} g(x) \tag{1}$$

is the conditional density of the exponential model with intensity parameter λ_0 multiplied by $g(x)$, a density of covariates. Therefore, the hazard function in this model, given the covariates $X = x$, is constant and has the form

$$\lambda_0 \exp(\beta' x).$$

The second model, the one generating observations, is defined by the density

$$\lambda(t) e^{\gamma x'} z e^{-\ddot{\mathrm{E}}(t) z e^{\gamma x'}} g(x) f(z), \tag{2}$$

where z is the frailty variable, $\gamma = (\gamma_1, \ldots, \gamma_k)$ indicates the true parameter value while $f(z)$ and $g(x)$ are respectively frailty and covariates densities. The cumulated baseline hazard $\Lambda(t)$ is assumed a power function t^α for some $\alpha > 0$. This model is of course more general, compared to (1): it contains frailty and its cumulated baseline hazard depends on time. Its hazard function, given the covariates x and frailty z, has then the form

$$z\lambda(t)\exp(\gamma x').$$

Notice that in model (2), the observed regressors and the frailty variable are assumed independent.

Suppose now that a statistician observes a random sample of survival times and the corresponding covariates vectors from the second model:

$$T_1, (X_{11}, X_{12}, \ldots, X_{1k}), \ldots, T_n, (X_{n1}, X_{n2}, \ldots, X_{nk}).$$

He does not observe frailties, he neither knows their distribution nor the distribution of covariates. All he knows is the structure of the second model generating these data. He would not know how to estimate the true values of the regression parameters γ.

What we show here is that the maximum likelihood method for the first model (the exponential one), applied to the observed data, yields scale consistent estimates of the regression parameters γ. The consistency means here that, as the sample size increases, estimators converge to the true parameter vector, multiplied by an unknown constant. It is a standard approach in statistics to prove such asymptotic consistency by verifying so called Fisher consistency condition. In our case it can be stated as follows.

If instead of λ_0 we take $\log(\lambda_0)$ as parameter of the exponential model, then the Fisher consistency holds if the equation

$$E\left(\left[1, X\right] - \left[1, X\right] T e^{\left[\log(\lambda_0), \beta\right]\left[1, X\right]'}\right) = 0 \tag{3}$$

is satisfied for $\beta = c\gamma$ for some scaling factor $c > 0$. The expected value is taken with respect to the true distribution from the second model. Notice that the expression under the expectation (3) is equal to the gradient of the log density (1):

$$\left[\log\left(\lambda_0\right), \beta\right]\left[1, x\right]' - t e^{\left[\log(\lambda_0), \beta\right]\left[1, x\right]'} + \log(g\left(x\right)),$$

with respect to β and $\log(\lambda_0)$. The second derivatives of the log densities (1) with respect to the parameters yield hessian

$$-\left[1, X\right]'\left[1, X\right] T e^{\left[\log(\lambda_0), \beta\right]\left[1, X\right]'}.$$

Its expectation is negative definite, implying strict concavity of the objective function

$$E\left(\left[\log\left(\lambda_0\right), \beta\right]\left[1, X\right]' - T e^{\left[\log(\lambda_0), \beta\right]\left[1, X\right]'}\right)$$

under very mild formal assumptions and lead to a unique solution of (3). Therefore, if there exist parameters maximizing the objective function, they are uniquely given. Further on, it will be assumed that all the integrals involved in formal calculations exist and are finite.

If the expectation in (3) is taken with respect to empirical distribution function, we obtain the estimation method, which in this case is equivalent to the maximum likelihood estimation for the exponential model. Estimates are then solutions to equation

$$\sum_{i=1}^{n}\left(\left[1, X_i\right] - \left[1, X_i\right] T_i e^{\left[\log(\lambda_0), \beta\right]\left[1, X_i\right]'}\right) = 0$$

with respect to β and $\log(\lambda_0)$.

Before the main result is stated, it is convenient to split equation (3) into two parts: the one corresponding to parameter λ_0 and resulting in explicit formula

$$\lambda_0 = \frac{1}{E\left(Te^{\beta X'}\right)}$$

and the other one related to regression parameters

$$EX - E\left(XT\lambda_0 e^{\beta X'}\right) = 0$$

giving

$$EX - \frac{E\left(XTe^{\beta X'}\right)}{E\left(Te^{\beta X'}\right)} = 0.$$

The theorem stated below shows that the scaled Fisher consistency condition for the maximum likelihood method, resulting from the exponential regression model, is always satisfied under the extended model for the regression parameters γ. Moreover, explicit relation between the scaling constant and the exponent in the cumulated baseline hazard is given.

Theorem (Scaled Fisher consistency)

If $\Lambda(t) = t^\alpha$ then for $\beta = \frac{1}{\alpha}\gamma$

$$EX - \frac{E\left(XTe^{\beta X'}\right)}{E\left(Te^{\beta X'}\right)} = 0.$$

Proof. A crucial step in establishing the scaled Fisher consistency consists in computing the value of $E\left(XTe^{\beta X'}\right)$. We have

$$E\left(XTe^{\beta X'}\right) = \int\int\int_0^\infty xte^{\beta x'} z\lambda\left(t\right)e^{\gamma x'}e^{-\ddot{\mathrm{E}}\left(t\right)ze^{\gamma x'}}dtg\left(x\right)dxf\left(z\right)dz.$$

The change in variable t yields

$$\int\int\int_0^\infty x\Lambda^{-1}\left(t\right)e^{\beta x'}ze^{\gamma x'}e^{-tze^{\gamma x'}}dtg\left(x\right)dxf\left(z\right)dz,$$

where Λ^{-1} denotes the inverse of Λ. Therefore plugging t^α for Λ leads to

$$\int\int xe^{(\beta+\gamma)x'}z\int_0^\infty t^{1/\alpha}e^{-tze^{\gamma x'}}dtg\left(x\right)dxf\left(z\right)dz$$

and again, change in variable t results in integral

$$\int\int xe^{\left(\beta-\frac{1}{\alpha}\gamma\right)x'} z^{-\frac{1}{\alpha}} \int_0^\infty t^{1/\alpha} e^{-t} dt g(x) dx f(z) dz$$

$$= \int\int xe^{\left(\beta-\frac{1}{\alpha}\gamma\right)x'} z^{-\frac{1}{\alpha}} \Gamma(1/\alpha+1) g(x) dx f(z) dz.$$

Taking $\beta = \frac{1}{\alpha}\gamma$, by independence of X and Z, we obtain

$$\frac{E\left(XTe^{\beta X'}\right)}{E\left(Te^{\beta X'}\right)} = \frac{\int\int xz^{-\frac{1}{\alpha}} \Gamma(1/\alpha+1) g(x) dx f(z) dz}{\int\int z^{-\frac{1}{\alpha}} \Gamma(1/\alpha+1) g(x) dx f(z) dz} = \int xg(x) dx.$$

This completes the proof of the theorem.

3. Simulation results

A Monte Carlo experiment was conducted to compare results of maximum likelihood estimation for the exponential regression model with the Cox partial likelihood estimator. The reference to the Cox model is clear as it is frequently applied in time to event data analysis. The R programming language was used and a general purpose *optim* procedure from the "stats" package for the maximum likelihood estimation (mle) in the exponential regression model with censored data was applied. The partial likelihood estimation was done with *cohph* procedure from the "survival" package. Two cumulated baseline intensities were used, $\Lambda(t) = t^2$ and $\Lambda(t) = t^{1/2}$, for the generating of data distribution (2). The choice was motivated by comparison of estimation differences between convex and concave cumulated hazard Λ.

Table 1. Results of the simulation experiment with a continuous frailty variable

	$\Lambda(t) = t^2$	$\Lambda(t) = t^{1/2}$
Parameter value	(1, –0.5, 0.5)	(1, –0.5, 0.5)
Scaled parameter value	(0.8165, –0.4082, 0.4082)	(0.8165, –0.4082, 0.4082)
$n = 50$		
Scaled mle exponential	(0.7860, –0.3956, 0.3913)	(0.7868, –0.3909, 0.3913)
	(0.1134, 0.1728, 0.1727)	(0.1168, 0.1753, 0.1750)
Scaled partial likelihood	(0.7919, –0.3985, 0.3949)	(0.7872, –0.3913, 0.3911)
	(0.1006, 0.1538, 0.1561)	(0.1154, 0.1746, 0.1744)

	$\Lambda(t) = t^2$	$\Lambda(t) = t^{1/2}$
	n = 100	
Scaled mle exponential	(0.8030, –0.3998, 0.3988)	(0.8031, –0.3987, 0.4011)
	(0.0775, 0.1228, 0.1237)	(0.0776, 0.1207, 0.1211)
Scaled partial likelihood	(0.8065, –0.4012, 0.4014)	(0.8032, –0.3994, 0.4013)
	(0.0675, 0.10704, 0.1070)	(0.0767, 0.1191, 0.1194)
	n = 300	
Scaled mle exponential	(0.8129, –0.4049, 0.4046)	(0.8120, –0.4058, 0.4064)
	(0.0435, 0.0692, 0.06996)	(0.0429, 0.0675, 0.0669)
Scaled partial likelihood	(0.8141, –0.4057, 0.4056)	(0.8123, –0.4055, 0.4066)
	(0.0366, 0.0585, 0.0581)	(0.0416, 0.0661, 0.0652)

Source: own elaboration

Table 2. Results of the simulation experiment with a discrete frailty variable

	$\Lambda(t) = t^2$	$\Lambda(t) = t^{1/2}$
Parameter value	(1, –0.5, 0.5)	(1, –0.5, 0.5)
Scaled parameter value	(0.8165, –0.4082, 0.4082)	(0.8165, –0.4082, 0.4082)
	n = 50	
Scaled mle exponential	(0.7831, –0.3944, 0.3887)	(0.7830, –0.3933, 0.3939)
	(0.1202, 0.1804, 0.1821)	(0.1184, 0.1775, 0.1780)
Scaled partial likelihood	(0.7912, –0.3974, 0.3939)	(0.7838, –0.3907, 0.3941)
	(0.1039, 0.1575, 0.1593)	(0.1189, 0.1775, 0.1791)
	n = 100	
Scaled mle exponential	(0.8001, –0.4020, 0.3985)	(0.8013, –0.4002, 0.4018)
	(0.0805, 0.1267, 0.1296)	(0.0803, 0.1222, 0.1222)
Scaled partial likelihood	(0.8047, –0.4035, 0.4014)	(0.8012, –0.4001, 0.4023)
	(0.0687, 0.1090, 0.1093)	(0.0798, 0.1221, 0.1221)
	n = 300	
Scaled mle exponential	(0.8105, –0.4057, 0.4063)	(0.8107, –0.4072, 0.4071)
	(0.0481, 0.0743, 0.0746)	(0.0436, 0.0689, 0.0676)
Scaled partial likelihood	(0.8126, –0.4061, 0.4068)	(0.8112, –0.4068, 0.4067)
	(0.0399, 0.0620, 0.0617)	(0.0429, 0.0683, 0.0671)

Source: own elaboration

The explanatory variable X was three dimensional standard normal and the true parameter $\gamma = (1, -0.5, 0.5)$. Three dimensions seem to be a minimum for this sort of studies as this gives sufficient flexibility for the regression parameters choice. The frailty variable was either squared standard normal plus 1 (Table 1) or three times binomial plus 1 with success probability 0.5 (Table 2), thus covering continuous and discrete case. The time variable was censored by an independent standard exponential variable, yielding about 15% and 30% rates of censoring in models with cumulated intensities $\Lambda(t) = t^2$ and $\Lambda(t) = t^{1/2}$ respectively, the reasonable proportions in practical studies. Estimations were repeated 5000 times for moderate sample sizes of 50, 100 and 300. Results are summarized in the table above.

Every section of the tables shows the mean values of the estimates scaled to length 1 (first row) and the corresponding standard deviations below. Simulation results indicate rapid bias decrease as the sample size increases and slightly smaller variability of the partial likelihood estimates. Analogous simulations for samples without censoring show smaller variability of exponential maximum likelihood estimation for binomial frailty and $\Lambda(t) = t^{1/2}$.

Another example presented here compares the two estimation methods for the Veteran's Administration lung cancer data – a data set described in detail in Kalbfleisch and Prentice (1980), used frequently to test different estimation methods. Interest in this particular data comes from the fact that it originates from a controlled clinical trial. There are a number of continuous covariates: Karnofsky rating, disease duration and age. The binary variables are prior therapy (yes = 1/no = 0) and treatment (standard = 1/test = 0). There are four cell types (adeno, large, small, squamous) which were coded as binary variables (squamous, small and adeno versus large).

Table 3. Comparison of partial likelihood and maximum likelihood exponential estimation results for the Veteran's Administration lung cancer data

Regressors	ple	z	scaled	exp	z	scaled
Karnofsky	–0.0328	–5.958	–0.0314	–0.0307	–6.044	–0.0316
Disease duration	0.0001	0.009	0.0000	0.0002	0.031	0.0003
Age	–0.0087	–0.936	–0.0083	–0.0063	–0.695	–0.0065
Prior therapy	0.0072	0.308	0.0068	0.0051	0.224	0.0052
Squamous	–0.4013	–1.420	–0.3839	–0.3672	–1.347	–0.3788
Small	0.4603	1.729	0.4403	0.4495	1.719	0.4637
Adeno	0.7948	2.624	0.7604	0.7438	2.527	0.7673
Treatment	0.2946	1.419	0.2818	0.2204	1.110	0.2274

Source: own elaboration

Minder and Bednarski (1996) have shown that a robust modification of the partial likelihood estimator leads to essentially different clinical conclusions there. Table 3 shows our estimation results. As in the simulated examples, the differences in estimates are relatively small. Moreover, the scaled and non-scaled values are quite close. There is no difference in significance for the explanatory regression variables. It is not however possible to determine, on the basis of this analysis, whether there is a non-constant frailty for the Veteran's Administration patients.

4. Conclusions

It is shown that the maximum likelihood method for the exponential regression model satisfies scaled Fisher consistency property for a large class of frailty models given by power cumulated baseline hazard function and arbitrary frailty distribution. The efficiency of the method is comparable in bias and variability of estimates to partial likelihood method for the Cox regression model. Estimation for the Veteran Administration lung cancer data shows also high consistency with Cox estimator. Extensions of the presented method to larger families of statistical models are under study.

References

Aalen O.O. (1992), *Modelling heterogeneity in survival analysis by the compound Poisson distribution*, "Annals of Applied Probability", vol. 2, no. 4, pp. 951–972.

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

Bednarski T. (1993), *Robust estimation in Cox regression model*, "Scandinavian Journal of Statistics", vol. 20, no. 3, pp. 213–225.

Cox D.R. (1972), *Regression models and life-tables (with discussion)*, "Journal of the Royal Statistical Society B", vol. 34, no. 2, pp. 187–220.

Henderson R., Oman P. (1999), *Effect of frailty on marginal regression estimates in survival analysis*, "Journal of the Royal Statistical Society B", vol. 61, no. 2, pp. 367–379.

Kalbfleisch J.D., Prentice R.L. (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Minder C.E., Bednarski T. (1996), *A robust method for proportional hazard regression*, "Statistics in Medicine", vol. 15, pp. 1033–1047.

Murphy S.A. (1994), *Consistency in a proportional hazard model incorporating a random effect*, "The Annals of Statistics", vol. 22, no. 2, pp. 712–734.

Ruud P. (1983), *Sufficient conditions for the consistency of maximum likelihood estimation despite misspecification of distribution in multinomial discrete choice models*, "Econometrica", vol. 51, no. 1, pp. 225–228.

Sasieni P.D. (1993), *Maximum weighted partial likelihood estimates for the Cox model*, "Journal of the American Statistical Association", vol. 88, pp. 144–152.

Stoker T. (1986), *Consistent estimation of scaled coefficients*, "Econometrica", vol. 54, no. 6, pp. 1461–1481.

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

Zgodna z dokładnością do skali estymacja parametrów w modelach regresji ze zmienną „frailty"

Streszczenie: W artykule omówiono atrakcyjną obliczeniowo metodę estymacji parametrów dla klasy modeli regresyjnych z nieobserwowaną zmienną „frailty". Dowiedziono, że estymator największej wiarygodności stosowany w klasycznym wykładniczym modelu regresji jest Fisherowsko zgodny z dokładnością do skali w rozważanym modelu „frailty". Przeprowadzone badania symulacyjne oraz analiza rzeczywistych danych wskazują na dobre własności asymptotyczne prezentowanej metody estymacji.

Słowa kluczowe: modele frailty, estymacja największej wiarygodności, Fisherowska zgodność

JEL: C13, C18

Received: 2018-02-18; verified: 2018-04-12. Accepted: 2018-08-07