

<https://doi.org/10.19195/2658-1310.27.2.5>

Magdalena Skolimowska-Kulig

ORCID: 0000-0002-4748-7624

Instytut Nauk Ekonomicznych, Uniwersytet Wrocławski

magdalena.skolimowska-kulig@uwr.edu.pl

Fisher consistent estimation of regression parameter in the proportional mean residual life model with frailty

Date of submission: 21.02.2020; date of acceptance: 28.08.2021

JEL classification: C13, C21, C51

Keywords: proportional mean residual life regression model, frailty models, Fisher consistent estimation

Abstract

Fisher consistent estimation of regression parameter in the proportional mean residual life model with frailty

In the article, we consider the Fisher consistent estimation of the regression parameters in the proportional mean residual life model with arbitrary frailty. It is discussed that conventional estimation procedures, such as the maximum likelihood estimation or Cox's approach, which are employed in common regression models, may also yield consistent inference in the extended models.

1. Introduction

The statistical analysis of lifetime data (or more general, time-to-event, event–history, or duration data) plays an important role in many fields of demography, biomedical studies, economics, actuarial science, reliability research and others. In most practical applications, the population under study consists of individuals who may differ in gender, socio-economic status, received treatment, and so forth. The regression framework allows to include information about these observed variables (covariates) in models of survival analysis. The most popularly used survival regression model is the Cox proportional hazards model (Cox, 1972), which investi-

gates the relationship of covariates and the time-to-event through the hazard function. It assumes that the covariates have a multiplicative effect on the hazard and that this effect is constant over time.

Naturally, the concept of regression allows for the entering of explanatory variables in order to describe their influence on the survival time. However, it is almost always impossible to include all influential factors, mainly because some of them cannot be measured or are ignored. In some cases, the omission of relevant explanatory variables may cause inferential perturbations that are out of researcher's control. A common remedy to this problem is to extend the model by including an unobserved random effect, a so-called frailty. The notion of frailty was introduced and applied to the population data by Vaupel, Manton, Stallard (1979). The classical and mostly frequently applied frailty model assumes a proportional hazards model that is conditional on a frailty variable. The hazard of an individual depends on an unobservable, age-independent random variable, which acts multiplicatively on the baseline hazard function. The usual approach to statistical inference with unobserved frailties assumes a parametric family of distributions for frailties, usually gamma, inverse gaussian or the power variance function family (see Wienke, 2011 for comprehensive review of frailty modelling in survival data analysis). 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; Hougaard, 2000).

Undoubtedly, modelling the frailty distribution is a remedy for inconsistent estimation of regression parameters but its limited choice relies more on analytical convenience than on experimental knowledge. Bednarski and Skolimowska-Kulig (2018, 2019) and Bednarski and Nowak (2021) took different approach. They considered the basic condition needed in sound statistical inference about parameters: the Fisher consistency of estimators, under no assumption about frailty distributions. In particular, they showed that the common procedures for the estimation of regression parameters in certain hazard-based survival models yield consistent up-to-scale estimators for the extensions of these models.

In this article, a similar attempt to the problem of the consistent estimation in oversimplified models is made for the regression model, based on the remaining lifetime. Recently, modelling the relation between the duration time and covariates through the remaining life has gained increasing attention from clinical researchers, but it also has been employed in some analyses of human populations and actuarial work, as well as in industrial reliability studies. However, the frailty variable is usually neglected in these models, mainly because including the random effects is challenging in estimation. Therefore, an examination of the accuracy of the employed estimation methods under the influence of an unknown factor seems to be of great importance. We reveal that the maximum likelihood estimator for the regression parameters in the linear exponential model is Fisher consistent

in the extended proportional mean residual life frailty model, while the estimator based on Cox's partial likelihood is Fisher consistent up to scale under an assumption on the distributional structure of covariates.

2. The proportional mean residual life regression model

A quantity of vital importance in understanding the survival, aging, or duration process, besides the survival function and the hazard (risk) rate, is the mean residual life function. For the nonnegative random survival time (the time to event) T with a survival function S and a finite mean, the mean residual life function at time $t \geq 0$ is defined as the expected remaining life of a subject given survival up to t :

$$m(t) = E(T - t | T > t).$$

It is worth underlining that this function completely determines the distribution with the finite mean.

The mean residual life function has a tremendous range of applications. In biomedical sciences, researchers use it to analyse survivorship, in economics it is applied for investigating landholding and optimal disposal of an asset. Actuaries apply the function to set rates and benefits for life insurances. In industrial reliability, it is a helpful tool in the planning of strategies for maintenance and replacement. For an extensive coverage of possible applications of the mean residual life, see Guess and Proschan (1988). In many ways, the mean residual life function provides a more natural basis for the modelling of the duration data than the hazard function. The former summarizes the entire residual life distribution, whereas the latter relates only to the risk of immediate failure.

Many early literatures on survival models based on the mean residual life function studied their stochastic behaviours and statistical inference under an assumption of homogeneous population. Oakes and Dasu (1990) proposed a semiparametric proportional mean residual life model for the analysis of reliability and survival data. Their model implies that the mean residual life functions for the two samples are in a constant ratio. To quantify and summarize the association between the mean residual function and covariates, Maguluri and Zhang (1994) extended this model to a regression context. The model specifies the conditional mean residual life function through the following relation

$$m(t | X = x) = m_0(t) \exp(\beta' x), \quad (1)$$

where X is the k -dimensional covariate vector (the vector of explanatory variables), β is the k -vector of regression parameters and $m_0(t)$ is the mean residual life function corresponding to unknown baseline distribution.

In general, there is no straightforward relationship between the proportional mean residual life model and the widely used Cox proportional hazards model (Cox, 1972) in which:

$$\lambda(t|X=x) = \lambda_0(t)\exp(\theta' x), \quad (2)$$

where, conditioned on X , $\lambda(t|X=x)$ is the hazard function of the time-to-event, $\lambda_0(t)$ is the baseline hazard and θ is the regression parameter. Oakes and Dasu (1990) showed that, when a model satisfies both the proportional mean residual life and the proportional hazards assumptions, then its underlying distribution belongs to the Hall-Winter class of distributions with linear residual life functions (Hall & Wellner, 1981). The special case of this situation appears for the exponential distribution, then $\theta = -\beta$. Furthermore, as pointed out by Maguluri and Zhang (1994), the mean residual life function equals the reciprocal of the hazard function of the recurrence times in the stationary renewal process with underlying survivor function of T , and that leads to the representation of (1) by the proportional hazards model of recurrence times of the form given by (2) with the regression parameter $\theta = -\beta$.

As an attractive alternative to the conventional hazard-based models, the mean residual life regression model provides a direct way to interpret the relationship between the expectation of the remaining lifetime and explanatory factors. The model is used mainly in biomedical studies in assessing life expectancy and measuring the effect of prognostic variables or a treatment. The sign of a regression parameter reflects whether the effect is positive, while its absolute value measures the magnitude of the effect. Maguluri and Zhang (1994) proposed estimation procedures for regression coefficients for uncensored survival data, which was later modified to accommodate censoring setting by Chen and Cheng (2005), who also developed semiparametric inference procedures to estimate the baseline mean residual life function. Huang, Xiang, Ha (2019) proposed a frailty model framework based on mean residual life regression to accommodate intracluster correlation, and a novel hierarchical quasi-likelihood method for estimation regression parameters.

3. The method

The notion of the Fisher consistency is crucial for many statistical formal bases. In its simplest form, it says that the expected value of model's score function (the derivative of the log-likelihood with respect to parameters) is zero at the true parameter value. The empirical version of this equation leads to such important properties of estimators as their asymptotic consistency and normality. In a more advanced framework, in which the estimation procedure is defined by a functional of a cumulative distribution function, the Fisher consistency means that the value of the functional at model's distribution is equal to the parameter of this distribution. The applicability of the notion is, of course, conditioned on the model's validity.

We consider two statistical models. The first one, the simple mean residual life regression model, gives the estimation methods. The second one is the extension of the proportional mean residual life model (1), which includes the frailty variable:

$$m(t|X=x, Z=z) = zm_0(t) e^{\beta'x}, \quad (3)$$

where z is an unobservable positive random factor (frailty) independent from the — dimensional covariate vector X , $m_0(t)$ is the unknown baseline mean residual life function and β is the k -vector of regression parameters. The above model describes the data generating mechanism, i.e., explains the relation between the duration time and explanatory variables.

We examine the Fisher's consistent estimation of the regression parameters in (3) in a case when the assumed model is oversimplified, particularly the random factor is neglected. We consider two estimating procedures proposed by Maguluri and Zhang (1994) for the usual proportional mean residual model (1) and uncensored data. The first one is derived from the maximum likelihood equation of the simple exponential regression model, while the second one is based on Cox's approach (Cox, 1972) for the underlying proportional hazards structure of the model.

4. Results

4.1. Maximum likelihood estimator

The basic approach presumes the following simple model and employs the maximum likelihood method to estimate the regression parameter β :

$$m(t|X=x) = m_0 e^{\beta'x}, \quad (4)$$

where the baseline mean residual life function m_0 is constant. Accordingly, the baseline distribution is exponential with the mean m_0 and the model is equivalent to the well-known exponential regression one with the regression parameter — β .

The condition of Fisher consistency of the regression parameters estimators takes the form:

$$E\left[X - \frac{E(XT\exp(-\beta'X))}{E(T\exp(-\beta'X))}\right] = 0, \quad (5)$$

where the expectations are taken with respect to the joint distribution of the variables. Evidently, the above condition holds if β is the true value of the parameter under the model given by (4). Furthermore, the condition also holds when the expectations in (5) are taken according to the joint distribution of T , X , and Z following the extended model (3). Consequently, the maximum likelihood estimator obtained under the assumption of the simple model is Fisher consistent for estimating the regression parameter in the proportional mean residual life model with frailty, where the baseline mean residual life function depending on time and the frailty variable are unknown.

The solution of the empirical version of (5):

$$\int \left[x - \frac{\int xt \exp(-\beta'x) dF_n(t, x)}{\int t \exp(-\beta'x) dF_n(t, x)} \right] dF_n(t, x) = 0$$

gives the considered maximum likelihood estimator, where is the empirical distribution function of the sample (T_i, X_i) , $i = 1, 2, \dots, n$.

4.2. Estimation based on Cox's approach

Cox proposed estimating the regression parameter in the proportional hazards model (2) by the solution of the equation:

$$\int \left[y - \frac{\int x \mathbb{I}\{v \geq w\} \exp(\theta'x) dF_n(v, x)}{\int \mathbb{I}\{v \geq w\} \exp(\theta'x) dF_n(v, x)} \right] dF_n(w, y) = 0 \quad (6)$$

where F_n denotes the empirical distribution function of a random sample. If in the above equation F_n is substituted by a joint distribution of the random variables generating the sample which follow the proportional hazards model with the true regression parameter θ_0 , then $\theta = \theta_0$ is the only solution of this equation (see Bednarski, 1993) and the Fisher consistency holds. Maguluri and Zhang (1994) used Cox's approach to estimate in the proportional mean residual life model (1). The procedure they proposed is based on the relation between the mean residual life function and the hazard rate of the forward recurrence time in a (hypothetical) stationary renewal process formed by renewals following the same distribution. More precisely, the proportional mean residual life model (1) implies Cox proportional hazards model for the forward recurrence time with the regression parameter $\theta = -\beta$. This means that the procedure leads to the Fisher consistent estimation of under the proportional mean residual life model.

Suppose it is known that the estimation method for the regression parameter in the mean residual life model described above is taken under the proportional mean residual life model with frailty (3). This means that the distributional structure of the data-generating mechanism is more complex than presumed. Bednarski and Nowak (2021) discussed the problem of inference based on the Cox regression model for various extensions of the model. In particular, they demonstrated that the partial likelihood estimator is Fisher consistent up to a scaling factor under a symmetry type distributional assumptions on regressors in the Cox model with frailty. The mean residual life regression model with frailty implies the Cox model with (reciprocal) frailty and the opposite regression parameter for the recurrence time. The partial likelihood method for the regression parameter in the oversimplified model applied to the observed data yields the Fisher scaled consistency of the esti-

mator in the mean residual life model with frailty if the vector of explanatory variables is spherically symmetric distributed. The scaled Fisher consistency means that the equation (6) with F_n substituted by the true distribution of (T, X', Z) under the model (3) holds $\theta = c\beta_0$, for where $c > 0$ is some scaling factor.

The estimation procedure of regression parameters in the common proportional mean residual life allows for straightforward estimation of the ratio of expected residual lives for any given two values of explanatory vectors. The Cox method used in the extended model leads to the scaled Fisher consistency, thus that ratio is only known up to the power c . However, even though the constant is unknown, still some comparison of expected residual lives remains possible.

The partial score equation (6) with F_n being the joint empirical distribution of the recurrence time and the vector of regressors could be used to derive the estimator of the regression parameter if the recurrence time was observable. However, it is only a theoretical possibility. To use the actual sample, Maguluri and Zhang (1994) noticed that the considered partial score equation may be rewritten in the following form:

$$\int \left[y - \frac{\int x(v-t)\mathbb{I}\{v \geq t\} \exp(-2\beta'x) dF_n(v,x)}{\int (v-t)\mathbb{I}\{v \geq t\} \exp(-2\beta'x) dF_n(v,x)} \frac{\int \mathbb{I}\{v \geq t\} \exp(-\beta'x) dF_n(v,x)}{\int v \exp(-\beta'x) dF_n(v,x)} \right] dt = 0,$$

where F_n denotes the empirical distribution of the sample (T_i, X'_i) , $i = 1, 2, \dots, n$, and the solution of the above equation gives the considered estimator.

5. Conclusions

The article discusses the problem of Fisher consistent estimation of regression parameters in the proportional mean residual life model with unknown frailty. The Fisher consistency is the minimum required for the reasonable statistical estimation and a primary step in establishing the asymptotic distribution of estimators. It is shown that despite the misspecification of the model — oversimplifying the baseline mean residual life function or omission of the unobservable explanatory variable, this inference remains possible. The standard maximum likelihood estimator derived under the simple exponential model assumption meets the aforementioned condition if the true model is more complex — the baseline distribution and the frailty variable are unknown. The second considered statistical approach — Cox's partial likelihood estimation, leads to the scaled Fisher consistency under a symmetry type distributional assumption on covariates. Thus, the estimation is consistent only up to some unknown scaling factor. However, it still allows for meaningful inference.

References

- Bednarski, T. (1993). Robust estimation in Cox regression model. *Scandinavian Journal of Statistics*, 20, 213–225.
- Bednarski, T., Nowak, P. (2021). Scaled Fisher consistency of partial likelihood estimator in the Cox model with arbitrary frailty. *Probability and Mathematical Statistics*, 41 (1), 77–87.
- Bednarski, T., Skolimowska-Kulig, M. (2018). Scaled consistent estimation of regression parameters in frailty models. *Acta Universitatis Lodzianis. Folia Oeconomica*, 5 (338), 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, 9–12.
- Chen, Y., Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data. *Biometrika*, 92 (1), 19–29.
- Cox, D. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 34, 187–220.
- Duchateau, L., Janssen, P. (2008). *The Frailty Model*. New York: Springer Verlag.
- Guess, F.P.F. (1988). Mean residual life: Theory and applications. In: P.R. Krishnaiah, C.R. Rao (eds.), *Handbook of Statistics*. 215–224. Amsterdam: Elsevier Science.
- Hall, W., Wellner, J. (1981). *Mean residual life*. In M. Csörgö (ed.), *Statistics and Related Topics: Proceedings of the international symposium held in Ottawa, Canada, May 5–7, 1980*. 169–184. Amsterdam: North-Holland.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer Verlag.
- Huang, R., Xiang, L., Ha, I. (2019). Frailty proportional mean residual life regression for clustered survival data: A hierarchical quasi-likelihood method. *Statistics in Medicine*, 38, 4854–4870.
- Maguluri, G., Zhang, C.H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 56 (3), 477–489.
- Oakes, D., Dasu, T. (1990). A note on residual life. *Biometrika*, 77 (2), 409–410.
- Vaupel, J., Manton, K., Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16, 439–454.
- Wienke, A. (2011). *Frailty Models in Survival Analysis*. Boca Raton, FL: CRC Press.