Goodness-of-fit tests for conditional models with time-dependent coefficients under censoring and truncation

Bianca Teodorescu* & Ingrid Van Keilegom

Institut de Statistique, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgique

In the context of semiparametric regression with censored data, a lot of models are available to account for the relationship between the survival time and a set of explanatory variables: Cox proportional hazards model, log-logistic model, etc. In a general way, these models can be written as linear regression models, where the dependent variable is a monotone transformation (ϕ) of the survival function (S). Different choices of this transformation give the models mentioned above.

In these models, the regression coefficients are often supposed to be constants. But in practice, the structure of the data might be more complex, in the sense that it is better to consider coefficients that can vary over time and also data that can be left-truncated. To be more precise, the following model is considered:

$$\phi(S(z|\boldsymbol{X})) = \boldsymbol{\beta}(z)\boldsymbol{X}, \qquad (1)$$

where ϕ is known, the survival-times are right-censored and/or left-truncated and the covariates $\boldsymbol{X}$ may be discrete and/or continuous.

A general method is applied to derive goodness-of-fit tests for this kind of models. For a given ϕ, $H_0 : \exists\boldsymbol{\beta}(z)$ such that (1) holds is tested against H_A : (1) does not hold for any $\boldsymbol{\beta}(z)$. The main idea is to compare a semiparametric estimator of the response, $\phi(\hat{S}(z|\boldsymbol{X}))$, with its parametric counterpart, $\hat{\boldsymbol{\beta}}(z)\boldsymbol{X}$ (same technique as in [1]). Where $\hat{S}$ is a conditional Kaplan-Meier type estimator with truncated data for S, see [2], and $\hat{\boldsymbol{\beta}}(z)$ is the least-squares estimator proposed in [3]. A large deviation between them indicates the lack of fit of the parametric form and thus the rejection of the null. The asymptotic distribution of the test statistic is established and a bootstrap procedure is proposed in order to get the critical values of the test. Some simulations and an application to real data are also provided.

Key words: semiparametric regression, generalized conditional linear models, time-dependent coefficients, least-squares estimator, U-processes.

References

[1] Cao, R. and González-Manteiga, W. (2008). Goodness-of-fit tests for conditional models under censoring and truncation. Journal of Econometrics, 143, 166-190.

[2] Iglesias-Pérez, C. and González-Manteiga, W. (1999) - Strong Representation of a Generalized Product-Limit Estimator for Truncated and Censored Data with Some Applications, *Nonparametric Statistics*, 10, 213-244

[3] Teodorescu, B., Van Keilegom, I. and Cao, R. (2008) - Semiparametric regression in survival analysis, (to appear in *The Annals of the Institute of Statistical Mathematics*)

[4] Teodorescu, B. and Van Keilegom, I. (2008) - Goodness-of-fit tests for conditional models in survival analysis, (under preparation)

[5] Teodorescu, B. and Van Keilegom, I. (2008) - Goodness-of-fit tests in survival analysis - a case study, (under preparation)