Multi-Parameter Regression Survival Models

Kevin Burke[1] and Gilbert MacKenzie[1,2]

[1] Centre for Biostatistics, University of Limerick, Ireland.
[2] CREST, ENSAI, Rennes, France.

Email: kevin.burke@ul.ie gilbert.mackenzie@ul.ie

Abstract: It is well known that the *proportional hazards* (PH) assumption is a simplifying assumption in survival analysis that may not always be appropriate. However, PH models are routinely fitted and inference is made on the data based on such models. A major flaw here is that if the data are non-PH then we will reach incorrect conclusions by making this assumption. For example we may find a covariate to be statistically insignificant when in fact it is important, but the model fails to pick this up. Even if a PH model *does* pick up the statistical significance of a non-PH covariate, the nature of the effect of the covariate on survival, as determined by this simplistic model, will clearly be incorrect. We introduce a regression-based extension of PH modelling to try an account for situations such as those described above and offer new, previously unavailable insights, into the data.

Keywords: Crossing hazards, Multi-parameter regression survival models, PH and non-PH models

1 Introduction

Generally, when modelling data the model has multiple parameters. Typically, we choose only to regress one of these parameters to measure the effect of the covariates. For example, in GLMs (McCullagh and Nelder, 1989) we regress the location parameter, $g(\mu) = X\beta$, whilst the dispersion parameter, σ, is often treated as being little more than a nuisance parameter. This view of the role of dispersion is disputed by Pan & MacKenzie (2003, 2006 & 2007) working in the longitudinal data modelling setting. Accordingly, in recent times there has been more work done in regressing dispersion parameters simultaneously with location parameters, however it is still not common practice among the statistical community.

A similar situation arises in survival analysis. One of the most popular models in this area is the *proportional hazards*, PH, model. Often this model is imposed on data even though it is known that, in reality, the data do not obey the PH assumption. Clearly the more *non*-PH data are, the less appropriate the model will be. The PH model is equivalent to regression of the *scale* parameter, say λ, in a model that has the proportional hazards property. Our new proposal is to simultaneously regress the *shape* parameter, say γ. This innovation thus generalizes the PH model to non-PH status and affords much more flexibility. The influence of covariates on the hazard ratio, which is constant in a PH model, can now change with time. The effect that the covariates have on the shape parameter may be of scientific interest in its own right. Here we will focus specifically on the Weibull model, although the methodology can easily be applied to other models.

2 Multi-Parameter Regression Survival Model

In a *parametric* PH model, we regress the scale parameter only. We propose the following multi-parameter regression:

$$g_1(\lambda) = x_1^T \beta, \qquad g_2(\gamma) = x_2^T \alpha, \tag{1}$$

where, g_1 and g_2 are appropriate link functions and we have also subscripted the corresponding covariate vectors to highlight the fact that the covariates regressing the scale and shape do not have to be the same.

2.1 MPR Weibull

The form of the Weibull distribution we will use is that presented in Collett (2003) which has hazard function $\lambda(t) = \lambda\gamma t^{\gamma-1}$ where λ,$\gamma > 0$. The hazard is decreasing for $\gamma < 1$, constant for $\gamma = 0$ and increasing for $\gamma > 1$.
The hazard function for the multi-parameter regression Weibull is

$$\lambda(t) = \exp(x_1^T \beta) \exp(x_2^T \alpha) t^{\exp(x_2^T \alpha)-1}. \tag{2}$$

We have used the log-link for both λ and γ here so that $\lambda, \gamma \in \mathrm{R}_+$. This generalization of the Weibull leads to a time-dependent hazard ratio for variables that are significant in the shape regression.
Moreover it is clear that the addition of a regression for the shape parameter allows the multi-parameter regression Weibull model to deal with crossing hazards data in a natural way, without recourse to the use of the frailty arguments adopted by MacKenzie & Ha (2007). Accordingly, use of this class of models, will render the analysis of crossing hazards data relatively routine.

2.2 Examples

In the presentation we analyse two data sets: a lung cancer data set collected in Northern Ireland between October 1991 and September 1992 (Wilkinson, 1995), in which we find some non-PH covariates, and the well-known gastric cancer data set of the Gastrointestinal Tumor Study Group (1982), where we observe the situation of crossing hazards. In both data sets we will illustrate the flexibility of the MPR Weibull model compared with the PH Weibull model.

3 Discussion

It has been found that the multi-parameter regression Weibull model indeed affords great flexibility and leads to better fits when compared with the standard proportional hazards model. This can be verified both graphically or more formally using likelihood ratio tests or AIC values. The extra generality leads of course to additional model selection issues and we describe their solution in the workshop presentation.

Acknowledgments: This work was supported, in part, by the SFI's (www.sfi.ie) BIO-SI research programme, grant number, **07MI012**. The first author is an IRCSET Scholar (www.ircset.ie) and the second is the Principal Investigator of BIO-SI (www.ul.ie/bio-si).

References

Collett, D. (2003) *Modelling Survival Data in Medical Research*, 2nd ed., Chapman & Hall/CRC.

MacKenzie G, & Ha ID. (2007) Modelling survival data with crossing hazards. *22nd IWSM Proceedings*, Barcelona, Spain, 416-420.

McCullagh, P. & Nelder, J. (1989) *Generalized Linear Models*, Chapman & Hall/CRC.

Pan J & MacKenzie G. (2003) On model selection for joint mean-covariance structures in longitudinal studies. *Biometrika*, **90**, 1, 239-244.

Wilkinson, Pauline (1995) *Lung Cancer in Northern Ireland 1991 - 1992*, PhD thesis, Queen's University Belfast.