The Stata Journal (2004)
4, Number 1, pp. 98–101

Flexible parametric alternatives to the Cox model: update

Patrick Royston
UK Medical Research Council

Abstract. Royston (2001) and Royston and Parmar (2002) introduced flexible parametric models for survival analysis, implemented in Stata through the ado-file `stpm` (Royston 2001). In the present article, `stpm` is updated to Stata 8.1 and has been shown to work correctly with Stata 8.2. To increase the reliability of the estimation procedure, the basis functions of the splines used to approximate the baseline distribution function have been orthogonalized.

Keywords: st0001_2, parametric survival analysis, proportional hazards, proportional odds, regression splines, orthogonal basis functions

1 Introduction

Royston (2001) and Royston and Parmar (2002) introduced flexible parametric models for survival analysis. These were implemented in Stata through the ado-file `stpm` (Royston 2001). Stated briefly, `stpm` extends the Cox proportional hazards (PH) model in three ways: by modeling the baseline distribution parametrically using a spline function of time; by providing, in addition to PH models, proportional odds and probit survival models; and by allowing regression coefficients for covariates to vary with time (`stratify()` option). This note reports some improvements to the software.

2 The models

To recapitulate, these models involve transformation of the survival function by a link function $g\left(.\right)$.

$$g\left\{S\left(t;\mathbf{z}\right)\right\} = g\left\{S_0\left(t\right)\right\} + \beta^{\mathrm{T}}\mathbf{z}$$

where $S_0\left(t\right) = S\left(t;\mathbf{0}\right)$ is the baseline survival function and β is a vector of parameters to be estimated for covariates $\mathbf{z}$. For the PH model, the link function $g\left\{S\left(t;\mathbf{z}\right)\right\}$ is chosen to be $\ln\left\{-\ln S\left(t;\mathbf{z}\right)\right\}$. Since according to a standard identity $-\ln S\left(t;\mathbf{z}\right)$ equals the cumulative hazard function, $H\left(t;\mathbf{z}\right)$, the PH model becomes

$$\begin{aligned}\ln H\left(t;\mathbf{z}\right) &= \ln H\left(t;\mathbf{0}\right) + \beta^{\mathrm{T}}\mathbf{z}\\ &= \ln H_0\left(t\right) + \beta^{\mathrm{T}}\mathbf{z}\\ &= s\left(t\right) + \beta^{\mathrm{T}}\mathbf{z}\end{aligned}$$

The function $s\left(t\right)$ is unspecified. We chose to represent it by a restricted cubic spline in log time:

$$s\left(t\right) = \gamma_0 + \gamma_1 \ln t + \gamma_2 v_1\left(\ln t\right) + \cdots + \gamma_{m+1} v_m\left(\ln t\right)$$

The integer m is the number of internal *knots* of the spline function, and $\ln t$, $v_1(\ln t)$, $\ldots$, $v_m(\ln t)$ are known as *basis functions* of the spline. Mathematical details are given by Royston and Parmar (2002).

3 Spline basis functions: problem and solution

The basis functions are highly correlated, which can sometimes cause `stpm` difficulties in estimating the parameters of the model quickly and reliably. For example, suppose that we have 100 observations of $\ln t$ equally spaced on the interval $[\,0.01, 1\,]$. We place $m = 3$ knots at 0.25, 0.5, and 0.75 and use the auxiliary command `frac_spl` to compute the untransformed spline basis functions. These turn out to have the following correlation matrix:

	$\ln t$	$v_1(\ln t)$	$v_2(\ln t)$	$v_3(\ln t)$
$\ln t$	1.000			
$v_1(\ln t)$	−0.922	1.000		
$v_2(\ln t)$	−0.942	0.997	1.000	
$v_3(\ln t)$	−0.966	0.987	0.996	1.000

A simple solution to this problem is to transform the basis functions linearly so that, after transformation, the correlations are zero. Gram–Schmidt orthogonalization is one way to do this, and it is available through the Stata command `orthog`; see [R] **orthog** in the Stata 8 manual for details. Orthogonalization is now built into `stpm`, although the option `noorthog` is provided for compatibility with earlier versions and for pedagogic reasons. In practice, it is not essential to transform $\ln t$, but only $v_1(\ln t), \ldots, v_m(\ln t)$.

4 How does orthogonalization affect the model?

Each orthogonalized basis function is a linear combination of the original basis functions plus a constant. The fitted spline function and the vector β of regression coefficients are unaffected by the change in basis functions. However, the regression coefficients for the basis functions may change following orthogonalization. These statements hold also for the more complex models involving the `stratify()` option. In these models, the coefficients of the basis functions may depend linearly on the covariates, **z**.

For example, a breast cancer dataset was supplied by Royston (2001) and used to exemplify aspects of modeling with `stpm`. A time-dependent odds model was fit, with results shown below:

(*Continued on next page*)

```
. stpm group2 group3, df(2) scale(odds) stratify(group2 group3)
(iteration log suppressed)
                                                  Number of obs   =        686
                                                  Wald chi2(2)    =      52.03
Log likelihood = -612.62274                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
s0           |
      group2 |  -2.443586    1.70277    -1.44   0.151    -5.780954    .8937813
      group3 |  -2.859845   1.646575    -1.74   0.082    -6.087073    .3673824
       _cons |   5.583426   1.605347     3.48   0.001     2.437003    8.729849
-------------+----------------------------------------------------------------
s1           |
      group2 |  -.2769887   .2216129    -1.25   0.211     -.711342    .1573647
      group3 |  -.3284494    .215424    -1.52   0.127    -.7506728     .093774
       _cons |   .5036676   .2068428     2.44   0.015     .0982631     .909072
-------------+----------------------------------------------------------------
xb           |
      group2 |   1.683409    .436123     3.86   0.000     .8286242    2.538195
      group3 |     2.8153   .4185716     6.73   0.000     1.994915    3.635686
       _cons |  -4.045346   .3831531   -10.56   0.000    -4.796313    -3.29438
------------------------------------------------------------------------------
Deviance =  1225.245 (686 observations.)
```

With the present update to `stpm`, the following output is obtained:

```
. stpm group2 group3, df(2) scale(odds) stratify(group2 group3)
(iteration log suppressed)
                                                  Number of obs   =        686
                                                  Wald chi2(2)    =      28.65
Log likelihood = -612.62274                       Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
s0           |
      group2 |  -1.203373   .7393709    -1.63   0.104    -2.652513    .2457677
      group3 |  -1.389217   .7146877    -1.94   0.052    -2.789979    .0115457
       _cons |   3.328259   .7000186     4.75   0.000     1.956248    4.700271
-------------+----------------------------------------------------------------
s1           |
      group2 |  -.5055262   .4044609    -1.25   0.211    -1.298255    .2872026
      group3 |  -.5994459   .3931657    -1.52   0.127    -1.370037    .1711447
       _cons |   .9192327   .3775043     2.44   0.015     .1793379    1.659128
-------------+----------------------------------------------------------------
xb           |
      group2 |   2.555608    .926667     2.76   0.006     .7393739    4.371842
      group3 |   3.849541   .8941082     4.31   0.000     2.097121    5.601961
       _cons |  -5.631324   .8748002    -6.44   0.000    -7.345901   -3.916747
------------------------------------------------------------------------------
Deviance =  1225.245 (686 observations.)
```

Everything appears to have changed, except for `Number of obs`, `Log likelihood`, and `Deviance`. That the log likelihood is identical shows that the fitted model has not changed. The model has merely been reparameterized.

Note that the constant term in the [xb] equation has changed from −4.045346 to −5.631324. This affects the results of predict <*varname*>, xb and follows from the orthogonalization of the basis functions, [s1], [s2], ... The functions are standardized to have mean 0 and variance 1.

5 Other changes to stpm

Orthogonalization of spline basis functions is the main change to stpm. In addition, the program has been updated to support version 8.1 and later of Stata, giving access to the current version of ml, Stata's maximum likelihood "engine". stpm no longer works with versions of Stata earlier than 8.1. An inappropriate choice of starting values for parameters occasionally caused stpm to report an infinite likelihood and stop with an error. This problem, resulting in no model being estimated, has been fixed. Finally, the help file has been updated.

Significantly, predict following stpm has been extended to provide the standard error of the log-hazard function for models with hazard scaling (fit with the scale(hazard) option of stpm). This permits calculation of confidence intervals for the hazard function through the expression $\exp\left[\ln\left(h\left(t\right)\right) \pm zs\left(t\right)\right]$ where $s\left(t\right)$ is the standard error of $\ln h\left(t\right)$ and is obtained via predict *varname*, hazard stdp and z is the appropriate standard normal deviate (1.96 for a 95% confidence interval). Additionally, predict now supports for all models the standard error of the derivative of the spline function via predict *varname*, dzdy stdp. With predict, the zero option is useful for giving baseline estimates and at() zero for getting estimates at particular values of the specified covariates with all other covariates set to zero.

In summary, users should find the new version more robust and faster than the first release. However, as always, please report anomalies and problems to the author.

6 References

Royston, P. 2001. Flexible parametric alternatives to the Cox model. *Stata Journal* 1(1): 1–28.

Royston, P. and M. K. B. Parmar. 2002. Flexible parametric-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 21: 2175–2197.

About the Author

Patrick Royston is a medical statistician with 25 years of experience, with a strong interest in biostatistical methodology and in statistical computing and algorithms. At present, he works in clinical trials and related research issues in cancer. Currently, he is focusing on problems of model building and validation with survival data, including prognostic factors studies; on parametric modeling of survival data; and on novel trial designs.