Nonlinear discrete-time hazard models for the rate of first marriage

Andy Batchelor[1], Heather L Turner[1] and David Firth[1]

[1] Department of Statistics, University of Warwick, Coventry CV4 7AL, United Kingdom. Email: heather.turner@warwick.ac.uk.

Abstract: We seek to model the hazard of entry into marriage for a sample of women in Ireland born between 1950 and 1973. Motivated by the work of Blossfeld and Huinink (1991), we propose a nonlinear discrete-time hazard model, which estimates the risk period and allows the effect of covariates on both the scale of risk and the age of maximum risk to be investigated.

Keywords: discrete-time survival analysis; non-proportional hazards; aliasing.

1 Introduction

In this paper we investigate the timing of first marriage for women in Ireland based on the Living in Ireland Surveys conducted by the Economic and Social Research Institute between 1994 and 2001. We limit our analysis to women born between 1950 and 1973, giving five, five-year cohorts who have passed the mean age at marriage for women in the full data set.

2 Linear Discrete-time Hazard Models

We first use the approach of Blossfeld and Huinink (1991), who proposed an exponential model for the hazard of first marriage, with baseline variables to control for the non-monotonic dependence of marriage rate on age:

$$r(t) = r_0 \exp\{\beta_L \log(age - 15) + \beta_R \log(45 - age) + \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2(t)'\boldsymbol{\beta}_2\}. \quad (1)$$

Here r_0 is the constant, baseline hazard; $\mathbf{x}_1$ and $\mathbf{x}_2$ are time-constant and time-varying covariates respectively, whilst $\log(age - 15)$ and $\log(45 - age)$ are the baseline variables that combine to produce a bell-shaped curve.
We only have the year of marriage, so we use episode splitting to generate yearly life course data, making appropriate adjustment for the month of birth. We then use the following discrete-time equivalent of Model 1:

$$C(r(t)) = \beta_0 + \beta_L \log(age - 15) + \beta_R \log(45 - age) + \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbf{x}_2(t)'\boldsymbol{\beta}_2, \quad (2)$$

where $C(r)$ is the complementary log-log transformation. Here age ranges from 15.04 to 44.96 years, so we keep the endpoints fixed at 15 and 45.

TABLE 1. Linear discrete-time hazard models.

	Model					
Variables	1	2	3	4	5	6
Intercept	-2.81	-17.92	-17.90	-19.31	-17.27	-17.21
Log(age - 15)		2.13	2.14	2.26	1.91	1.89
Log(45 - age)		3.63	3.67	4.14	3.70	3.67
Class s/skilled manual			-0.13	-0.10	-0.08	
Class skilled manual			-0.13	-0.06	-0.03	
Class non manual			-0.26	-0.22	-0.16	
Class low professional			-0.21	-0.18	-0.10	
Class high professional			-0.48	-0.43	-0.29	
Class missing			-0.07	-0.08	-0.02	
Cohort (54,59]				0.03	0.03	0.03
Cohort (59,64]				-0.08	-0.07	-0.07
Cohort (64,69]				-0.58	-0.55	-0.55
Cohort (69,74]				-1.30	-1.23	-1.23
In education					-1.52	-1.56
Deviance	13483	12414	12388	12086	11971	11981
Residual df	29866	29864	29858	29854	29853	29859

As far as possible, we follow Blossfeld and Huinink (1991) in building a model for our data, adding the baseline variables first, then social class, cohort and education variables. Our results are presented in Table 1. We find that women in later cohorts are less likely to marry and that the risk of marriage is significantly less whilst women are in education. Social class becomes insignificant when the education status is taken into account. Adding the final level of education does not significantly improve the model.

2.1 Nonlinear Discrete-time Hazard Models

We first consider extending Model 2 by defining the endpoints of the bell curve as parameters to be estimated:

$$\beta_0 + \beta_L \log(age - \alpha_L) + \beta_R \log(\alpha_R - age) \tag{3}$$

However we find that there is aliasing amongst the parameters in Equation 3, such that perturbations of one parameter can be compensated for by changes in the other parameters.

We therefore consider the following re-parameterization in which the aliasing is reduced:

$$\gamma - \exp(\delta) \left\{ \frac{(\nu - \alpha_L) \log\left(\frac{\nu - \alpha_L}{age - \alpha_L}\right) + (\alpha_R - \nu) \log\left(\frac{\alpha_R - \nu}{\alpha_R - age}\right)}{(\nu - \alpha_L) \log\left(\frac{\nu - \alpha_L}{\nu - D - \alpha_L}\right) + (\alpha_R - \nu) \log\left(\frac{\alpha_R - \nu}{\alpha_R - \nu + D}\right)} \right\} \tag{4}$$

Now the rate of marriage has a maximum of $C^{-1}(\gamma)$ at age ν and tends to zero as the age approaches α_L or α_R. The sharpness of the peak is captured

TABLE 2. Nonlinear discrete-time hazard models.

	Model				
Variables	6	7	8	9	10
Intercept(γ)	-2.12	-1.96	-1.68	-1.81	-2.31
Peak age (ν)					
Intercept	25.11	25.09	24.76	24.61	16.00
Education level (years)					0.78
Peakedness (δ)	-0.47	-0.45	-0.31	-0.41	-0.20
Left endpoint (α)	13.77	13.74	13.40	12.04	12.35
Class s/skilled manual		-0.13	-0.10		
Class skilled manual		-0.14	-0.06		
Class non manual		-0.26	-0.22		
Class low professional		-0.21	-0.19		
Class high professional		-0.49	-0.43		
Class missing		-0.07	-0.09		
Cohort (54,59]			0.03	0.03	0.06
Cohort (59,64]			-0.08	-0.07	-0.04
Cohort (64,69]			-0.58	-0.56	-0.53
Cohort (69,74]			-1.31	-1.25	-1.19
In education				-1.55	-0.65
Education level (years)					0.05
Deviance	12394	12368	12060	11960	11813
Residual df	29863	29857	29853	29858	29856

by δ, since the rate of marriage will be $C^{-1}(\gamma - \exp(\delta))$ at age $\nu - D$, where D is a fixed distance from ν, which we take to be 5 years.

Using the re-parameterization, we find that the fitted models are not significantly different from models in which $\alpha_R \to \infty$, where the baseline model is then:

$$\gamma - \exp(\delta) \left\{ \frac{(\nu - \alpha) \log \left(\frac{\nu - \alpha}{age - \alpha} \right) + age - \nu}{(\nu - \alpha) \log \left(\frac{\nu - \alpha}{\nu - D - \alpha} \right) - D} \right\} \tag{5}$$

Repreating the analysis of the previous section with this baseline model leads to a signficant improvement over the equivalent fixed endpoint models (Models 6 to 9 in Table 2).

We can improve the model further by including the additive effect of education level on both the maximum rate of marriage (γ) and the age at which this maximum is reached (ν), leading to a non-proportional hazard model (Model 10, Table 2). We represent the level of education by the equivalent years spent in education, based on averages from the data. We can see from the corresponding hazard and survival curves in Figure 1 that an increase in education level delays the age at which the marriage rate peaks and increases the maximum marriage rate, so that women with a higher education level eventually overtake those with a lower education level in terms of the proportion that marry.

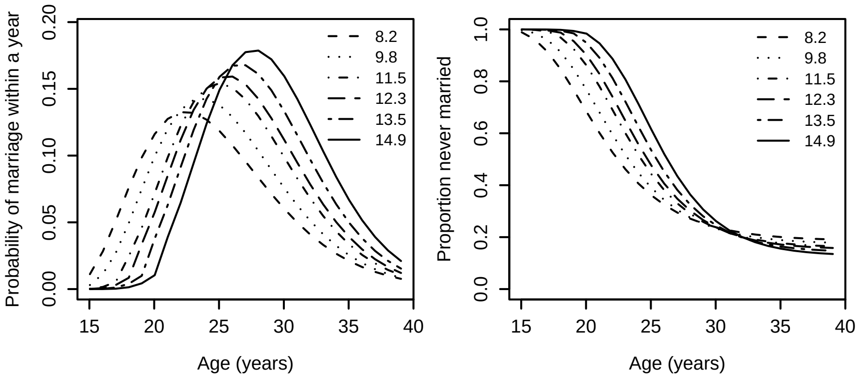

FIGURE 1. Hazard and survival curves under Model 10, for the $(59, 64]$ year cohort and skilled manual class, by increasing education level (equivalent years).

3 Summary

The nonlinear discrete-time hazard models we propose allow the risk period to be estimated and the effect of covariates on both the scale of risk and the age of maximum risk to be investigated. We find the latter to be important in describing the effect of education level on the risk of entry into marriage.

Software: The generalized nonlinear models described in this paper were fitted using the *R* package *gnm* (Turner and Firth, 2007).

Acknowledgments: The work of Heather Turner and David Firth was supported by ESRC Professorial Fellowship RES-051-27-0055 and by the ESRC National Centre for Research Methods, Lancaster-Warwick Node (ref RES-576-25-5020). The data are from The Economic and Social Research Institute Living in Ireland Survey Microdata File (©Economic and Social Research Institute). We acknowledge Carmel Hannan for introducing us to this application and providing background on the data.

References

Blossfeld, H.-P., and Huinink, J. (1991). Human Capital Investments or Norms of Role Transition? How Women's Schooling and Career Affect the Process of Family Formation. *American Journal of Sociology*, **97**, 143-168.

Turner, H.L. and Firth, D. (2007). Generalized nonlinear models in R: An overview of the gnm package. Documentation in the *gnm* package, `http://cran.r-project.org`.