This article was downloaded by: [University of Oxford]

On: 27 October 2010

Access details: Access Details: [subscription number 773573598]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

Mathematical Population Studies

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713644738>

Bayesian Adjustment of Anticipatory Covariates in Analyzing Retrospective Data

GEBRENEGUS GHILAGABER^a; JOHAN KOSKINEN^{b,c}

^a Department of Statistics, Stockholm University, Sweden ^b Swedish Institute for Social Research (SOFI), Stockholm University, Sweden ^c Department of Psychology, School of Behavioral Science, University of Melbourne, Australia

To cite this Article GHILAGABER, GEBRENEGUS and KOSKINEN, JOHAN(2009) 'Bayesian Adjustment of Anticipatory Covariates in Analyzing Retrospective Data', Mathematical Population Studies, 16: 2, 105 — 130

To link to this Article: DOI: 10.1080/08898480902790171

URL: <http://dx.doi.org/10.1080/08898480902790171>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Bayesian Adjustment of Anticipatory Covariates in Analyzing Retrospective Data

Gebrenegus Ghilagaber

Department of Statistics, Stockholm University, Sweden

Johan Koskinen

Swedish Institute for Social Research (SOFI), Stockholm University, Sweden, and Department of Psychology, School of Behavioral Science, University of Melbourne, Australia

In retrospective surveys, records on important variables such as the respondent's educational level and social class refer to what is achieved by the date of the survey. Such variables are then used as covariates in investigations of behavior such as marriage and divorce in life segments that have occurred before the survey. To what extent can any change in the behavior be attributed to the misclassification of respondents across the various levels of the anticipatory variable? To what extent do they reflect real differences in the behavior across the levels? The connection is obtained by a Bayesian adjustment, by specifying a continuous-time Markov model for the incompletely observed time-varying anticipatory covariates, and by implementing standard Bayesian data augmentation techniques. The issues are illustrated by estimating effects of educational level on risks of divorce in a multiplicative piecewise-constant hazard model. Results show that ignoring the time-inconsistency of anticipatory variables may seriously plague the analyses because the relative risks across the anticipatory educational level are overestimated.

Keywords: anticipatory analysis; Bayesian analysis; divorce; education; event-history analysis; MCMC; retrospective surveys

1. INTRODUCTION

Anticipatory covariates are variables whose values refer to what is attained by the date of the survey but are used to explain behavior in life course which took place before the survey. Highest educational level and social class at survey time are typical examples

Address correspondence to Gebrenegus Ghilagaber, Department of Statistics, Stockholm University, Stockholm SE-106 91, Sweden. E-mail: Gebre@stat.su.se

of anticipatory variables. Such variables are common in many retrospective studies because the data collection focuses on, for example, birth or employment histories but contain no history on educational careers or social class mobility.

In subsequent analyses of such retrospective survey data, these variables are used as regressors in modelling some outcome variable such as rates of marriage or divorce, events which took place long before the survey. This causes a *time inconsistency* because data that pertain to the date of the survey become less and less informative, the further from the date of the survey is the date of the event of interest.

For instance, if the level of education achieved at the time of the survey is used as a covariate in modelling a demographic outcome (marriage, childbearing, or divorce) which took place before the survey, educational progress is likely to occur between the time of the event (e.g., entry into marriage) and the survey. To what extent can the changes in patterns across educational levels be attributed to changes in the distribution of respondents across the various levels of education, and to what extent do they reflect real differences in behavior of different levels of education? Ignoring the anticipatory nature of such variables potentially produces biased estimates.

While many investigators are aware of this problem, our knowledge about the strength and direction of the biases in realistic situations is still too scanty. Hoem (1996) uses a case study on Swedish women to warn us that the use of anticipatory covariates may ruin a sensible analysis but concludes that the adverse effects may be less harmful for women who marry after age 20. Kravdal (2004) suggests stochastic imputation of both educational level and enrollment to minimize the adverse effects of incomplete educational histories in assessing the importance of education for fertility. In their study of mortality clustering in India using a retrospective history of births and deaths, Arulampalam and Bhalotra (2006) exclude any anticipatory (current-dated) regressors: household asset, toilet facility, electricity, or access to piped water at the date of the survey. This may not be the optimal solution because much valuable information may be lost by ignoring such variables. This view is also shared by Hoem and Kreyenfeld (2006a, 2006b) who argue that anticipatory variables may provide some useful information. These authors acknowledge that their alternative strategy is weak because it is based on unrealistic assumptions and fails to produce the information expected from anticipatory analysis. Faucett et al. (1998) concluded that a Bayesian missing data approach, apart from employing a more sound treatment of the uncertainty stemming from partially unknown covariates, gave interval estimates with superior coverage as compared to imputation.

We come up with bias-corrected estimates of effects of anticipatory covariates. We define a joint model for the outcome variable and the partially unobserved anticipatory variable. Using Bayesian methods we draw inference based on the posterior distributions of model parameters given the observations. We model risks of divorce among 1342 Swedish men born between 1936 and 1964 in a piecewise constant hazard model (Breslow and Day, 1975; Hoem, 1987). The results show that lack of proper account of anticipatory variables can seriously plague the analyzes because the relative risks of divorce across the anticipatory educational level are overestimated.

In Section 2 we provide a brief introduction of the standard multiplicative model with piecewise-constant baseline hazard. Sections 3 and 4 are devoted to a step-by-step description of our Bayesian approach: In Section 3, we extend the piecewise constant hazard model to a conditional hazard model, and introduce a covariate process with a view to adjust for its anticipatory nature; while in Section 4, we outline the inference procedure. Our model is fitted to data, in Section 5, and its performance is compared with that of a standard model which uses anticipatory covariates, as well as with one which is based on a reduced dataset.

2. THE STANDARD MULTIPLICATIVE TWO-FACTOR HAZARD MODEL

For a sample of U individuals, consider J educational levels and let D_{ij} be the total number of occurrences, say divorces, at duration of marriage i ($i = 1, \dots, I$) for the j th educational level ($j = 1, 2, \dots, J$) for T_{ij} years of observed exposure to the risk (of divorce). The covariate indexed by i is the time variable (duration of marriage) measured from the date of marriage until the date of divorce (for those who got divorced) or until the interview date (for those still in marriage).

Define

$$D_{i+} = \sum_{j=1}^J D_{ij}, \quad D_{+j} = \sum_{i=1}^I D_{ij} \quad (1)$$

and

$$D_{++} = \sum_{i=1}^I \sum_{j=1}^J D_{ij} = \sum_{i=1}^I D_{i+} = \sum_{j=1}^J D_{+j}, \quad (2)$$

and let T_{i+} , T_{+j} , and T_{++} represent similar quantities for the exposure variable T . Usually, divorce risks are assumed to be *piecewise constant* in each of the I time intervals but may vary between the intervals.

The time to divorce is assumed to follow a *piecewise exponential distribution* for each educational level. The density function of the time to divorce in duration group i for a person k with educational level j is given by

$$f(t_{ijk}) = \lambda_{ij} \exp(-\lambda_{ij} t_{ijk}), \quad (3)$$

where λ_{ij} is the hazard of divorce at marriage duration i and for individuals with educational level j . The populations defined by the J educational levels have been observed over a fixed time period, and censoring is possible so long as it is noninformative in the sense of Lagakos (1979).

A multiplicative model for the λ_{ij} is

$$\lambda_{ij} = \beta_i \alpha_j, \quad (4)$$

where the divorce rates are obtained from multiplicative contributions of the i th duration group (β_i) and j th level of education (α_j). A model of this form is discussed in Breslow and Day (1975) while Hoem (1987) reviews the statistical theory behind the model.

This model has $I + J - 1$ parameters in general (namely $\beta_1, \beta_2, \dots, \beta_I$, and $\alpha_2, \alpha_3, \dots, \alpha_J$), for α_1 is tied down by the normalization $\alpha_1 = 1$. α_j measures the relative risk of divorce for individuals with educational level j (relative to those with $j = 1$), while $\lambda_{i1} = \beta_i \alpha_1 = \beta_i(1) = \beta_i$ is the risk of divorce at duration-group i in the baseline educational group ($j = 1$).

To construct the likelihood function when Eq. (4) holds, we first define D_{ijk} as an indicator variable of whether the k th member having the j th level of education is divorced ($D_{ijk} = 1$) or is still in marriage ($D_{ijk} = 0$) in the i th duration of marriage. From Eq. (3) and (4), the contribution to the likelihood of the subsample of individuals in the i th duration-group and having the j th level of education is

$$\Lambda_{ij} = \prod_k ((\beta_i \alpha_j)^{D_{ijk}} \exp(-t_{ijk} \beta_i \alpha_j)) = (\beta_i \alpha_j)^{D_{ij}} \exp(-T_{ij} \beta_i \alpha_j), \quad (5)$$

where

$$D_{ij} = \sum_{k=1}^U D_{ijk}, \quad \text{and} \quad T_{ij} = \sum_{k=1}^U t_{ijk}. \quad (6)$$

The likelihood for the entire sample is the product of the Λ_{ij} over all levels of i and j :

$$\Lambda = \prod_{i=1}^I \prod_{j=1}^J \Lambda_{ij} = \prod_{i=1}^I \prod_{j=1}^J ((\beta_i \alpha_j)^{D_{ij}} \exp(-T_{ij} \beta_i \alpha_j)). \quad (7)$$

From Eq. (7), the maximum likelihood estimates of the baseline hazards $\hat{\beta}_i$ and relative hazards $\hat{\alpha}_j$ are functions of the total number of events D_{ij} and the exposure times T_{ij} . The misclassification of events or times of exposure into wrong intervals or into wrong levels of the covariate, as it is the case with an anticipatory covariate, leads to incorrect estimates. Our method is intended to adjust for the anticipatory nature of covariates in order to reduce such serious errors in estimating the parameters β_i and α_j .

3. ADJUSTING ANTICIPATORY COVARIATES USING BAYESIAN ANALYSIS

The parameters β_i and α_j in Eq. (4) are usually assumed unknown and to be estimated from data. Here, these parameters along with the anticipatory covariate will be treated as variables with some prior distribution. In order to specify the joint model for data and covariates, we need to reformulate the hazard rate model of Eq. (4). For any given realization of the covariate model, the conditional model for data must be consistent with the one for which all covariate values were observed.

Educational attainment can be viewed as time-varying and the covariate model comprises two ingredients. Firstly, the anticipatory nature of data, the fact that observations are made at one point in time, motivates a continuous-time model for the covariates. Second, it seems plausible to exclude a respondent would lower his educational level and hence educational attainment ought to be nondecreasing. These two elements hinder us from using here the procedure in Faucett et al. (1998), where a missing covariate was used to account for the partially missing smoking status.

We assume that the educational career is independent of the time of marriage. This assumption implies that observations are missing at random (MAR) in the sense of Rubin (1976), as an observation on the covariate is missing whenever the highest reported educational level is completed after the time of marriage. The probability that a datum is missing is independent of the missing data (Rubin, 1976; Little and Rubin, 1987; Schafer, 1997). In the context of our illustration it is conceivable that the educational progress would differ for married and unmarried individuals, in which case the MAR assumption would be violated. How to best specify a joint model for education and marriage is an empirical issue. How deviations from MAR may be addressed in the anticipatory adjustment procedure and the sensitivity of our result to violations of MAR are not pursued here. Interested readers

can consult Lillard et al. (1994) where procedures for modelling jointly multiple trajectories in the life course are described.

3.1. The Conditional Piecewise Exponential Hazard Rate Model

For individual k , consider a continuous-time Markov chain $\{X_k(t) : t \in \Delta = [0, \tau_k^1]\}$, on the finite outcome space $\mathcal{J} = \{0, \dots, J\}$. The elements of $\mathcal{J}$ can be different educational levels with J being the highest possible educational level. The process is modelled as being right-continuous and nondecreasing. Taking these two elements together we consider the processes in terms of the union $\bigcup_{j=0}^{\max_t X_k(t)} \Delta_j^*$ of distinct nonoverlapping intervals

$$\Delta_j^* = [t_{j,0}, t_{j,1}) = \{t \in \Delta : X_k(t) = j\}, \quad (8)$$

where the lower and upper bounds are defined such that

$$X_k(t_{j,0}) = j \quad \text{and} \quad X_k(t_{j,1}) = X_k(t_{j+1,0}) = j + 1, \quad (9)$$

respectively. Conditional on a realization X_k of the process, we define the right-continuous Markov chain $\{Y_k(t) : t \in \Delta\}$, defined on the outcome space $\mathcal{Y} = \{0, 1\}$, which is also nondecreasing. An example is given in Figure 1, where the educational level, $x_k(t)$, (thin line) stays at level 0 until just after $t = 1.5$, after which it jumps to educational level 1. The data process, for example, marriage status, $y_k(t)$ (thick line), remains in state 0 for a little while, longer than the educational level process but then jumps to the marriage status 1. Δ is partitioned into intervals Δ_i , $i \in \mathcal{I} = \{1, \dots, I\}$, representing for example different stages of marriage. Conditional on X_k we form new intervals through the intersection of the intervals Δ_j^* and Δ_i :

$$\Delta_{ij} = \Delta_i \cap \Delta_j^*. \quad (10)$$

The intervals are functions of x_k but the notational dependence on k and the corresponding variable are omitted. For instance, if $\Delta_i = [3, 6)$ and $\Delta_j^* = [4, 7)$, then $\Delta_{ij} = [4, 6)$ is the period when an individual had been married between 3 and 6 years and had achieved educational level j . The process $\{Y_k(t) : t \in \Delta\}$ can be thought of as a nonrecurring event and for each interval Δ_{ij} , we define the time to event t_{ijk}^* with exponential density

$$f(t_{ijk}^* | \lambda_{ij}) = \lambda_{ij} \exp(-\lambda_{ij} t_{ijk}^*), \quad (11)$$

FIGURE 1 A sketch of the underlying processes (covariate process $x_k(t)$, thin line, response variable process, $y_k(t)$, thick line).

for $\lambda_{ij} > 0$. For each interval Δ_{ij} we define a variable D_{ijk} indicating whether the event takes place in interval Δ_{ij} or not. Equivalently, if we let $m(a, b)$ be the length of the interval (a, b) , then

$$D_{ijk} = \mathbf{1}_{\{t_{ijk}^* < m(\Delta_{ij})\}}. \quad (12)$$

We do not allow for withdrawal without having experienced the event (Holford, 1976) at this point. We assume that the events are nonrecurring and denoting by $\partial\Delta_{ij}$ and $\partial(i, j)$, respectively, the interval $\Delta_{uv} = [t_a, t_b)$ and pair (u, v) , such that t_b is the lower bound of Δ_{ij} , we have

$$\Pr(D_{ijk} = 0 \mid D_{\partial(i,j)k} = 1) = 1 \quad (13)$$

and

$$\begin{aligned} & \Pr(D_{ijk} = 1 \mid D_{\partial(i,j)k} = 0, D_{\partial(i,j)k} = 0, \dots, D_{0,0,k} = 0) \\ &= \Pr(t_{ijk}^* < m(\Delta_{ij})) \\ &= 1 - \exp(-m(\Delta_{ij})\lambda_{ij}). \end{aligned}$$

Eq. (13) follows from the assumption that the process $Y_k(t)$ is nonrecurring and Eq. (14) from the properties of the exponential distribution. We also have the alternative characterization

$$D_{ijk} = \max\{Y_k(t) : t \in \Delta_{ij}\} - \max\{Y_k(t) : t \in \partial\Delta_{ij}\}. \quad (14)$$

As the time to event has been truncated to the right in the upper interval bound, we define the *exposure time* as

$$t_{ijk} = \min\{t_{ijk}^*, m(\Delta_{ij})\}. \quad (15)$$

The process $\{Y_k(t), X_k(t)\}_{t \in \Delta}$ completely determines $\{D_{ijk}, t_{ijk}\}$. Given X , $\{D_{ijk}, t_{ijk}\}$ completely determines $\{Y_k(t)\}_{t \in \Delta}$. An example of observation $\{Y_k(t), X_k(t)\}_{t \in \Delta}$ is given in Figure 1. The dependent process $\{Y_k(t)\}_{t \in \Delta}$ in Figure 1 is represented by the thick band and is superimposed on the covariate process $\{X_k(t)\}_{t \in \Delta}$. Forming the intervals Δ_j^* according to Eq. (8), we plot the process in terms of $\{D_{ijk}, t_{ijk}\}$ in a Lexis diagram in Figure 2.

In summary, for a given realization x_k , we have a piecewise exponential hazard rate model with hazard rate λ_{ij} for each interval

FIGURE 2 The Lexis diagram of the conditional process of $\{D_{ijk}, t_{ijk}\}$ (marriage duration intervals Δ_i , and educational level intervals Δ_j^*).

Δ_{ij} and density given by

$$p(y | \lambda, x) = \prod_{i,j} \lambda_{ij}^{D_{ijk}} \exp(-\lambda_{ij} t_{ijk}), \quad (16)$$

where the product is taken over relevant intervals (i, j) so that respondents who have experienced the event are removed because they are no longer at risk of experiencing the event of interest. For a sample of independent observations $\{D_{ijk}, t_{ijk}\}$, products over k in Eq. (16) reduce to the likelihood given in Eq. (7).

3.2. Modelling the Covariate Process

By proposing a model for the covariate process we provide a way of predicting what is missing. A simple way of implementing this is through multiple imputation. Thus, the educational status of individuals may be imputed according to the covariate-model specified in Eq. (17), and these imputed values can then be used as time-varying covariates in a Cox proportional hazards semi-parametric model. Analysis from different imputations would then be combined in standard fashion (Schafer, 1997). Imputation of missing time-varying covariates in the Cox proportional hazards model is treated in, for example, Paik and Tsai (1997) and Chen (2002).

We prefer a fully parametric model for covariates to a deterministic or stochastic imputation of missing covariate values (Kravdal, 2004). This allows for a consistent way of jointly analyzing the dependent variable process and the covariate process. It is for example hard to say how the approximation entailed by imputation affects inference. Available information on the covariates is incorporated into the joint model and subsequently updated in the estimation of the model parameters. If the uncertainty stemming from having observed the covariate values is neglected partially, the uncertainty about the parameter estimates is underestimated.

For a sample U of individuals, let k , as before, indicate an observation for the unit $k \in U$. For a simple model for $\{X_k(t)\}_{t \in \Delta^k}$, assume that the educational level follows a continuous-time Markov chain and that the time spent in the state $x \in \mathcal{J}$ is exponential with rate η_{x+1} for $x < J$, and that $x = J$ is an absorbing state. Denote the $x_k(\tau_k^1)$ holding times of $\{X_k(t)\}_{t \in \Delta^k}$ by $s_{1k}, \dots, s_{x_k(\tau_k^1)k}$. In the notation of the previous section $s_{jk} = m(\Delta_j^*)$. A convenient way to model the holding times is to assume that, independent of $j \in \mathcal{J}$, $s_{jk} \sim \exp(\eta_j)$.

By modelling the exposure time, which is the duration spent in each educational level, with an exponential distribution, we *adjust the*

exposure times to divorce at each educational level. Given the vector of parameter values $\eta = (\eta_1, \dots, \eta_J)^T$, the density of x_k is given by

$$p(x_k | \eta) = C(x_k) \prod_{j=0}^{x_k(\tau_k^1)} \eta_j \exp(-s_{jk} \eta_j), \quad (17)$$

where

$$C(x_k) = \begin{cases} \exp(-s_{x_k(\tau_k^1)+1,k} \eta_{x_k(\tau_k^1)+1}) & \text{if } x_k(\tau_k^1) < J \\ 1 & \text{otherwise,} \end{cases} \quad (18)$$

stems from the censoring.

A drawback of this formulation is that the model is based on the tacit assumption that sooner or later all individuals will end up in the absorbing state (i.e., attain the highest educational level). This is unrealistic and could introduce a noise. If individuals reach their highest levels of education very early and if j is relatively far from the absorbing state J , the observed time during which these individuals remained in j should not bring any piece of information above the times between transitions. We introduce another unobserved latent variable, indicating whether an individual's reported educational level is "terminal" or not. In our empirical illustration, this is obtained with Bernoulli trial.

4. INFERENCE IN THE ADJUSTED MODEL

4.1. Posterior Distributions

In addition to the covariate model and the model for data, prior distributions for the unknown parameters help us perform statistical inference. The objective is to obtain the posterior distributions for the parameters once we have a sample U with independent observations on event times and highest reported values on the covariate (education).

For individual $k \in U$, we observe $\{Y_k(t)\}_{t \in \Delta^k}$, and at times $\tau_k \in \Delta^k$, the highest achieved educational level, $x_k(\tau_k)$. The highest achieved educational level for example provides us with the information that $x_k(t) = x_k(\tau_k)$ for $t \in \tilde{\Delta}^k = [\tau_k, \tau_k^1)$. Let the indicator function

$$\mathbf{1}_{\{x_k(t) \preceq x_k\}}^{\tilde{\Delta}^k} = \min_{t \in \tilde{\Delta}^k} \{\mathbf{1}_{\{x_k(t) = x_k(\tau_k)\}}\} \quad (19)$$

indicate whether or not $\{x_k(t)\}$ is concordant with the observation $x_k(\tau_k)$, the educational level as described by $\{x_k(t)\}$ does not change after k has achieved his highest educational level. Assuming that λ

and η are independent of each other a priori, with prior distributions $\pi(\lambda)$ and $\pi(\eta)$, the joint posterior of $X = \{x_k(t) : t \in \Delta^k, k \in U\}$, λ and η given data $y = \{y_k(t) : t \in \Delta^k, k \in U\}$ and $\{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}$, is proportional to

$$\pi(\lambda)\pi(\eta) \prod_{k \in U} p(y_k | \lambda, x_k) p(x_k | \eta) \mathbf{1}_{\{x_k(t) \preceq \tilde{x}_k\}}. \quad (20)$$

4.2. Outline of the Inference Scheme

Exact inference for model variables is hampered by the fact that the joint posterior distribution $\pi(\lambda, \eta, X | \{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y)$ in Eq. (20) is only specified up to an analytically intractable multiplicative constant. We use MCMC to simulate this distribution (Gilks et al., 1996). We write the full conditional posterior distributions

$$\pi(\theta_s | \theta_{-s}, \text{Data}) \quad (21)$$

for each block θ_s of posterior variables $\theta_1, \dots, \theta_S$, up to a multiplicative constant. The subscript “ $-s$ ” refers to all other indices except s . Here $\theta_1, \dots, \theta_S$ include parameters λ and η , as well as X .

We produce a sequence of draws $\{\theta_1^{(g)}, \dots, \theta_S^{(g)}\}_{g=0}^G$, by circling through all the S coordinates, sequentially performing draws

$$\theta_s^{(g)} \sim \theta_s | \theta_1^{(g)}, \dots, \theta_{s-1}^{(g)}, \theta_{s+1}^{(g-1)}, \dots, \theta_S^{(g-1)}, \text{Data} \quad (22)$$

for each $g > 0$.

For some of the coordinates the full conditional posterior $\theta_s | \theta_{-s}, \text{Data}$ is a standard distribution and we perform a draw directly, a Gibbs updating step. For other coordinates, we only have

$$\pi(\theta_s | \theta_{-s}, \text{Data}) \propto f(\theta_s | \theta_{-s}, \text{Data}), \quad (23)$$

for which we use a Metropolis updating step, proposing a move to θ_s^* , drawn from the distribution $q(\theta_s^* | \theta_s)$, and accepting this with probability $\alpha = \min\{1, A\}$, where

$$A = \frac{f(\theta_s^* | \theta_{-s}, \text{Data}) q(\theta_s | \theta_s^*)}{f(\theta_s | \theta_{-s}, \text{Data}) q(\theta_s^* | \theta_s)}. \quad (24)$$

The distribution $q(\theta_s^* | \theta_s)$ used for drawing new candidate states is called the proposal distribution.

The resulting sequence of draws $\{\theta_1^{(g)}, \dots, \theta_S^{(g)}\}_{g=0}^G$ is a sample from the joint posterior distribution of $\theta_1, \dots, \theta_S$ given Data. This is guaranteed only as G tends to infinity, but the generated sample can

be treated as an “exact” sample from the posterior distribution with the usual caveats regarding convergence and auto-correlation (Gilks et al., 1996; Tierney, 1994). Point estimates of any relevant quantity $g(\theta_1, \dots, \theta_S)$ can be approximated by the corresponding ergodic mean of this quantity over the sample, $\bar{g}(\theta_1, \dots, \theta_S) = G^{-1} \sum g(\theta_1^{(g)}, \dots, \theta_S^{(g)})$.

4.3. The Covariate Process and Its Parameters

4.3.1. Exposure Times and Educational Careers

The total number of intervals of the type $\Delta_j^{*,k}$ in Eq. (8) is fixed at $x_k(\tau_k^1) = x_k(\tau_k)$. However, once the highest achieved educational state is recorded, the total number of intervals constructed from Eq. (8) varies with the location of the latent transitions in the covariate process. For some $(i, j) \in \mathcal{I} \times \mathcal{J}$, we can have $\Delta_{ij}^k = \emptyset$, which forbids defining the exposure time t_{ijk} . Whenever $\Delta_{ij}^k = \emptyset$, we set $t_{ijk} = m(\Delta_{ij}^k) = 0$, and $D_{ijk} = 0$ by convention. These “structural zeros” correspond to the rectangles in the Lexis diagram which are never entered by the process ($\Delta_{2,2}$ in Figure 2).

When the highest reported educational level is observed

$$x_k(\tau_k) = j \quad (25)$$

implies that

$$x_k(t) = j \quad \text{for } t \in \tilde{\Delta}^k \quad \text{and} \quad x_k(t) < j \quad \text{for } t < \tau_k. \quad (26)$$

When the time of transition is not observed, we have

$$x_k(t) = j \quad \text{for } t \in \tilde{\Delta}^k \quad (27)$$

but

$$x_k(\tau_k - \epsilon) = j \quad \text{for some } \epsilon > 0. \quad (28)$$

4.3.2. Conditional Posterior Distribution of Educational Careers

For observed $\{Y_k(t)\}_{t \in \Delta^k}$, and $\{x_k(t)\}_{t \in \tilde{\Delta}^k}$ for $k \in U$, and parameters θ and η , the fully conditional posterior distribution of x can be treated separately for each individual k . The fully conditional posterior of x_k is proportional to

$$\begin{aligned} & w(x_k; y_k, \lambda, \eta, \{x_k(t)\}_{t \in \tilde{\Delta}^k}) \\ &= p(y_k | \lambda, x_k) p(x_k | \eta) \mathbf{1}_{\{x_k(t) \preceq x_k\}} \\ &= \prod_{i,j} \lambda_{ij}^{D_{ijk}} \exp(-t_{ijk} \lambda_{ij}) \prod_{j=0}^{x_k(\tau_k^1)} \eta_j \exp(-s_{jk} \eta_j) C(x_k) \mathbf{1}_{\{x_k(t) \preceq x_k\}}. \end{aligned} \quad (29)$$

As mentioned earlier, the sequence $\{X_k(t) : t \in \Delta\}$ is completely determined by its holding times $s_k = (s_{1k}, \dots, s_{x_k(\tau_k^1)_k})$, and we could generate sequences s_k of holding times from some density q to implement a Metropolis-Hastings updating step for the unobserved educational careers. As $w(x_k; y_k, \lambda, \eta, \{x_k(t)\}_{t \in \tilde{\Delta}^k}) > 0$ for sequences concordant with $x_k(\tau_k)$, it is unnecessary to generate sequences of holding times for elements different from $x_k(\tau_k)$.

When the differences between the η_j s are small, given s_k with the implied sequence x_k in the current iteration, we propose a move to s'_k with the implied sequence x'_k , drawn from the re-scaled Dirichlet distribution with density

$$q(s'_k) = \frac{\Gamma(\sum_{j=1}^{x_k(\tau_k)} \nu_j)}{\prod_{j=1}^{x_k(\tau_k)} \Gamma(\nu_j)} \tau_k^{1-\sum \nu_j} \prod_{j=1}^{x_k(\tau_k)} (s'_{jk})^{\nu_j-1}, \quad (30)$$

such that $\sum s'_{jk} = \tau_k$, and where the parameters $\nu_1, \dots, \nu_{x_k(\tau_k)}$ can depend on η or s_k .

The move is accepted with probability $\min\{1, A\}$, for

$$A = \frac{w(x'_k; y_k, \lambda, \eta, x_k(\tau_k)) q(s_k)}{w(x_k; y_k, \lambda, \eta, x_k(\tau_k)) q(s'_k)}. \quad (31)$$

When the time of transition is not observed, we need an additional holding time for the time elapsed from $\min(\Delta_{x_k(\tau_k)}^*)$ until τ_k .

If too many proposed moves are rejected we introduce dependence between the current state of the latent educational career and the proposed educational career by “centering” the proposal distribution over the current state. As the Dirichlet distribution has one parameter for each coordinate, a suitable way of centering the proposal distribution over the current state is by fixing the first moment. For s'_k following a $\text{Dirichlet}(\nu)$, the expected values are given by $E(s'_k) = \tau_k \nu_j / \sum_{j=1}^{x_k(\tau_k)} \nu_j$. If we let $\nu_j = s_{jk}$ in drawing s'_k , then the proposal distribution of s'_k given s_k is centered over s_k and the ratio of proposal densities is given by

$$\begin{aligned} \ln \frac{q(s_k)}{q(s'_k)} &= \sum (\ln \Gamma(s_{jk}) - \ln \Gamma(s'_{jk})) + (\sum (s_{jk} - s'_{jk})) \ln \tau_k \\ &\quad + \sum (s'_{jk} - 1) \ln s_{jk} - \sum (s_{jk} - 1) \ln s'_{jk}. \end{aligned} \quad (32)$$

4.3.3. Conditional Posterior Distribution of the Parameters in Educational Careers

As η and λ are independent of each other a priori, the only relevant information for updating the covariate model parameters is provided by the realizations on the educational careers. This is reflected in

the fully conditional distribution of η given all the rest, which is proportional to

$$\begin{aligned} \pi(\eta) \prod_{k \in U} p(y_k | \lambda, x_k) p(x_k | \eta) \mathbf{1}_{\{x_k(t) \succeq x_k\}} \\ \propto \pi(\eta) \prod_{k \in U} \prod_{j=1}^{x_k(\tau_k^1)} \eta_j \exp(-s_{jk} \eta_j) C(x_k) \\ = \pi(\eta) \prod_{j=1}^J \eta_j^{n_j} \exp(-s_{j+} \eta_j), \end{aligned} \quad (33)$$

where $n_j = \sharp\{k \in U : x_k(t) = j, \text{ for some } t \in \Delta^k\}$, and $s_{j+} = \sum_{k \in U} s_{jk}$. If η_j has conjugate $\text{gamma}(\gamma_1^j, \gamma_2^j)$ prior, then the fully conditional posterior of η given the rest is $\text{gamma}(\gamma_1^j + n_j, \gamma_2^j + t_{j+}^j)$.

4.4. Conditional Distributions for the Model Parameters

We have not dealt with the parameters of interest yet. This part of the inference procedure is straightforward. It is the way a Bayesian inference scheme would be implemented, if we had completely observed the relevant part of the educational careers, τ_k^0 , following the onsets. This portion of the updating scheme requires that we have classified the exposure times (t_{ijk}) and the event indicators (D_{ijk}). This was also the intention in modelling the piecewise exponential hazard rate model, conditional on the realized educational career.

While the comprehensive model for data and covariates seems complex, the part of the inference procedure involving the model parameters is straightforward and involves standard distributions.

Consider a model with multiplicative hazard rates, $\lambda_{ij} = \beta_i \alpha_j$ such that

$$\begin{aligned} \pi(\beta_i | \beta_{-i}, \alpha, y, x, \eta) &\propto \pi(\lambda) \prod_{i,j} \lambda_{ij}^{D_{ij}} \exp(-t_{ij} \lambda_{ij}) \\ &\propto \pi(\lambda) \beta_i^{D_{i+}} \exp\left(-\beta_i \sum_j t_{ij} \alpha_j\right), \end{aligned} \quad (34)$$

where $D_{i+} = \sum_j D_{ij}$. With conjugate prior $\text{gamma}(\tilde{\beta}_i^{(1)}, \tilde{\beta}_i^{(2)})$, the fully conditional posterior is

$$\beta_i | \beta_{-i}, \alpha, y, x, \eta \sim \text{gamma}\left((\tilde{\beta}_i^{(1)} + D_{i+}), \left(\tilde{\beta}_i^{(2)} + \sum_j t_{ij} \alpha_j\right)\right). \quad (35)$$

Analogously, with α_j gamma($\tilde{\alpha}_j^{(1)}, \tilde{\alpha}_j^{(2)}$) a priori,

$$\alpha_j | \alpha_{-j}, \beta, y, x, \eta \sim \text{gamma}\left((\tilde{\alpha}_j^{(1)} + D_{+j}), (\tilde{\alpha}_j^{(2)} + \sum_i t_{ij} \beta_i)\right). \quad (36)$$

The shape parameter for β_i in the fully conditional posterior consists of a part which represents our prior knowledge, $\tilde{\beta}_i^{(1)}$, and a part which represents data, D_{i+} . The latter is simply a count of the total number of individuals having experienced the event in duration interval i . A hyper-parameter $\tilde{\beta}_i^{(1)}$ equal to 1 has the weight of one observation and when, as in the case with the priors used in the empirical illustration, we set $\tilde{\beta}_i^{(1)} = 1/1000$, the weight assigned to the prior information is a thousandth of an observation. Analogous interpretations hold for the scale (hyper-)parameter $\tilde{\beta}_i^{(2)}$ and its associated function of the exposure times, $\sum_j t_{ij} \alpha_j$.

4.5. Accounting for Long-Term Survivors

The covariate model with exponentially distributed holding times is limited because (i) we are not at liberty to set the location and scale separately and (ii) the same educational state is assumed to be terminal for all individuals although some individuals may never attain the highest educational state. We have the so-called long-term survivors (individuals who will never experience the event of interest; in our case, progressing to the next possible educational level). The first problem is alleviated by assuming that the time during which individual $k \in U$ stays in educational state $x - 1$ is gamma(ζ_x, η_x) (instead of exponential). The second problem is couched in terms of varying absorbing states J_k . When the educational process for an individual jumps from a state x to $x + 1$, we flip a coin with probability of success ϕ_{x+1} . If the coin-flip is a success we let $J_k = x + 1$, individual k becomes a long-term survivor, otherwise the process continues. The covariate model is

$$\begin{aligned} p(x_k, J_k | \eta, \zeta, \phi) &= \prod_{j=1}^{\min(x_k(\tau_k^1)-1, J_k-1)} (1 - \phi_j) C(x_k, J_k) \\ &\times \prod_{j=0}^{x_k(\tau_k^1)} \frac{\eta_j^{\zeta_j}}{\Gamma(\zeta_j)} \exp(-s_{jk} \eta_j) \mathbf{1}_{\{J_k \geq x_k(\tau_k^1)\}}, \end{aligned} \quad (37)$$

and the censoring is a function of both the highest achieved state $x_k(\tau_k^1)$ and the terminal state J_k :

$$C(x_k, J_k) = \begin{cases} \Gamma(\zeta_j, \eta_j; s_{x_k(\tau_k^1)+1, k}) & \text{if } x_k(\tau_k^1) < J_k \\ \phi_{x_k(\tau_k^1)} & \text{otherwise,} \end{cases} \quad (38)$$

where

$$\Gamma(a, b, t) = \frac{b^a}{\Gamma(a)} \int_t^\infty x^{a-1} e^{-xb} dx. \quad (39)$$

The fully conditional posterior of the state parameters is proportional to

$$\pi(\phi) \prod_{j=1}^{J-1} \phi_j^{J^{(j)}} (1 - \phi_j)^{n - J_+^{(j)}}, \quad (40)$$

where

$$J^{(j)} = \#\{k \in U : J_k = j\}, \quad J_+^{(j)} = J^{(1)} + \dots + J^{(j)}, \quad \text{and} \quad n = |U|. \quad (41)$$

Consequently, when we a priori let ϕ_j be independent beta($\tilde{\phi}_j^1, \tilde{\phi}_j^2$), for $j = 1, \dots, J-1$, the fully conditional posterior of ϕ_j is = beta($\tilde{\phi}_j^1 + J^{(j)}, \tilde{\phi}_j^2 + n - J_+^{(j)}$).

Because of the censoring factor $C(x_k, J_k)$, the fully conditional posterior of the parameters in the models for the times to educational progress, ζ and η , is not as straightforward as when we had exponential holding times. It is given by

$$\begin{aligned} \pi(\eta, \zeta) \prod_{k \in U} \Gamma(\zeta_j, \eta_j; s_{x_k(\tau_k^1)+1, k})^{1_{\{J_k > x_k(\tau_k^1)\}}} \prod_{j=0}^{x_k(\tau_k^1)} \frac{\eta_j^{\zeta_j}}{\Gamma(\zeta_j)} s_{jk}^{\zeta_j-1} \exp(-s_{jk} \eta_j) \\ = \pi(\eta, \zeta) \prod_{k \in U} \Gamma(\zeta_j, \eta_j; s_{x_k(\tau_k^1)+1, k})^{1_{\{J_k > x_k(\tau_k^1)\}}} \\ \times \prod_{j=0}^J \left(\frac{\eta_j^{\zeta_j}}{\Gamma(\zeta_j)} \right)^{n_j} \left(\prod_{k \in U} s_{jk} \right)^{\zeta_j-1} \exp(-s_{j+} \eta_j), \end{aligned} \quad (42)$$

where n_j is defined as before and where

$$s_{j+} = \sum_{k \in U: J_k > x_k(\tau_k^1)} s_{jk} \quad (43)$$

with the sum taken over all no long-term survivors.

We illustrate the models with empirical data on divorce among Swedish men born between 1936 and 1964.

5. EFFECT OF EDUCATION ON DIVORCE

5.1. Background

In the study of effects of educational level attained by the time of marriage on the risks of divorce, based on retrospective survey data, we assume that some of the respondents have improved their educational level between the time of marriage and the date of interview. At each recorded educational level, except the lowest one, some individuals at the time of marriage belonged to a lower level. If those with lower education have lower risks of divorce, then we suspect that the empirical risks of divorce corresponding to the recorded educational levels overestimate the true rates because many of the respondents would be misclassified to higher educational levels. The degree of overestimation depends on the extent of educational progress and the actual values of the relative risks, both unknown to us. The aim of this section is to estimate the degree of overestimation.

5.2. The Dataset

The dataset is an extract from the 1985 Mail Survey of Swedish Men, where information was collected on background variables as well as entry into and exit from marital and nonmarital unions, and childbearing. Our work is based on 1,312 ever-married men who were either divorced or still married by the survey time. These men have been exposed to the risk of divorce. Their distribution with age at marriage and with age of educational completion is shown in Figure 3. An appreciable amount of the men (those below the diagonal) have completed their education after marriage. Anticipatory analysis, common in the analysis of such type of data, amounts to moving the values below the diagonal in Figure 3 to the left.

A cross tabulation of the sample (Table 1) shows differentials in percentage divorced across the anticipatory status of education. Therefore, investigating the role of misclassification on such differentials is worthwhile.

5.3. Models

The time variable (duration of marriage in years) was categorized into five intervals: $0-1^-$, $1-2^-$, $2-3^-$, $3-6^-$, and 6 years and over. The lowest educational level (primary level) was set as baseline level. Its corresponding parameter (relative hazard) was set to $\alpha_1 = 1$. Vague gamma(1/1000, 1/1000) priors were assumed for all parameters,

FIGURE 3 Age of completed education versus age at marriage for a sample of 1,312 Swedish men.

except for ϕ , the coordinates of which were assumed to be a priori independent $\text{beta}(1, 1)$, uniform on the interval $(0, 1)$.

In order to make the Bayesian correction scheme comparable to the anticipatory approach, the latter is also carried out within a Bayesian

TABLE 1 Data Structure for the Sample of 1312 Swedish Men

Anticip. status	Highest edu. lev.	Marital status at survey			% Div.
		Married	Divorced	Total	
Non-anticip.	Primary	371	71	442	16
	Secondary	433	55	488	11
	Post-secondary	116	21	137	15
	Subtotal	920	147	1067	14
Anticipatory	Primary	—	—	—	—
	Secondary	66	28	94	30
	Post-secondary	120	31	151	21
	Subtotal	186	59	245	24
Total		1106	206	1312	16

framework. Using anticipatory covariates in fact corresponds to “back-dating” the times of highest educational achievement, τ_k , for a number of individuals. For individual k such that $\tau_k > \tau_k^0$, $x_k(\tau_k^0)$ is set to $x_k(\tau_k)$. In Figure 3 this corresponds to moving the marks for the individuals below the diagonal reference line to the left. After these manipulations, the analysis is carried out as if $\tau_k \leq \tau_k^0$ for all k .

Another alternative to the Bayesian correction scheme and the anticipatory approach is to limit the analysis to the respondents for which we have completely observed the relevant exposure times, only the individuals k for which $\tau_k \leq \tau_k^0$. These individuals are found on or above the diagonal in Figure 3. Although this seemingly leaves data unaltered in the sense that we do not manipulate covariate values or data but only reduce the set of respondents, it is unclear what type of systematic errors this might create beyond the obvious loss of information. We call this the *reduced data* approach (“available-case” analysis, Little and Rubin, 1987).

Neither the anticipatory manipulations nor the Bayesian correction change the observed marginal occurrences D_{i+} and exposures T_{i+} . For the reduced data approach these margins are reduced due to the reduction of the total number of respondents. For the anticipatory approach and the Bayesian correction, the marginal total number of occurrences in age-of-marriage group i and the corresponding exposure time are the same.

The time at which the highest educational level is achieved is irrelevant as long as it precedes the time of marriage. However, the time of highest educational level is relevant for calculating exposure times whenever it occurs after the time of marriage.

5.4. Results

Figure 4 contains estimates of baseline and relative risks of divorce across the three models. The upper-right panel shows that the relative risk of divorce for people with post-secondary educational level (α_3) is overestimated when we use reduced data or anticipatory version of the covariate education. The extent of overestimation is more pronounced in the case of reduced data. Similarly, the estimate of the second marriage duration (β_2) is underestimated in reduced data. Point estimates of this estimation, which corresponds to the expectations with respect to the posteriors in Figure 4 are given in Tables 2 and 3. The point estimates of λ_{ij} in Table 4 are obtained as the MCMC

FIGURE 4 Posterior distributions for model parameters. Point estimates and 95% highest posterior density intervals for reduced data (circle), anticipatory data (triangle), and the Bayesian covariate model (square).

estimators of $E(\beta_i \alpha_j \mid \{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y)$

$$\frac{1}{G} \sum_{g=1}^G \beta_i^{(g)} \alpha_j^{(g)}, \tag{44}$$

and correspondingly for $E(\beta_i/\beta_{i*} \mid \{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y)$. This is not the same as taking products or ratios of the corresponding point

TABLE 2 Estimates of Relative Risks α_j (Relative to α_1) for Bayesian Covariate Model, Anticipatory Analysis, and Reduced Data Analysis

	Highest educational level		
	Primary	Secondary	Post-sec.
Bayesian corr.	1	0.98	1.07
Anticipatory	1	1.13	1.37
Reduced	1	0.96	1.58

TABLE 3 Estimates of Baseline Risks β_i (Relative to β_1) for Bayesian Covariate Model, Anticipatory Analysis, and Reduced Data Analysis

Marriage duration	Bayesian correlation	Anticipatory	Reduced
0–1	1.00	1.00	1.00
1–2	1.83	1.83	0.92
2–3	2.19	2.21	2.10
3–6	2.70	2.73	2.37
6–6+	2.08	2.11	1.87

estimates, a procedure that does not take the inter-dependencies between parameters into consideration.

In Tables 2 and 3 α_3 is overestimated by 51% (1.58 vs 1.07) while β_2 is underestimated by 91% (0.92 vs 1.83) in the model with reduced data compared with the Bayesian adjusted model. The overestimation of α_3 reflects that values D_{ij}/t_{ij} in Table 5 are lower in the Bayesian adjusted procedure than in the reduced analysis for $j = 3$ and at all durations i . In the reduced analysis, it follows from the missing data mechanism that individuals who were divorced and had the highest educational level, $j = 3$, were never exposed to being married at educational levels $j = 1, 2$ (also compare the figures 11% and 15%

TABLE 4 Estimates of λ_{ij} (Relative to $\lambda_{1,1}$) from Bayesian Covariate Model, Anticipatory Analysis, and Reduced Data Analysis

		Highest educational level		
Marriage duration		Primary	Secondary	Post-sec.
0–1	Bayes (B)	1.00	0.98	1.07
	Anticipatory (A)	1.00	1.13	1.37
	Reduced (R)	1.00	0.96	1.58
1–2	B	1.83	1.79	1.96
	A	1.83	2.08	2.50
	R	0.92	0.88	1.45
2–3	B	2.19	2.13	2.33
	A	2.21	2.51	3.02
	R	2.10	2.01	3.33
3–6	B	2.70	2.63	2.88
	A	2.73	3.10	3.73
	R	2.37	2.27	3.76
6–6+	B	2.08	2.03	2.22
	A	2.11	2.40	2.89
	R	1.87	1.80	2.99

TABLE 5 Occurrences and Exposures by Levels of Highest Achieved Educational Level and Marriage Duration for a Sample of Swedish Men (for the Bayesian covariate model the posterior expectations for occurrences and exposures are used)

Mar. dur.	Edu. lev.	D_{ij}			t_{ij}		
		Bayes	Anticip.	Red.	Bayes	Anticip.	Red.
0–1	1	4.9	4	4	573	437	437
	2	4.1	4	4	576	570	476
	3	0.0	1	0	143	286	135
	Total	9.0	9	8			
1–2	1	5.7	2	2	539	420	420
	2	6.3	5	2	533	534	441
	3	2.0	7	2	154	273	126
	Total	14.0	14	6			
2–3	1	9.4	7	7	509	407	407
	2	3.6	5	3	503	508	418
	3	3.0	4	3	159	257	115
	Total	16.0	16	13			
3–6	1	22.7	16	16	1,351	1,130	1,130
	2	22.3	22	16	1,306	1,327	1,076
	3	8.0	15	7	481	681	285
	Total	53.0	53	39			
6–6+	1	47.4	42	42	3,842	3,631	3,631
	2	45.8	47	30	3,404	3,360	2,542
	3	21.8	25	9	1,534	1,788	498
	Total	114.0	114	81			

in Table 1). The underestimation of β_2 in the reduced analysis is due to the fact that many observed occurrences in the second time interval, $i = 2$, pertain to individuals who attained their highest educational levels after they were married (8 of 14 as shown in Table 5).

The distribution of the respondents over different combinations of marriage duration and educational level is given automatically in the reduced data scheme and in the anticipatory analysis. The details of this are provided in Table 5. As the allocation of “uncertain” respondents is stochastic in the Bayesian approach, one way of investigating the allocation of these respondents is through posterior expected values, which are given in Table 5.

Moreover,

$$\tilde{D}_{ij} \equiv E_{\text{Bayes-corr}}(D_{ij} \mid \{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y) \geq D_{ij}^{\text{reduced}}, \tag{45}$$

where D_{ij}^{reduced} is the total number of occurrences calculated for all k such that $\tau_k \leq \tau_k^0$. Often $\tilde{D}_{ij} < D_{ij}^{\text{anticip}}$ because D_{ij}^{anticip} incorporates some cases k for which we know that divorce happened before τ_k , and consequently the corresponding $D_{i, x_k(\tau_k), k} = 0$ (and not $D_{i, x_k(\tau_k), k} = 1$ which is the case in the anticipatory analysis). This is reflected in the table where the overestimation of the total number of divorces for respondents with post-secondary education is well marked in the anticipatory analysis. This discrepancy is further illustrated by the point estimates of rates in Table 4. For instance, in Table 5, marriage group 3–6 contains 15 occurrences (divorces) with post-secondary education in the anticipatory model, but only 8 in the Bayesian adjusted model. The corresponding relative risks shown in Table 4 are 3.73 and 2.88.

As only one model is considered and as prior distributions were chosen on the basis on convenience rather than on substantive considerations, posterior predictive distributions are more appropriate for assessing model fit than Bayesian model selection procedures. We follow Gilks et al. (1996) and define the posterior predictive p -value as $\Pr\{K(y^{\text{rep}}; \theta) \geq K(y^{\text{obs}}; \theta) | y^{\text{obs}}\}$, where $K(\cdot; \theta)$ is a discrepancy measure and y^{rep} denotes a replicate observation from the posterior predictive distribution. For the piecewise exponential model it is natural to define the discrepancy in terms of the occurrences D_{ij} with a corresponding χ^2 -like statistic:

$$K(y; \theta) = \sum_{i,j} \frac{(D_{ij} - E(D_{ij} | \theta))^2}{\text{Var}(D_{ij} | \theta)}. \quad (46)$$

As described in Gilks et al. (1996), the corresponding p -value can be calculated as the MCMC mean of $\mathbf{1}_{\{K(y^{\text{rep},(g)}; \theta^{(g)}) \geq K(y^{\text{obs}}; \theta^{(g)})\}}$ over the posterior sample, where the replicate $y^{\text{rep},(g)}$ is generated from the model conditional on $\theta^{(g)}$. In our present case, the p -value was 0.59 indicating a good fit of the model to the data.

The marginal posterior distributions of the covariate model parameters are shown in Figure 5. As all absorbing states were assumed to be observed, the posterior distributions of the probabilities of staying at the primary school level (ϕ_1) and at the secondary school level (ϕ_2) were straightforward to obtain. In addition, because we employed “vague” priors, the two posteriors are centered tightly over the observed relative frequencies of stayers. Results not reported here indicate that the observed distribution of educational achievements is reasonably well reproduced by the covariate model defined by the parameters η_j and ζ_j in Figure 5.

FIGURE 5 Posterior distributions for covariate model parameters.

6. CONCLUSION

Anticipatory analyses explaining current behavior by future outcomes in life-course research are problematic because they do not follow the temporal order of events. Event history data collected at enormous cost lack history on important explanatory variables such as education and often social status. It is the investigator's responsibility to seek appropriate procedures to minimize, if not eliminate, the problems due to such errors in design before any attempt is made to estimate the parameters of interest. We addressed this issue and proposed analytic procedures.

Our specific problem has been that some individuals in our sample have achieved their reported highest educational level after marriage. They should have had lower educational level at the time of marriage than at the time of the survey. We have little idea of how much lower it should be but we do have information on the age of the individual and year at which he achieved the reported highest educational

level. Using the education variable causes biases in the estimated relative hazards but the strength and direction of this bias is unclear. The main goal of our investigation was to come up with numerical estimates of the direction and strength of this bias.

We proposed a Bayesian approach in order to use the available information. We specified a continuous-time Markov model for the incompletely observed time-varying anticipatory covariate and implemented standard Bayesian data augmentation techniques in estimating the adjusted baseline and relative risks.

Empirical findings for our case study indicate that the failure to account for the anticipatory nature of the covariate leads to overestimate the relative risks of divorce across educational levels. The extent of overestimation is higher among individuals with the highest educational level because these are overrepresented among those who completed their reported highest educational level after marriage.

ACKNOWLEDGEMENTS

A draft version of this paper was presented in Session 39 (Advances in Methodology) at the 71st Annual Meeting of the Population Association of America (PAA), Los Angeles, California, March 30–April 1, 2006. We are grateful to Prof. Thomas W. Pullum and Prof. Wiji Arulampalam for valuable comments at the meeting. The authors would also like to express their appreciation to the Editor-in-Chief Noël Bonneuil and anonymous referee for their substantive and editorial comments that improved the paper substantially.

REFERENCES

- Arulampalam, W. and Bhalotra, S. (2006). Sibling death clustering in India: State dependence versus unobserved heterogeneity. *Journal of the Royal Statistical Society–Series A (Statistics in Society)*, 169: 829–848.
- Breslow, N.E. and Day, N.E. (1975). Indirect standardization and multiplicative models for rates, with reference to the age adjustment of cancer incidence and relative frequency data. *Journal of Chronic Diseases*, 28: 289–303.
- Chen, H.Y. (2002). Double-semiparametric method for missing covariates in Cox regression models. *Journal of the American Statistical Association*, 97: 565–576.
- Faucett, C.L., Schenker, N., and Elashoff, R.M. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, 93: 427–437.
- Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, 6: 733–807.
- Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

- Hoem, J.M. (1987). Statistical analysis of a multiplicative model and its application to the standardization of vital rates: A review. *International Statistical Review*, 55: 119–152.
- Hoem, J.M. (1996). The harmfulness and harmlessness of using anticipatory regressor: How dangerous is it to use education achieved as of 1990 in the analysis of divorce risks in earlier years? *Yearbook of Population Research in Finland*, 33: 34–43.
- Hoem, J.M. and Kreyenfeld, M. (2006a). Anticipatory analysis and its alternatives in life-course research. Part 1: The role of education in the study of first childbearing. *Demographic Research*, 15: 461–484.
- Hoem, J.M. and Kreyenfeld, M. (2006b). Anticipatory analysis and its alternatives in life-course research. Part 2: Two interacting processes. *Demographic Research*, 15: 485–498.
- Holford, T.R. (1976). Life tables with concomitant information. *Biometrics* 32: 587–597.
- Kravdal, Ø. (2004). An illustration of the problems caused by incomplete education histories in fertility analyses. *Demographic Research*, S3: 135–154.
- Lagakos, S.W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, 35: 139–156.
- Lillard, L.A., Panis, C., and Upchurch, D. (1994). Interdependencies Over the Life Course: Women's Fertility, Marital, and Educational Experiences. Working Paper 94-17, Labour and Population Program, Santa Monica, CA: The RAND Corporation.
- Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley.
- Paik, M.C. and Tsai, W.Y. (1997). On using the Cox proportional hazards model with missing covariates. *Biometrika*, 84: 579–593.
- Rubin, D.B. (1976). Inference and missing data (with discussion). *Biometrika*, 63: 581–592.
- Schafer, J.L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- Tierney, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 22: 1701–1762.