Research Report

Department of Statistics

No. 2006:9

Bayesian Adjustment of Anticipatory Covariates in the Analysis of Retrospective Data

Gebrenegus Ghilagaber
Johan Koskinen

Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden

Bayesian Adjustment of Anticipatory Covariates in Analyzing Retrospective Data*

Gebrenegus Ghilagaber[†] and Johan Koskinen[‡]

*A draft version of this paper was presented in Session 39 (Advances in Methodology) at the 71^{st} Annual Meeting of the Population Association of America (PAA), Los Angeles, California, 30 March - 1 April, 2006. We are grateful to Prof. Thomas W. Pullum and Prof. Wiji Arulampalam for valuable comments at the meeting.

[†]Gebrenegus Ghilagaber (corresponding author), Department of Statistics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: Gebre@stat.su.se

[‡]Johan H. Koskinen, Swedish Institute for Social Research (SOFI), Stockholm University, Sweden; and Department of Psychology, School of Behavioural Science, University of Melbourne, Parkville Victoria 3010, Australia. E-mail: johank@unimelb.edu.au

Abstract

In retrospective surveys, records on important variables like respondent's educational level and social class refer to what is achieved by the date of the survey. Such anticipatory variables are then used as covariates in investigations of behavior - such as marriage and divorce - in life segments that have occurred before the survey-date. A question worth investigating in such situations is, thus, as to what extent any changes in the behavior under study can be attributed to misclassification of respondents across the various levels of the anticipatory variable; and to what extent they reflect real differences in the behavior across the levels. In this paper, we propose a Bayesian approach to address this important question and correct the biases in estimates of covariate effects due to the use of anticipatory variables. This is accomplished by specifying a continuous-time Markov model for the incompletely observed time varying covariates and then implementing standard Bayesian data augmentation techniques. The issues are illustrated by estimating effects of educational level on divorce-risks within the framework of a multiplicative piece-wise-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.

1. Introduction

Anticipatory covariates are variables whose value refers to what is attained by the date of the survey (interview); but are used to explain behaviour in life-course that took place before the survey. Highest educational level and social class at survey time are typical examples of anticipatory variables. Such variables are common in many (possibly most) retrospective studies simply because the data-collection procedure focuses on, say, 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 modeling some outcome variable like rates of marriage or divorce - events that took place long before the survey date. It is obvious that this causes a *time-inconsistency* problem because, in such cases, data that pertain to the date of the survey become less and less informative, the further from the date of survey is the date of the event of interest.

For instance, suppose education-level achieved by survey date is used as a covariate in modelling a demographic outcome (such as marriage, childbearing, or divorce) which took place before the survey. Educational progress is likely to occur between the time of the event (say entry into marriage) and the survey date. An interesting question would then be as to what extent the changes in patterns of the phenomena being studied across educational levels can 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 behaviour of different levels of education. In other words, ignoring the anticipatory nature of such variables potentially produces biased estimates and, thus, incorrect conclusions with respect to the effect of the covariate on the phenomena under study.

While many investigators are aware of of this problem, our knowledge about the strength and direction of the biases in realistic situations is still too scanty. Among the few works on the topic, Hoem (1996) uses a case study on Swedish women and warns that 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. Alho (1996), on the other hand, uses simple linear regression framework to show that the magnitude of the bias introduced by the use of anticipatory covariates cannot be guaranteed to remain small. Kravdall (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 retrospective history of births and deaths, Arulampalam and Bhalotra (2003; 2006) suggest not to include any anticipatory (current-dated) regressors - household asset, toilet facility, electricity or access to piped water at the date of the survey - in their analyses. In our view, this may not be the optimal solution because much valuable information may be lost by totally ignoring such variables. This view is also shared by more recent works (Hoem & Kreyenfeld, 2006a; 2006b) who argue that anticipatory variables may provide some useful summary information, anyway. Further, they use empirical examples to address the issue and propose alternative strategy. They hold, however, that their alternative strategy is weak because it is based on unrealistic assumptions and that it fails to produce the summary information that is expected of anticipatory analysis. For a related problem 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 e.g. imputation.

In the present paper, we propose a more general framework to address this important problem and come up with bias-corrected estimates of effects of anticipatory covariates. We define a joint model for the out-come variable and the partially unobserved anticipatory variable and using Bayesian methods we draw inference based on the posterior distributions of model parameters given what we have observed. The issues addressed are illustrated by modeling divorce risks among 1342 Swedish men born 1936-1964 in the framework of a piece-wise constant hazard model (Breslow and Day, 1975; Hoem, 1987). The results show that lack of proper account of anticipatory variables may seriously plague the analyses because the relative risks of divorce across the anticipatory educational level are overestimated.

The rest of the paper is organized as follows. In Section 2 we provide a brief introduction of the standard multiplicative model with piece-wise-constant baseline hazard. Sections 3 and 4 are devoted to a step-by-step description of our proposed Bayesian approach - in Section 3, we extend the piece-wise-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 proposed model is then fitted to a data set, in Section 5, and its performance is compared with that of a standard model which uses anticipatory covariates, as well as with one that is based on a reduced data set. Section 6 summarizes the contents of the paper and provides some concluding remarks.

2. The standard multiplicative two-factor hazard model

For a sample of individuals, consider J educational levels and let D_{ij} be the number of occurrences, say divorces, at marriage-duration i ($i = 1, ..., I$) for the j^{th} educational level ($j = 1, 2, ..., J$) for T_{ij} years of observed exposure to the risk (of divorce). Note here that the covariate indexed by i is the grouped-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

$$D_{i+} = \sum_{j=1}^{J} D_{ij}, \qquad D_{+j} = \sum_{i=1}^{I} D_{ij}, \tag{1}$$

and

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

and let T_{i+}, T_{+j}, and T_{++} represent similar quantities for the exposure variable T. Usually, divorce risks are assumed to be *piece-wise constant* in each of the the I time intervals but may vary between the intervals. In other words, the time to divorce is assumed to follow *piece-wise exponential distribution* for each educational level. Thus, the density function of the time to divorce in duration group i for a person k with educational level j is given by

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

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

A multiplicative (log-linear) model for the λ_{ij} arises when we express it as

$$\lambda_{ij} = \beta_i \alpha_j \tag{4}$$

whereby 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 has been suggested for many situations. A brief discussion of its merits has been given by Breslow and Day (1975) while Hoem (1987) reviews the statistical theory behind the model.

This model has $I + J - 1$ parameters in general (namely β_1, $\beta_2, ..., \beta_I$, and $\alpha_2, \alpha_3, ..., \alpha_J$), for α_1 is tied down by the normalization $\alpha_1 = 1$. In such formulation, α_j measures the relative super-/sub-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 standard (baseline) educational group ($j = 1$).

To construct the likelihood function when (4) holds, we first define D_{ijk} as an indicator variable of whether the k^{th} sample 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. Using (3) and (4) the contribution, to the likelihood, of the sub-sample of individuals in the i^{th} duration-group and having the j^{th} level of education can then be obtained as

$$\Lambda_{ij} = \prod_k \left[(\beta_i \alpha_j)^{D_{ijk}} \exp\left(-t_{ijk} \beta_i \alpha_j\right) \right] = (\beta_i \alpha_j)^{D_{ij}} \exp\left(-T_{ij} \beta_i \alpha_j\right) \qquad (5)$$

where

$$D_{ij} = \sum_k D_{ijk}, \quad \text{and} \quad T_{ij} = \sum_k t_{ijk}.$$

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

$$\Lambda = \prod_i \prod_j \Lambda_{ij} = \prod_i \prod_j \left\{ (\beta_i \alpha_j)^{D_{ij}} \exp\left(-T_{ij} \beta_i \alpha_j\right) \right\} \qquad (6)$$

As we can note from equation (6), the Maximum Likelihood estimates of the baseline hazards ($\widehat{\beta}_i$) and relative hazards ($\widehat{\alpha}_j$) are direct functions of the number of events (D_{ij}) and the exposure times (T_{ij}). Thus, misclassification of the events and/or exposure-times into wrong intervals or, most importantly, into wrong levels of the covariate - as is the case with anticipatory covariates - will lead to incorrect estimates of the model parameters. This could potentially ruin the purpose of the analysis. The method we propose below is intended to adjust for the anticipatory nature of covariates in order to reduce, if not eliminate, such serious errors in estimating the parameters β_i and α_j.

3. Adjusting Anticipatory Covariates Using Bayesian Analysis

In a classical approach, the parameters β_i and α_j in equation (4) are assumed to be unknown but fixed quantities to be estimated from data. In the present paper, 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 above (the model for data) based on a model for the covariates such that for any given realisation of the covariate model, the conditional model for data is consistent with that which was assumed for data when all covariate values were observed. Apart from defining the covariate proces, the next two sections introduce some necessary notation so that the underlying processes are in accordance with the piecewise formulation in (6).

Educational attainment may be viewed as time-varying and, thus, the covariate model comprises two key 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. Secondly, it seems plausible not to allow for a respondent to lower his educational level and hence the evolution of the educational attainment ought to be non-decreasing. These two elements rule out using approaches like that used in Faucett et al. (1998), where a missing covariate process was used to account for partially missing smoking status.

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 $\mathcal{J} = \{0, \ldots, J\}$. The elements of $\mathcal{J}$ can, for example, be different educational levels with J being the highest possible educational level. The process is modeled as being right-continuous and non-decreasing. Taking these two elements together we may consider the processes in terms of the union $\bigcup_{j=0}^{\max_t X_k(t)} \Delta_j^*$ of distinct non-overlapping intervals

$$\Delta_j^* = [t_{j,0}, t_{j,1}) = \{t \in \Delta : X_k(t) = j\}, \tag{7}$$

where the lower and upper bounds are defined such that

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

respectively. *Conditional* on a realisation 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 non-decreasing. An example of such realisation 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, e.g. 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 marriage status 1. In order to define this process in more detail, we assume that Δ is partitioned into intervals Δ_i, $i \in \mathcal{I} = \{1, \ldots I\}$, representing for example different stages of marriage. Conditional on X_k we can form new intervals through the intersection of the intervals Δ_j^* and Δ_i

$$\Delta_{ij} = \Delta_i \cap \Delta_j^*.$$

The intervals are really functions of x_k but the notational dependency on k and the corresponding variable is supressed here. 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 non-recurring event and for each interval Δ_{ij}, we define the time to event t_{ijk}^* with exponential density

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

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

$$D_{ijk} = \mathbf{1}\left\{t_{ijk}^* < m(\Delta_{ij})\right\}.$$

Note that we do not allow explicitly for withdrawal without having experienced the event (Holford, 1976) at this point. We generally assume that the events are non-recurring, and denoting by $\partial\Delta_{ij}$ and $\partial(i, j)$ respectively, the

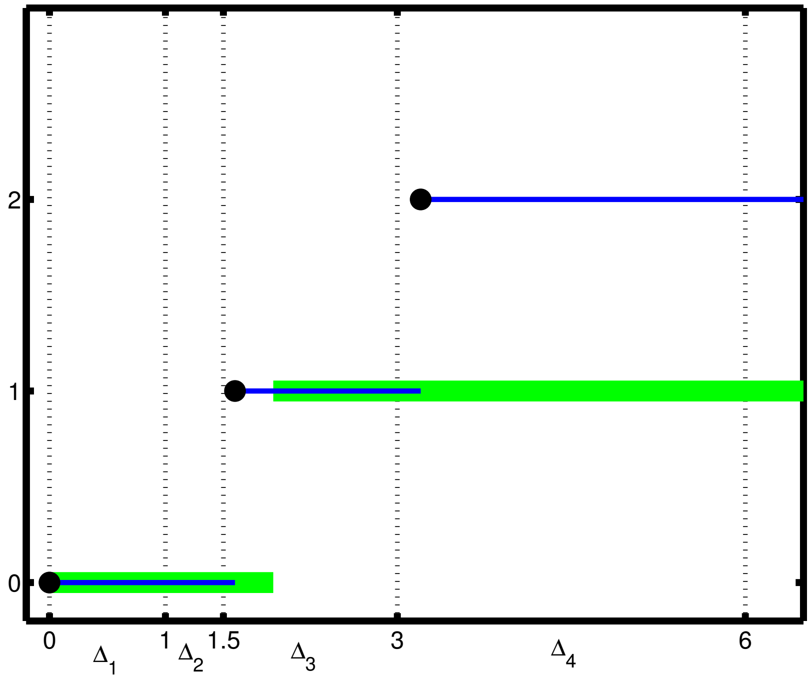

Figure 1: A sketch of the underlying processes (covariate process $x_k(t)$, thin line, response variable process, $y_k(t)$, thick line)

interval $\Delta_{uv} = [t_a, t_b)$ and pair (u, v), such that t_b is the lower bound of Δ_{ij}, we have

$$\Pr\big(D_{ijk} = 0 | D_{\partial(i,j)k} = 1\big) = 1 \tag{8}$$

and

$$\begin{aligned} & \Pr\big(D_{ijk} = 1 | D_{\partial(i,j)k} = 0, D_{\partial(\partial(i,j))k} = 0, \ldots, D_{0,0,k} = 0\big) \quad (9) \\ = \; & \Pr\big(t^*_{ijk} < m(\Delta_{ij})\big) \\ = \; & 1 - \exp\left(-m(\Delta_{ij})\lambda_{ij}\right). \end{aligned}$$

Equation (8) follows from the assumption that the process $Y_k(t)$ is non-reccurring and equation (9) follows from the properties of the exponential distribution. We also have the alternative characterization

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

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

$$t_{ijk} = \min\left\{t^*_{ijk}, m(\Delta_{ij})\right\}.$$

The process $\{Y_k(t), X_k(t)\}_{t\in\Delta}$ completely determines $\{D_{ijk}, t_{ijk}\}$, and hence, given X, $\{D_{ijk}, t_{ijk}\}$ completely determines $\{Y_k(t)\}_{t\in\Delta}$. An example of a hypothetical 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. (7), we can plot the process in terms of $\{D_{ijk}, t_{ijk}\}$ in a *Lexis* diagram, as shown in Figure 2.

In summary, for a given realisation x_k, we have a piecewise exponential hazard rate model with hazard rate λ_{ij} for each interval Δ_{ij} and density given by

$$p(y|\lambda, x) = \prod_{i,j} \lambda_{ij}^{D_{ijk}} \exp\left\{-\lambda_{ij} t_{ijk}\right\} \tag{10}$$

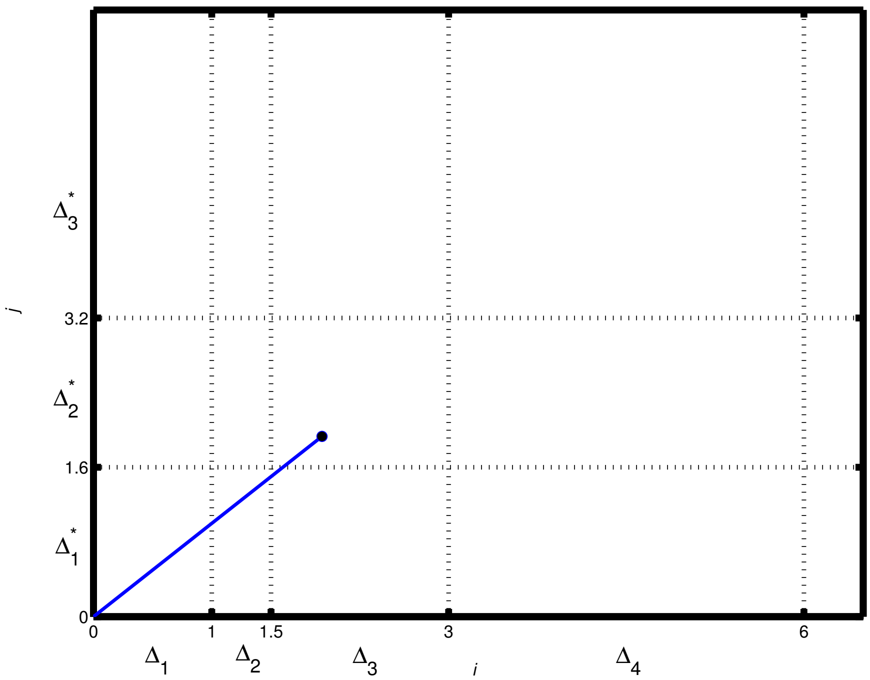

Figure 2: The Lexis diagram of the conditional process of $\{D_{ijk}, t_{ijk}\}$ (marriage duration intervals Δ_i, and educational level intervals Δ_j^*)

where the product is taken over relevant intervals (i, j) so that respondents who have experienced the event are removed since they are no longer at risk of experiencing the event of interest. For a sample of independent observations $\{D_{ijk}, t_{ijk}\}$, we note that products over k in equation (10) reduces to the likelihood given in equation (6).

3.2 Modeling the covariate process

We now proceed to suggest a simple model for the covariate process. As will be discussed below, this particular covariate model, in its simplest form, is rarely adequate. Here it serves as a reference for the presentation of the inference proceedure and additional elements may easily be added later.

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 analysing the dependent variable process and the covariate process. It is for example hard to say exactly how the approximation that imputation entails would affect hypothesis testing. A consistent approach opens for probabilistic model selection. In addition, available information on the covariates is incorporated into the joint model and subsequently updated in the process of estimating the model parameters from the data. If the uncertainty that stems from not having fully observed the covariate values is neglected the uncertainty about the model parameter estimates will be 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 model 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 $s_{1k}, \ldots, s_{x_k(\tau_k^1)k}$. 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, i.e. the duration spent in each educational level, via an exponential distribution, we are *adjusting the exposure times to divorce at each educational level.* Given the vector of parameter values $\eta = (\eta_1, \ldots, \eta_J)^T$, the density of x_k is then given by

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

where

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

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 (attain the highest educational level). This is highly unrealistic and it could also potentially introduce an unnecessary noise. For instance, if some individuals reach their highest level of education very early and this educational level is relatively far from the absorbing state J, the observed time these individuals remained in this educational state should not, in principle, contribute any information to the times between transitions since we know these individuals are not going to make another transition. In order to deal with this issue, we need to introduce yet another unobserved latent variable - an indicator of whether an individual's reported educational level is "terminal" or not. In our empirical illustration, this is achieved via simple Bernoulli trials though it is possible to use other models like the sequential probit model.

4. Inference in the adjusted model

4.1 Posterior distributions

In addition to the two components presented above (a covariate model and a model for data), we need to specify prior distributions for the unknown parameters in order to 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

$$\mathbf{1}\{x_k(t) \overset{\tilde{\Delta}^k}{\simeq} x_k\} = \min_{t\in\tilde{\Delta}^k}\{\mathbf{1}\{x_k(t) = x_k(\tau_k)\}\},$$

indicate whether $\{x_k(t)\}$ is concordant with the observation $x_k(\tau_k)$, i.e. the educational level as described by $\{x_k(t)\}$ does not change after k has acheived his or her highest educational level. Assuming that λ and η are independent a priori, with prior distributions $\pi(\lambda)$ and $\pi(\eta)$, respectively, 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

$$\pi(\lambda)\pi(\eta)\prod_{k\in U} p(y_k|\lambda, x_k)p(x_k|\eta)\mathbf{1}\{x_k(t) \overset{\tilde{\Delta}^k}{\simeq} x_k\}. \tag{11}$$

Before we address the inferential issues related to the above model (the particular way our data and covariates are combined), we outline, below, the main ideas in Markov chain Monte Carlo (MCMC) methodology.

4.2 An outline of the inference scheme

Exact inference for model variates 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 (11) is only specified up to an analytically intractable multiplicative constant. We may however use MCMC to simulate from this distribution (see e.g. Gilks, Richardson & Spiegelhalter, 1996, for an introduction to Markov chain Monte Carlo techniques). More specifically, we are able to write the full conditional posterior distributions

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

for each block θ_s of posterior variates $\theta_1, \ldots, \theta_S$ - at least up to a multiplicative constant. The subscript $-s$, is henceforth used as a notational shorthand referring to all other indices except s. Here $\theta_1, \ldots, \theta_S$ include both parameters λ, and η, as well as X.

This means that we may produce a sequence of draws $\{\theta_1^{(g)}, \ldots, \theta_S^{(g)}\}_{g=0}^{G}$, by circling through all the S coordinates, sequentially performing draws

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

for each $g > 0$.

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

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

in which case 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

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

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

The resulting sequence of draws $\{\theta_1^{(g)}, \ldots, \theta_S^{(g)}\}_{g=0}^G$ is a sample from the joint posterior distribution of $\theta_1, \ldots, \theta_S$ given $Data$. Naturally, this is guaranteed only as G tends to infinity but the generated sample may for all intents and purposes be treated as an "exact" sample from the posterior distribution with the usual caveats regarding convergence and autocorrelation (for details see Gilks et al., 1996; Tierney, 1994). This means that point estimates of any relevant quantity $g(\theta_1, \ldots, \theta_S)$ can be approximated by the corresponding ergodic mean of this quantity over the sample, $\bar{g}(\theta_1, \ldots, \theta_S) = G^{-1} \sum g(\theta_1^{(g)}, \ldots, \theta_S^{(g)})$.

4.3 The covariate process and its parameters

4.3.1 Exposure times and educational careers

Recall that the number of intervals of the type $\Delta_j^{*,k}$ in (7) is fixed at $x_k(\tau_k^1) = x_k(\tau_k)$. However, once the highest achieved educational state is recorded, the number of intervals constructed from Eq. (7) varies with the location of the latent transitions in the covariate process. More specifically, for some $(i,j) \in \mathcal{I} \times \mathcal{J}$, we may have $\Delta_{ij}^k = \emptyset$, which may cause a conceptual difficulty in defining the exposure time t_{ijk}. To deal with this we apply the convention that whenever $\Delta_{ij}^k = \emptyset$, we set $t_{ijk} = m(\Delta_{ij}^k) = 0$, and $D_{ijk} = 0$. These "structural zeros" may be thought of as corresponding to the rectangles in the Lexis diagram that are never entered by the process (see e.g. $\Delta_{2,2}$ in Figure 2).

In actual empirical fitting of the model to data we have to make a distinction between two cases: (i) where the exact point in time when the highest reported educational level is observed or (ii) where it is not observed.

In the first case,

$$x_k(\tau_k) = j$$

implies that

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

In the second case (when the exact time of transition is not observed), we have

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

but

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

In what follows, the main focus will be on the first case.

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 of x may be treated separately for each individual k. In general terms, the fully conditional posterior of x_k is proportional to

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

As mentioned earlier, the sequence $\{X_k(t) : t \in \Delta\}$ is completely determined by its holding times $s_k = (s_{1k}, \ldots, s_{x_k(\tau_k^1)k})$, and thus, in principle, we need only be able to generate sequences s_k of holding times from some density q, to be able to implement a Metropolis-Hastings up-dating step for the unobserved educational careers. Since we know that $w(x_k; y_k, \lambda, \eta, \{x_k(t)\}_{t\in\tilde{\Delta}^k}) >$

0 for sequences that are 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, we may, given s_k with the implied sequence x_k in the current iteration, propose a move to s'_k with the implied sequence x'_k, drawn from the rescaled Dirichlet distribution with density

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

for the case (i), such that $\sum s'_{jk} = \tau_k$, and where the parameters $\nu_1, \ldots, \nu_{x_k(\tau_k)}$ may or may not depend on η or s_k.

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

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

For case (ii), we need an additional holding time for the time that has elapsed from $\min(\Delta^*_{x_k(\tau_k)})$ until τ_k.

If the drawing of the latent educational careers is sensitive to the choice of proposal distribution in the sense that too many proposed moves are rejected we may introduce dependency between the current state of the latent educational career and the proposed educational career by "centering" the proposal distribution over the current state. Since 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 Dirichlet(ν), 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 will be centered over s_k and the ratio of proposal densities is given by

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

4.3.3 Conditional posterior distribution of the parameters in educational career

Since η and λ were independent a priori, the only relevant information for updating the covariate model parameters is provided by the realisations on the educational careers. This is reflected in the fully conditional distribution of η given all the rest, which is proportional to

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

where $n_j = \sharp\{k \in U : x_k(t) = j$, for some $t \in \Delta^k\}$, and $s_{j+} = \sum_{k \in U} s_{jk}$. If η_j has conjugate prior that is gamma(γ_1^j, γ_2^j), then the fully conditional posterior of η given the rest is gamma$(\gamma_1^j + n_j, \gamma_2^j + t_+^j)$.

4.4 Conditional distributions for the model parameters

So far we have not dealt with the parameters of interest for the research question. This part of the inference procedure is straightforward and it is the way a Bayesian inference scheme would be implemented had we had completely observed the educational careers (at least the relevant part that follows the onsets, τ_k^0). Hence, this portion of the updating scheme only requires that we have classified, correctly or using the anticipatory approach, 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.

It is worth noting at this point that while the comprehensive model for data and covariates may seem complex, the part of the inference procedure involving the model parameters is extremely straightforward and only involves standard distributions.

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

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

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

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

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

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

The conjugate character of this part also illustrates nicely the influence of the prior distributions. The shape parameter for β_i in the fully conditional posterior consists of a part that represents our prior knowledge, $\tilde{\beta}_i^{(1)}$, and a part that represents data, D_{i+}. The latter is simply a count of the number of individuals that have experienced the event in age-interval i. A hyper-parameter $\tilde{\beta}_i^{(1)}$ equal to 1 thus roughly 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), as mentioned previously, the same educational state is assumed to be terminal for all individuals although some individuals may never proceed to the highest educational state. In other words, 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 individual $k \in U$ stays in educational state $x-1$ is gamma(ζ_x, η_x) (instead of exponential). The second problem may be 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 may then be written as

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

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

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

where

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

The fully conditional posterior of the state parameters is proportional to

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

where

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

Consequently, when we a priori let ϕ_j be independent beta$(\tilde{\phi}_j^1, \tilde{\phi}_j^2)$, for $j = 1, \ldots, J-1$, 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 in between educational progress, ζ and η, is not as straightforward as before when we had exponential holding times. In the present case, it is given by

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

where n_j is as defined before, and where

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

with the sum taken over all cases that are not long-term survivors.

Below, we illustrate the above models with empirical data divorce risks among Swedish men born 1936-64.

5. Empirical illustration: effect of education on divorce risks

5.1 Background

In the study of impacts of educational level attained by the time of marriage on the risks of divorce based on retrospective survey data, we may safely argue that some of the respondents have improved their educational-level between the time of marriage and the date of interview. Therefore, at each recorded educational level (except the lowest one) there are some individuals who, at the time of marriage, belonged to a lower level. If it is found that those with lower education have lower risks of divorce, then it is plausible to suspect that the empirical risks of divorce corresponding to the recorded educational-levels are overestimates of 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 of which are unknown to us. The aim of the illustration in this section is, therefore, to shed light on this issue and come up with numerical estimates of the degree of overestimation.

5.2 The Data Set

The data set used in this illustration is an extract from the 1985 Mail Survey of Swedish men, where information was collected on background variables as well as detailed retrospective history related to family dynamics (entry into and exit from marital and non-marital unions, as well as childbearing). The entire data set consisted of over 3000 men but our present illustration is based on 1312 ever-married me who were either divorced or still married by the survey time. These are men who have been exposed to the risk of divorce, i.e. individuals who have at some point been married and have complete and reliable values on the variables of interest. A distribution of these men across the ages at which they married and at which they reported to have completed their recorded highest educational level, is shown in Figure 3. As is clearly shown in the Figure, an appreciable amount of the men (those below the diagonal) have completed the attained highest educational level after they have married (have anticipatory values on the education variable). Anticipatory analysis - the tradition that is common in the analysis of such

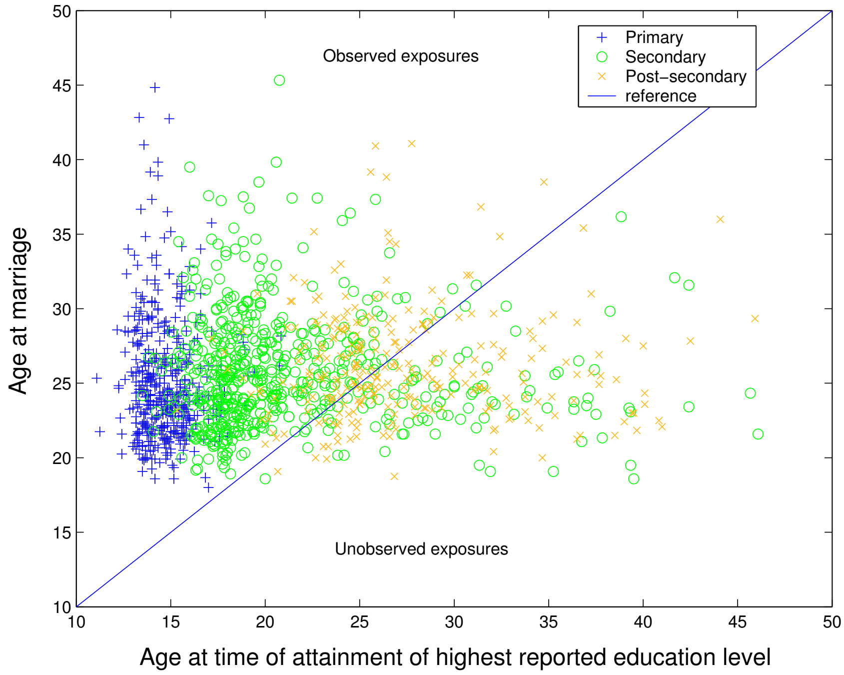

Figure 3: The age at which the highest educational level was achieved plotted against the age at marriage for a sample of Swedish men

type of data - amounts to moving these values (those below the diagonal in Figure 3) to the left, all the way until the diagonal reference line.

Further, a cross tabulation of the sample (see Table 1) shows differentials in percentage divorced across the anticipatory status of education. Investigating the role of misclassification on such differentials is, therefore, a worthwhile effort.

5.3 Models

The time variable (duration of marriage in years) has been categorized into five intervals: 0–1, 1–2, 2–3, 3–6, and 6^+ years. The lowest educational level (primary-level) was set as a base-line level and, thus, its corresponding

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
	Sub total	920	147	1067	14
Anticipatory	Primary	–	–	–	–
	Secondary	66	28	94	30
	Post-Secondary	120	31	151	21
	Sub total	186	59	245	24
Total		1106	206	1312	16

Table 1: Summary of data structure for the sample of Swedish men

parameter (relative hazard) was set to $\alpha_1 = 1$. Vague gamma$(1/1000, 1/1000)$ priors were assumed for all parameters, except for ϕ, the coordinates of which were assumed to be a priori independent beta$(1, 1)$, i.e. 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 framework. While this, at least superficially, may seem like not altering data, using anticipatory covariates in fact corresponds to "back-dating" the times of highest educational achievement, τ_k, for a number of individuals. More specifically, for individuals k such that $\tau_k > \tau_k^0$, when $x_k(\tau_k^1)$ is used anticipatory, $x_k(\tau_k^0)$ 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, all the way to the reference line. After these manipulations, the analysis is carried out as described above, as if $\tau_k \leq \tau_k^0$ for all k.

Yet another alternative to the Bayesian correction scheme and the anticipatory approach would be to limit the analysis to the respondents for which we have completely observed the relevant exposure times, i.e. 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 not immediately clear what type of systematic errors in the analysis this might create beyond the obvious loss of information. We may call this the *reduced data* approach (c.f. "available-case" analysis, Little and Rubin, 1987, sec. 3.3).

It is to be noted that neither the anticipatory manipulations, nor the Bayesian correction, changes the observed marginal occurrences D_{i+} and exposures T_{i+}. For the reduced data approach these marginals are reduced due to the reduction of the number of respondents but for both the anticipatory approach and the Bayes-correction the marginal number of occurrences in age-of-marriage group i and the corresponding exposure time is the same. How these are distributed among educational categories is shown in Table 5.

As may be understood from the description of the hazard rate model implemented here, 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 we have fitted. 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

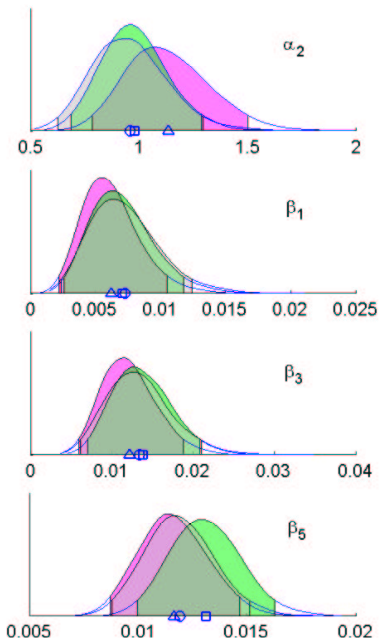

Figure 4: Posterior distributions for model parameters. Point estimates and .95 Highest posterior density intervals for reduced data (circle,gray), anticipatory data (triangle,red), and the Bayesian covariate model (square,green)

	Highest educational level		
	Primary	Secondary	Post-sec.
Bayesian corr.	1	.98	1.07
Anticipatory	1	1.13	1.37
Reduced	1	.96	1.58

Table 2: Estimates of Relative risks α_j (relative to α_1) for Bayesian covariate model, Anticipatory, and Reduced

of reduced data. Similarly, the estimate of the second marriage duration (β_2) is underestimated in reduced data. Point estimates of these over/under estimation, which correspond 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 estimators of $E[\beta_i\alpha_j|\{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y]$

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

and correspondingly for e.g. $E[\beta_i/\beta_{i*}|\{x_k(t) : t \in \tilde{\Delta}^k, k \in U\}, y]$. Note that this is not the same as taking products or ratios of the corresponding point estimates (a procedure that does not take the interdependencies between parameters into consideration).

In Tables 2 and 3 we note, among others, that 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 as compared with the Bayes-adjusted model.

As mentioned above, the distribution of the respondents over different combinations of Marriage duration and Educational level is given automatically in the reduced data scheme and when using anticipatory analysis. The details of this are provided in Table 5. Since the allocation of “uncertain” respondents is stochastic in the Bayesian approach, one way of investigating

Marr.dur.	Bayesian corr.	Anticipatory	Reduced
$0-1$	1	1	1
$1-2$	1.83	1.83	.92
$2-3$	2.19	2.21	2.10
$3-6$	2.70	2.73	2.37
$6-$	2.08	2.11	1.87

Table 3: Estimates of Baseline risks β_i (relative to β_1) for Bayesian covariate model, Anticipatory, and Reduced

		Highest educational level		
Marriage duration		Primary	Secondary	Post-sec.
$0-1$	Bayes (B)	1	.98	1.07
	Anticipatory (A)	1	1.13	1.37
	Reduced (R)	1	.96	1.58
$1-2$	B	1.83	1.79	1.96
	A	1.83	2.08	2.50
	R	.92	.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
$6-$	B	2.08	2.03	2.22
	A	2.11	2.40	2.89
	R	1.87	1.80	2.99

Table 4: Estimates of λ_{ij} (relative to $\lambda_{1,1}$) from Bayes covariate model, Anticipatory, and Reduced

the allocation of these respondents is via posterior expected values. These posterior expected values are also given in Table 5. It may be noted that

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

where D_{ij}^{reduced} is the number of occurrences calculated for all k such that $\tau_k \leq \tau_k^0$. It is also given that often $\tilde{D}_{ij} < D_{ij}^{\text{anticip}}$ since D_{ij}^{anticip} incorporates some cases k for which we know the time of divorce is 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 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 occurences (divorces) with post-secondary education in the anticipatory model, but only 8 in the Bayes-adjusted model. The corresponding relative risks as shown in Table 4 are 3.73 and 2.88, respectively.

The marginal posterior distributions of the covariate model parameters are shown in Figure 5. Since we assumed that all the absorbing states were 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, since 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 acheivements is reasonably well reproduced by the covariate model defined by the parameters η_j and ζ_j in Figure 5.

6. Summary and Concluding Remarks

It is obvious that inference procedures that attempt to explain current behaviour by future outcomes (anticipatory analyses) in life-course research

		D_{ij}			t_{ij}		
Mar. dur	Edu. lev.	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	1	0	143	286	135
	Total	9	9	8			
$1-2$	1	5.7	2	2	539	420	420
	2	6.3	5	2	533	534	441
	3	2	7	2	154	273	126
	Total	14	14	6			
$2-3$	1	9.4	7	7	509	407	407
	2	3.6	5	3	503	508	418
	3	3	4	3	159	257	115
	Total	16	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	15	7	481	681	285
	Total	53	53	39			
$6-$	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	114	81			

Table 5: Occurences 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)

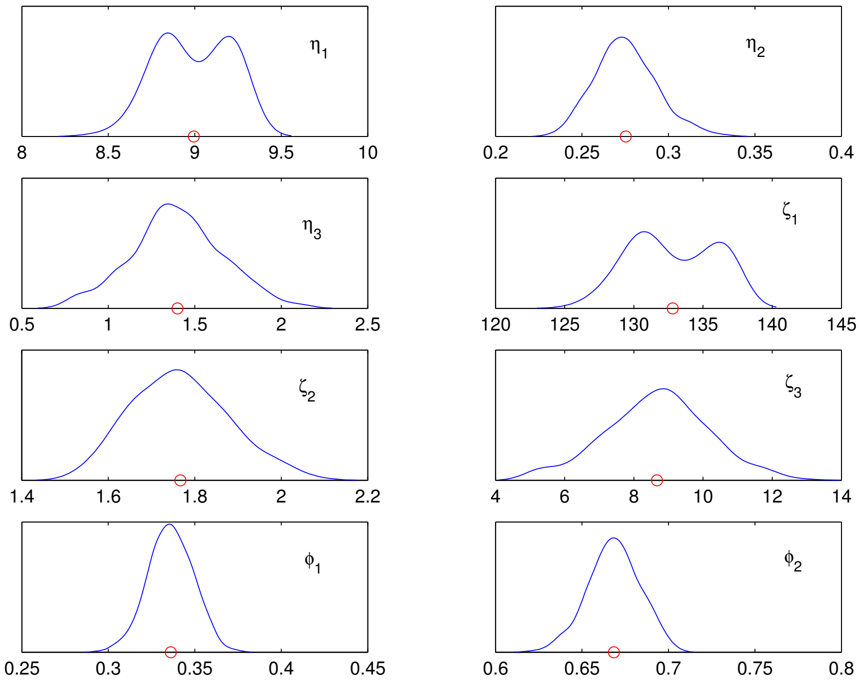

Figure 5: Posterior distributions for covariate model parameters

are problematic because they don't follow the temporal order of events. On the other hand, due to practical reasons, it is more often than not that event history data collected at enormous cost lack history on important explanatory variables such as education and social class. It is then the investigator's responsibility to seek appropriate procedures to minimise, if not eliminate, the problems due to such errors in design before any attempt is made to estimate the parameters of interest. The primary aim of this paper has been to address this issue and propose appropriate analytic procedures.

Our specific problem has been that some individuals in our sample have achieved their reported highest educational level after they have married. Thus, they should have had lower educational level at the time of marriage than what they reported at the time of the survey. We have little or no idea as to how much lower it should be but we do have information on the age of the individual and year he achieved the reported highest educational level. We know that using the education variable as it is causes biases in the estimated relative hazards but the strength and direction of this bias was unclear. The main goal of our investigation has, therefore, been to come up with numerical estimates of the direction and strength of such bias.

To achieve our task, we proposed a Bayesian approach in order to makes use of available information. This is accomplished by specifying a continuous-time Markov model for the incompletely observed time-varying anticipatory covariate and implementing standard Bayesian data augmentation techniques in estimating the adjusted baseline and relative risks.

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

The work presented here is not, however, without limitations. As briefly mentioned earlier, educational career is likely to have long-time survivors, individuals that never proceed to the highest educational level. The standard approach for dealing with long-time survivors is by applying mixture models (e.g. Yamaguchi, 1992). In the case of Sweden, there is considerable prior knowledge about educational carers to draw on, not only in terms of the times individuals spend in different states but also in terms of determinants of educational progress. With this in mind a natural further elaboration of the covariate model would be to include additional covariates in modelling the educational careers.

A model that has proved useful for educational progress of individuals is the sequential probit model (Mare, 1979; 1980). This is a more realistic model for education and only a little more complex than the simplistic model presented here.

A little more complicated but still more realistic would be to model the bivariate process directly. For the purposes of the analysis of the present paper, modeling the events conditionally on education may not be a bad approximation but it seems plausible to assume that it is not only educational level that influences divorce risk but that the converse is also true (marital status influences educational progress). This approach shares some similarities with the way other investigators handled missing covariates in survival analysis with the accelerated failure time model; jointly modeling the failure time and covariates (see, for instance, Faucett, Schenker & Elashoff, 1998; Cho & Schenker, 1999; Meng & Schenker, 1999).

As mentioned before, the joint modelling of covariates and the response variable of interest facilitates a fully parametric approach to testing hypotheses and addressing the issue of model mis-specification. For differently specified hazard models the mappings of section 3.1 might, for example, have to be changed but the approach is still fairly general. Of interest could be to test whether the risk of divorce increases with marriage duration, $\beta_1 < \beta_2 < \cdots < \beta_5$, or if the risk is constant over marriage duration groups, $\beta_1 = \beta_2 = \cdots = \beta_5$. In other words, a host of questions and problems remains to be investigated in order to make full use of the potential of the fully Bayesian approach:

- the prior distributions that were used here were for example specified solelely from the point of view of convenience. A principled approach for testing hypotheses regarding data requires that more effort is put into specifying prior distributions.
- in the present work we have treated ordinal time-varying covariates but in some applications we may have interval-level time-varying covariates. A covariate model for these interval level covariates would differ substantially from the one presented here since it would probably not be possible to model the evolution of the covariates as conveniently in terms of holding times and jumps. There are several candidate models in standard statistical theory, such as various diffusion processes, but things may be complicated if we introduce a monotony akin to the one presented here.

With the above limitations in mind, however, we were able to make some important conclusions and comparisons. It is our hope that the findings in this paper make a modest contribution to the existing knowledge on how to handle anticipatory covariates and serve as a stimulation to investigators to reproduce the procedures we present here on their own data sets.

References

Alho, J. M. (1996). A note on the use of anticipatory covariates in event history analysis. *Yearbook of Population Research in Finland,* **33**: 328-332.

Arulampalam, W. and Bhalotra, S. (2003). Sibling death clustering in India: genuine scarring vs unobserved heterogeneity. *Discussion Paper* **No. 03/552,** Department of Economics, University of Bristol.

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.

Cho, M. and Schenker, N. (1999). Fitting the log-F accelerated failure time model with incomplete covariate data. *Biometrics*, **55**: 380–387

Cho, M., Schenker, N., Taylor, J. M. G., and Zhuang, D. (2001). Survival Analysis with Long-term Survivors and Partially Observed Covariates. *The Canadian Journal of Statistics,* **29**: 421-436.

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.

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 First Childbearing. *Demographic Research*, Volume 15, Article 16, pp. 461 - 484 (29 November 2006). Available online at http://www.demographic-research.org/volumes/vol15/16/. (accessed 26 May 2007)

Hoem, J. M. and Kreyenfeld, M. (2006b), Anticipatory Analysis and its Alternatives in Life-Course Research. Part 2: Two Interacting Processes. *Demographic Research*, Volume 15, Article 17, pp. 485 - 498 (29 November 2006). Available online at http://www.demographic-research.org/volumes/vol15/17/. (accessed 26 May 2007)

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*. Special Collection 3, Article 6 (pp. 135-154). Available online at http://www.demographic-research.org/special/3/6/S3-6.pdf (accessed 26 May 2007).

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. *Biometrics*, **35**: 139-156.

Little, R. J. A., and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: John Wiley.

Meng, X. and Schenker, N. (1999). Maximum likelihood estimation for linear regression models with right censored outcomes and missing predictors. *Computational Statistics and Data Analysis*, **29:** 471–483.

Mare. R. D. (1979). Social Background Composition and Educational Growth. *Demograph,* **16**: 55-71.

Mare. R. D. (1980). Social Background and School Continuation Decisions. *Journal of the American Statistical Association,* **75**: 295-305.

Tierney, L., (1994). Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22**: 1701-1762.

Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of permanent employment" in Japan. *Journal of the American Statistical Association*, **87**: 284–292.