Taylor & Francis
Taylor & Francis Group

Dynamic Bayesian adjustment of anticipatory covariates in retrospective data: application to the effect of education on divorce risk

Parfait Munezero 🔟 and Gebrenegus Ghilagaber 🔟

Department of Statistics, Stockholm University, Stockholm, Sweden

ABSTRACT

We address a problem in inference from retrospective studies where the value of a variable is measured at the date of the survey but is used as covariate to events that have occurred long before the survey. This causes problem because the value of the current-date (anticipatory) covariate does not follow the temporal order of events. We propose a dynamic Bayesian approach for modelling jointly the anticipatory covariate and the event of interest, and allowing the effects of the anticipatory covariate to vary over time. The issues are illustrated with data on the effects of education attained by the survey-time on divorce risks among Swedish men. The overall results show that failure to adjust for the anticipatory nature of education leads to elevated relative risks of divorce across educational levels. The results are partially in accordance with previous findings based on analyses of the same data set. More importantly, our findings provide new insights in that the bias due to anticipatory covariates varies over marriage duration.

1. Introduction

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

Such variables are common in many retrospective studies because the data collection focuses on, say, birth or employment histories but contain no history on educational careers or social class mobility. The use of anticipatory covariates causes problems in inference because they are treated as fixed variables although, in fact, they are inherently time-varying.

Consider, for instance, a retrospective survey where the interest is to investigate differentials in the risk of divorce across educational levels attained before marriage but the

available information is only respondents' highest educational level at the time of the survey. We have information on the age of individuals and the year at which they achieved the reported highest educational level. But, we have no idea of their educational level at the time of marriage on which we want to base our investigation. At least some should have had lower educational level at the time of marriage than what they have reported at the time of the survey but we have no idea of how much lower it was.

The main goal of this paper is to propose a method of restoring the temporal order of the educational level and, thereby, disentangle the effects due to misclassification of the educational level into wrong categories from those that are attributable to real differences in educational levels.

The two recent studies that attempted to adjust for the anticipatory nature of educational level [9,10] are based on the assumption of proportional risks (constant relative risks) of divorce for educational levels across marriage durations. In the present study, we propose a dynamic Bayesian adjustment model that (i) allows covariate effects to vary over time and (ii) adjusts for the bias due to misclassification of anticipatory educational levels.

Our proposed methodology allows for modelling jointly the risk of divorce and the process of educational level transitions. We use a reversed time Markov chain process to estimate the most probable educational level before exposure to the event of interest. These estimated values are, then, used to compute adjusted effects of the covariate on the event under investigation.

Results show that ignoring the anticipatory nature of education leads to increased educational gradients in divorce risks with significant difference between those with secondary and tertiary education. But, such difference disappears when we adjust for the anticipatory nature of education and we find no significant differences in divorce risks across educational levels.

In Section 2, we introduce the data set and the specific problem to be addressed. In Section 3, we present the proposed dynamic Bayesian adjustment model. In Section 4, we fit the model, and present our results and compare them with those from previous analyses of the same data set. We summarise our findings by way of discussion in Section 5.

2. The data set and our specific problem

2.1. The data set

The data set on which we base our illustration is a subset from a Mail Survey of Swedish men where data on attitudes towards children and family was collected from about 3200 men [17]. Apart from attitudes and family plans the survey also collected other information such as parental background as well as respondents' highest educational level attained at the time of the survey. We have chosen this data for our illustration because it was analysed in previous studies [9,10] with which we want to compare our results.

The usable records for our present case are 1312 ever married men who have either divorced before the survey date (events) or were still married (censored). The variable of interest for our present purpose is educational level and its effects on the risk of divorce. Table 1 presents the data classified by whether the individual completed his reported educational level before he married (the non-anticipatory cases) or after he married (anticipatory cases).

Table 1. Distribution of sample across when education was completed.

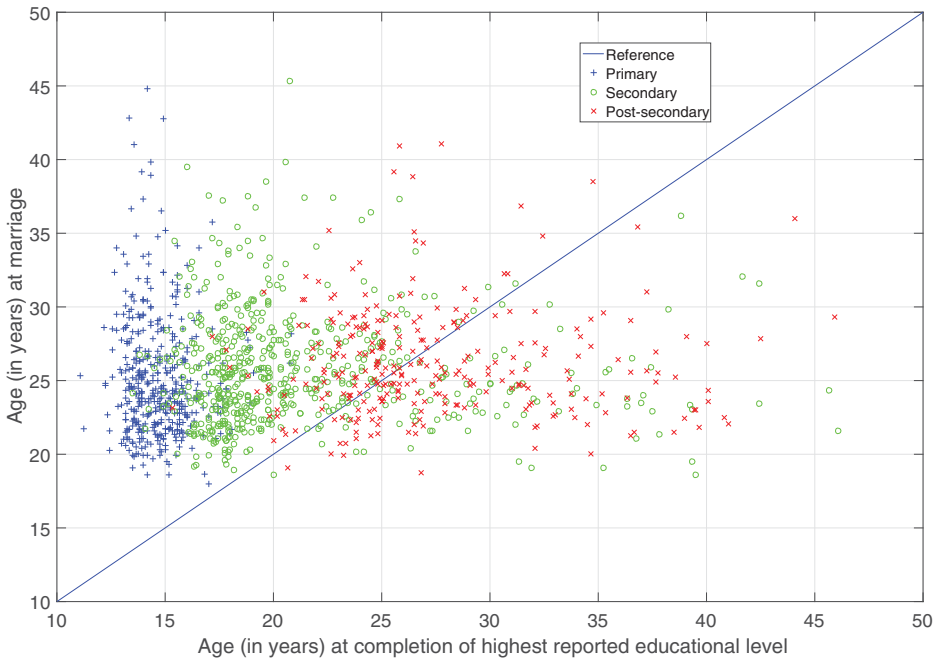| Completed	Education	Status			
		Still married	Divorced	Total	% divorced
Before marriage	Primary	371	71	442	16
(non-anticip.)	Secondary	433	55	488	11
	Post Secon.	116	21	137	15
	Sub-total	920	147	**1067**	14
After marriage	Primary	–	–	–	–
(anticipatory)	Secondary	66	28	94	30
	Post-Secon.	120	31	151	21
	Sub-total	186	59	**245**	24
	Total	1106	206	**1312**	16

Figure 1. Distribution of the data by age at marriage and at completion of reported education. The observations below the main diagonal are anticipatory cases.

Thus 245 of the 1312 respondents (close to 20%) are anticipatory cases in the sense that they completed their highest reported educational level after they married. These are also shown to the right of the diagonal line in Figure 1. As would be expected, all of those who reported to have primary level education at the time of the survey have completed this level before they married (indicated by blue '+' in Figure 1). For those who reported to have secondary or post-secondary education at the survey time, on the other hand, some have attained the reported educational level after they married. The exact figures are that 16% of those with secondary level education (the green 'o' in Figure 1), and 52% of those with post-secondary education (the red 'x' in Figure 1) are anticipatory cases.

Overall, 442 of the 1312 respondents (33.69%) have reported primary level education, 582 (44.36%) have reported they have secondary level education, while the rest 288

(21.95%) have reported to have achieved post-secondary-level education by the survey time.

It can also be noted that out of the entire sample of 1312 respondents (who had at least primary level education by the survey time), 870 (66.31%) progressed to secondary level education. Of these, 288 (33.10%) progressed further to post-secondary level education. These figures are relevant for comparison with our estimated transition rates between educational levels to be presented in Section 4.2.

2.2. The specific problem

Our specific problem is that some individuals have achieved their reported highest educational level after marriage and, in fact, a smaller proportion have completed the reported educational level after they divorced.

The question that remains to answer is what to do with these respondents whose educational level does not follow the temporal order of the event of interest. The simplest option is to discard them from further analyses as in [2]. We label this approach as the *reduced model* in the following sections but argue that it leads to loss of information and bias due to potential selection. Another option, which is the common practice in the literature, is to 'blind oneself' and proceed with analyses of the entire data set ignoring the anticipatory nature of the covariate. We call this alternative the *anticipatory model* in subsequent sections. We also argue that even this approach leads to biased estimates of educational gradients in divorce because potential changes in educational levels (between exposure to the risk and the event or censoring) are ignored. Thus we propose a model that uses a reversed time Markov chain process to estimate the most probable educational level before exposure to the risk of divorce. These estimated values are then used to compute adjusted effects of education on divorce risk. This approach is what we call the *adjusted model* in the rest of the paper. Other methods that require data imputation methods include [14–16,18] while [5,9,10] are examples of approaches based on joint modelling of the risk and the education process.

3. Dynamic Bayesian adjustment of anticipatory covariates

3.1. Dynamic hazards model

We observe data $D = (t_i, d_i, x_i, t_{mi}, t_{x_i})$, $i = 1, \ldots, n$, where $t_i = \min(T_i, C_i)$ represents the observed marriage duration for individual i which is either event time T_i (years between marriage and divorce) or a censoring time C_i (years between marriage and the survey date). The censoring indicator, d_i, takes the value 1 if the ith respondent was divorced by the survey time and the value 0 if he was still married (censored). The x_i is the highest education level reported at the time of the survey and have one of three levels (1 for primary level, 2 for secondary level and 3 for post-secondary level). The t_{mi} and t_{x_i} are, respectively, the dates the ith respondent married and completed his highest reported education.

We also introduce a variable Z which is the educational level achieved by the date of marriage. In the reduced and anticipatory models, $Z = x$, while in the adjusted model, Z is a latent variable whose value needs to be estimated together with other parameters of interest.

If Z is known and fixed, the marriage duration is modelled by assuming a continuous rate of divorce,

$$\lambda\left(t|Z_i\right) = \lambda_0\left(t\right)\exp\left(\beta_1(t)z_{i1} + \beta_2(t)z_{i3}\right), \tag{1}$$

where $\lambda_0(t)$, $\beta_1(t)$, $\beta_2(t)$ are functions of time representing, respectively, the baseline hazards and the effects of education levels on divorce risk at marriage duration t, $z_{il} = 1$ if $Z_i = l$, and $z_{il} = 0$ otherwise, for $l = 1, 3$. Here, secondary ($l = 2$) is set as the reference level.

One common approach of parametrizing time-varying covariate's effects is the semi-parametric piece-wise exponential (PE) model which expresses them using a Gaussian random walk process [6,12,23]. This is a special case of the spline representation described in [4,13,22].

The PE model partitions the time interval (in our case, marriage duration) into smaller intervals, $I_j = [\tau_{j-1}, \tau_j)$, $j = 1, \ldots, J$, where $\tau_0 = 0 < \tau_1 < \cdots < \tau_J = \max(T)$, and assumes that effect parameters and baseline hazard functions are constant within each interval I_j but can vary between intervals. Thus, for our case, we define $\lambda_0(t) = \lambda_{0j}$ (where $\lambda_{0j} > 0$) $\beta_1(t) = \beta_{1j}$ and $\beta_2(t) = \beta_{2j}$ for $t \in I_j$.

The continuous hazard function in Equation (1) is, therefore, discretized in terms of interval-specific hazards in the log-linear form:

$$\ln \lambda_{ij} = \theta_{0j} + \beta_{1j}z_{i1} + \beta_{2j}z_{i3} \tag{2}$$

$$= \boldsymbol{\theta}_j\mathbf{z}_i, \tag{3}$$

where $\mathbf{z}_i = (1, z_{i2}, z_{i3})'$, and $\boldsymbol{\theta}_j = (\theta_{0j}, \beta_{1j}, \beta_{2j})$ is the new regression parameter-vector within interval I_j and contains an intercept $\theta_{0j} = \ln(\lambda_{0j})$. The PE assumption allows to express the likelihood function in interval I_j as a product of the contributions of individuals in that interval [6]:

$$f_j\left(\mathbf{t}_j|\boldsymbol{\theta}_j, \mathbf{Z}_j\right) = \prod_{i=1}^{n_j} \lambda_{ij}^{d_{ij}} \exp\left(-\lambda_{ij}t_{ij}\right) \tag{4}$$

where λ_{ij} is defined in Equation (2), n_j is the number of men at risk in I_j (who are still married at the end of I_{j-1}), $t_{ij} = \min\{\tau_j, t_i\}$, d_{ij} is an interval-specific censoring indicator ($d_{ij} = 1$ if event occurs in I_j and $d_{ij} = 0$ otherwise), $Z_j = (Z_1, \ldots, Z_{n_j})$ and $t_j = (t_{1j}, \ldots, t_{n_j,j})$ are, respectively, the vectors of education profiles at marriage date and exposure times of men at risk in I_j.

As we can note from Equation (4) the likelihood functions in each interval 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 marriage intervals or, most importantly, into wrong levels of the covariate – as is the case with our anticipatory education – will lead to incorrect estimates of the parameters. This, in turn, can potentially ruin the purpose of the analysis.

Further, a plot of life-table estimates of the hazards of divorce for the three educational levels (see the supplemental material) shows that the hazards cross each other across marriage duration – thereby violating the proportional hazards assumption.

Below, we propose a method that both relaxes the proportional hazards assumption and adjusts for the anticipatory nature of the educational level.

3.2. Joint likelihood for the marriage duration and the latent educational level

To adjust for the anticipatory education level, we express the joint likelihood of marriage duration, T, and the latent education level achieved by marriage date, Z, for respondent i as

$$f_j\left(t_{ij}, Z_i | \boldsymbol{\theta}_j, x_i\right) = f_j\left(t_{ij} | \boldsymbol{\theta}_j, Z_i\right) p\left(Z_i | x_i\right). \tag{5}$$

Here, $f(t_{ij} | \boldsymbol{\theta}_j, Z_i)$ is the conditional likelihood function (4), and $p(Z_i | x_i)$ models the latent education level achieved by the date of marriage given the highest education level reported at the survey time.

To model Z, we define the probability that the level of education of respondent i at his marriage date was l given that he reported an education level $x_i = k$ at the survey time, as a function of the time $s_i = t_{x_i} - t_{m_i}$ between the date of marriage t_{m_i} and the date of completion of the reported highest educational level t_{x_i}:

$$p\left(Z_i = l | x_i = k\right) = \psi_{kl}\left(s_i\right), \ i = 1 \cdots, n; \quad \text{and} \quad l, k = 1, 2, 3. \tag{6}$$

We model Z by tracing backward in time the paths of education level progress taking into account that education is a non-decreasing process. This means, given the highest education level $x = k$ reported at the survey time, the plausible education level that could have been attained by marriage time is $l \le k$. Thus $p(Z_i = l | x_i = k)$ can be expressed in terms of the transition probability matrix

$$\Psi\left(s_i\right) = \begin{bmatrix} 1 & 0 & 0 \\ \psi_{21}\left(s_i\right) & \psi_{22}\left(s_i\right) & 0 \\ \psi_{31}\left(s_i\right) & \psi_{32}\left(s_i\right) & \psi_{33}\left(s_i\right) \end{bmatrix}, \tag{7}$$

for $k = 1, 2, 3$ (rows) and $l = 1, 2, 3$ (columns). Note that the probabilities sum to 1 in each row and that the primary education-level ($x = 1$) is an absorbing state because, based on evidence from the data, all individuals who reported to have primary-level education at the survey time have completed it before they married.

For a given row in Equation (7), Z follows a categorical distribution

$$p\left(Z_i | x_i = k\right) = \prod_{l=1}^{k} \left[\psi_{kl}\left(s_i\right)\right]^{z_{il}}, \tag{8}$$

where z_{il} are indicators of the education levels ($z_{il} = 1$ if $Z_i = l$ and $z_{il} = 0$ otherwise).

3.3. Estimation of the transition probabilities

We assume that x and Z are realizations from a continuous Markov chain $\{\tilde{X}(\tilde{t}) : \tilde{t} \in (t_b, t_s]\}$, where t_b and t_s are, respectively, the birth date and the survey date, and $\tilde{X}(\tilde{t})$ is defined on an outcome space of three states, $\Omega = \{1, 2, 3\}$, representing the three levels of education. This means $x = \tilde{X}(t_x)$ and $Z = \tilde{X}(t_m)$ where, as defined before, t_x and t_m denote the date at which the reported highest education level was attained and the date of marriage, respectively. Because x, t_x and t_m are observed, we only need to focus on the paths of the chain $\tilde{X}(\tilde{t})$ in the time frame $[t_m, t_x]$ to make inference about Z.

Since $t_m \leq t_x$ (for the anticipatory cases), and by definition $Z = \tilde{X}(t_m)$, it follows that Z can be obtained from reversing all paths of the chain $\tilde{X}(\tilde{t})$ passing by the state $x = \tilde{X}(t_x)$ at time t_x. This induces a reversed Markov chain $\{\tilde{X}(\tilde{s}) : \tilde{s} \in [t_m, t_x]\}$ with initial state $x = \tilde{X}(t_x)$ and final state $Z = \tilde{X}(t_m)$ which is the most plausible educational level achieved by the date of marriage. Note that the time frame $[t_m, t_x]$ can be translated to $[0, t_x - t_m]$, which allows to express the reversed Markov chain as $\{\tilde{X}(s) : s \in (0, t_x - t_m]\}$ implying that $x = \tilde{X}(0)$ and $Z = \tilde{X}(t_x - t_m)$.

Assuming it is not possible to skip an education level, transitions in $\{\tilde{X}(s) : s \in (0, t_x - t_m]\}$ can only occur between two consecutive states in a decreasing way as shown in the diagram below .
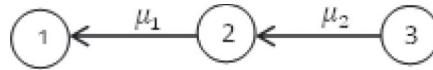
The process can begin either in $\tilde{X}(0) = 2$ or $\tilde{X}(0) = 3$ and it is assumed that during an infinitesimal small time Δs, reversed transitions from the state $\tilde{X}(s) = 3$ to the state $\tilde{X}(s + \Delta s) = 2$ occur at a rate of μ_2, while those from $\tilde{X}(s) = 2$ to $\tilde{X}(s + \Delta s) = 1$ (the absorbing state) occur at a rate of μ_1. Here, the rate is defined as the number of men completing the educational level in question per unit time. Therefore, both μ_1 and μ_2 are defined on $[0, +\infty)$.

The Markovian assumption implies that the probability of being in the state $l \in \Omega$ at a time $s + \Delta s$ depends only on the current state at time s, which can be expressed as

$$p_l(s + \Delta s) = \sum_{k=1}^{3} p_k(s) p_{kl}(s, s + \Delta s), \qquad (9)$$

where $p_l(\cdot)$ is the probability of being in the state $l \in \Omega$ at a specified time, and $p_{kl}(s, s + \Delta s)$ is the transition probability from the state k to the state l during an infinitesimal small time Δs.

In matrix notation, the expression in Equation (9) can be written as

$$P(s + \Delta s) = P(s) P(s, s + \Delta s), \qquad (10)$$

where $P(s, s + \Delta s)$ is a 3×3 matrix of transition probabilities $p_{kl}(s, s + \Delta s)$, and $P(\cdot) = [p_1(.) \; p_2(\cdot) \; p_3(\cdot)]$ is a vector of the marginal probabilities of the states at a specified time.

In order to complete the adjusted model, we need to define the transition probabilities $p_{kl}(s, s + \Delta s)$. We define them considering two facts: (1) transitions can only occur between consecutive states and (2) the states are not instantaneous, which means that before a transition occurs the process stays in the current state for some time.

These probabilities can be modelled, following Gross *et al.* [11], as

$$p_{kl}(s, s + \Delta s) = \begin{cases} 1, & k = 1, l = 1 \\ \mu_1 \Delta s, & k = 2, l = 1 \\ 1 - \mu_1 \Delta s, & k = 2, l = 2 \\ \mu_2 \Delta s, & k = 3, l = 2 \\ 1 - \mu_2 \Delta s, & k = 3, l = 3 \\ 0, & \text{elsewhere} \end{cases} \qquad (11)$$

These can then be used to estimate the conditional probabilities defined in Equation (7) as follows (see the supplemental material for derivation and more details on how these transition probabilities are estimated).

$$
\begin{aligned}
\psi_{21} &= 1 - \exp\{-\mu_1 s\}, \\
\psi_{22} &= \exp\{-\mu_1 s\}, \\
\psi_{31} &= \frac{\mu_1}{\mu_2 - \mu_1} \exp\{-\mu_2 s\} - \frac{\mu_2}{\mu_2 - \mu_1} \exp\{-\mu_1 s\} + 1, \\
\psi_{32} &= \frac{\mu_2}{\mu_2 - \mu_1} \left[\exp\{-\mu_1 s\} - \exp\{-\mu_2 s\}\right], \\
\psi_{33} &= \exp\{-\mu_2 s\}.
\end{aligned}
\tag{12}
$$

Thus, our final adjusted model is the joint likelihood,

$$
f_j\left(t_j, \mathbf{Z}_j | \boldsymbol{\theta}_j, x\right) = \prod_{k=1}^{3} \prod_{l=1}^{k} \prod_{i=1}^{n_j} \lambda_{ij}^{d_{ij}} \exp\left(-\lambda_{ij} t_{ij}\right) \psi_{kl}^{x_{ik} z_{il}},
\tag{13}
$$

where $\ln \lambda_{ij} = \boldsymbol{\theta}_j \mathbf{z}_i$, ψ_{kl} are the estimates in Equation (12) using $s_i = t_{x_i} - t_{m_i}$, and $x_{ik} = 1$ if the reported highest education level is $k \in \{1, 2, 3\}$ and $x_{ik} = 0$, otherwise. In the adjusted model, Z can either be the reported highest level or a lower level depending on the value of s. For instance, if a respondent reported post-secondary education at survey but we see that this was attained 2 years after marriage ($s = 2$), then his educational level at marriage date would be either post-secondary or secondary. The probability of primary, in this case, would be nearly zero as it requires more than 2 years to complete both secondary and post-secondary levels.

The adjusted model in Equation (13) can be seen as a generalization of the reduced and anticipatory models since they can be obtained as special cases by setting $\psi_{kl} = 1$ in Equation (13).

3.4. The prior distribution

To complete specification of the adjusted model, we specify prior distributions for the effect parameter paths $\boldsymbol{\theta}_{1:J}$ and the transition rates μ.

To insure smoothness in the regression parameters across different intervals I_j, we assume a prior of the form

$$
p\left(\boldsymbol{\theta}_{1:J}\right) = p\left(\boldsymbol{\theta}_1\right) \prod_{j=2}^{J} p\left(\boldsymbol{\theta}_j | \boldsymbol{\theta}_{j-1}\right),
\tag{14}
$$

where $p(\boldsymbol{\theta}_j | \boldsymbol{\theta}_{j-1})$ is the Gaussian random walk,

$$
\boldsymbol{\theta}_j = \boldsymbol{\theta}_{j-1} + \boldsymbol{\epsilon}_j, \quad \boldsymbol{\epsilon}_j \sim N\left(0, V_j\right),
\tag{15}
$$

and V_j is the variance that controls the evolution of the effect parameters through the intervals. Often, V_j is considered constant throughout the intervals I_j and a prior is set on it [7,12,21,23]. It is also possible to estimate V_j by using the discount procedure in [24]. We

follow such procedure in this paper where we assume V_j is proportional to the covariance matrix, C_{j-1}, of the posterior of the parameter $\boldsymbol{\theta}_{j-1}$ in the preceding interval I_{j-1}:

$$V_j = \left(\frac{1-\alpha}{\alpha} \right) C_{j-1},$$

where $0 < \alpha < 1$ is the discounting factor regulating the amount of information transferred from the previous to the current interval. As $\alpha \to 0$, then $V_j \to \infty$, leading the prior (15) to become non-informative; hence no communication among intervals. On the other hand, as $\alpha \to 1$, then $V_j \to 0$, which leads to a static evolution of the effect parameters. Here, we set $\alpha = 0.45$, following [20], to allow the effect parameters evolve smoothly and adapt to any local change that may occur in the hazard function. Further, we set the initial distribution $p(\boldsymbol{\theta}_1)$ to a non-informative normal distribution $N(0, 100 \times I_3)$, where I_3 is a three-dimensional unit matrix, to express lack of information in the first interval.

Finally, given that μ is defined on the positive real line, we propose independent gamma priors for the transition rates $\mu = (\mu_1, \mu_2)$:

$$\mu \sim \text{Gamma}(a_1, b_1) \times \text{Gamma}(a_2, b_2) \tag{16}$$

where a_1, a_2, b_1 and b_2 are positive real numbers which we set to: $a_1 = a_2 = b_1 = b_2 = 0.001$. In this way, we assume that *a priori* the rates are nearly zero and let the data dictate the optimal rates. Implicitly, the prior (16) suggests that we start *a priori* with the anticipatory model as μ_1 and μ_2 are close to zero and Z is close to the highest reported educational level.

3.5. The posterior distribution

The joint posterior distribution for the parameters in the adjusted model can be expressed as

$$p\left(\boldsymbol{\theta}_{1:J}, Z_{1:n}, \mu | t_{1:n}, s_{1:n}, x_{1:n}\right) = p\left(\boldsymbol{\theta}_{1:J} | t_{1:n}, Z_{1:n}\right) p\left(Z_{1:n} | s_{1:n}, x_{1:n}, \mu\right) p\left(\boldsymbol{\theta}_{1:J}\right) p\left(\mu\right), \tag{17}$$

where $\boldsymbol{\theta}_{1:J} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)$, $t_{1:n} = (t_1, \ldots, t_n)$ denotes the observed marriage duration, $Z_{1:n} = (Z_1, \ldots, Z_n)$ and $x_{1:n} = (x_1, \ldots, x_n)$ represent, respectively, the educational level at marriage time and the reported highest educational level for all n individuals. The parameter $\boldsymbol{\theta}_{1:J}$ is high-dimensional (where the dimension grows with J); therefore, estimating it jointly may be computationally inefficient.

One possibility to address this issue is to estimate the marginal posterior of $\boldsymbol{\theta}_j$, commonly referred to as the filtering distribution, recursively through time as follows:

$$p\left(\boldsymbol{\theta}_j | \mathbf{t}_{1:j}, \mathbf{Z}_{1:j}\right) \propto f_j\left(\mathbf{t}_j | \mathbf{Z}_j, \boldsymbol{\theta}_j\right) p\left(\boldsymbol{\theta}_j | \mathbf{t}_{1:j-1}, \mathbf{Z}_{1:j-1}\right), \tag{18}$$

where $\mathbf{t}_{1:j-1} = (\mathbf{t}_1, \ldots, \mathbf{t}_{j-1})$, $\mathbf{Z}_{1:j-1} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_{j-1})$, $\mathbf{Z}_j$, $\mathbf{t}_j$ are, respectively, vectors of educational levels and exposure times of individual who are at risk in I_j, and

$$p\left(\boldsymbol{\theta}_j | \mathbf{t}_{1:j-1}, \mathbf{Z}_{1:j-1}\right) = \int p\left(\boldsymbol{\theta}_j | \boldsymbol{\theta}_{j-1}\right) p\left(\boldsymbol{\theta}_{j-1} | \mathbf{t}_{1:j-1}, \mathbf{Z}_{1:j-1}\right) d\boldsymbol{\theta}_{j-1}. \tag{19}$$

The conditional posterior of the rate parameter, μ, can be expressed as

$$p\left(\mu | Z_{1:n}, s_{1:n}, x_{1:n}\right) = \prod_{k=1}^{3} \prod_{l=1}^{k} \prod_{i=1}^{n} \left[\psi_{kl}\left(s_i\right)\right]^{x_{ik} z_{il}} p\left(\mu\right), \tag{20}$$

where $p(\mu)$ is the joint prior of μ, and ψ_{kl} are the transition probabilities defined in Equation (7) and estimated in Equation (12).

Lastly, the conditional posterior of the latent parameter Z can be expressed as

$$p\left(Z_i = l | t_i, , s_i, \mu, \boldsymbol{\theta}_{1:J}, x_i = k\right) \propto \prod_{j=1}^{J} f_j\left(t_{ij} | Z_i, \boldsymbol{\theta}_j\right) \psi_{kl},$$

$$k = 1, \ldots, 3, \quad l = 1, \ldots, k, \; i = 1, \ldots, n. \tag{21}$$

3.6. Inference

Since the reduced and anticipatory models are nested within the adjusted model, we discuss only the method of sampling from the adjusted model. The same procedure applies to the former models but with some non-applicable steps skipped.

To sample from the joint posterior distribution of the adjusted model, we apply a Gibbs sampler that follows three steps at each iteration $m = 1, 2, \ldots$

(1) Sample $\mu^m | Z_{1,\ldots,n}, s_{1,\ldots,n}, x_{1,\ldots,n}$
(2) Sample $Z_i^m | t_i, \mu^m, \theta_{1:J}, x_i$
(3) Sample $\boldsymbol{\theta}_{1:J}^m | t_{1,\ldots,n}, \mu^m, Z_{1,\ldots,n}^m, x_{1,\ldots,n}$

Starting from an initial path of the effect parameters $\boldsymbol{\theta}_{1:J}^0$ and an initial education level Z^0, we sample the vector of the rates μ^m from its conditional posterior distribution given in Equation (20). Then, for each individual, we sample an education level Z^m from the categorical distribution in Equation (8) with probability proportional to the expression in Equation (21). Finally, we sample a path $\boldsymbol{\theta}_{1:J}^m$ from the conditional posterior in Equation (18).

The first and last steps cannot be implemented directly because the expressions in Equations (18) and (20) are not tractable. We replace the first step by a Metropolis–Hastings kernel. Thus, at the mth iteration, μ^m is proposed from a proposal distribution $g(\mu|\mu^{m-1})$ and it is accepted with probability proportional to

$$\min\left(1, \frac{p\left(\mu^m | z, x\right) g\left(\mu^{m-1} | \mu^m\right)}{p\left(\mu^{m-1} | z, x\right) g\left(\mu^m | \mu^{m-1}\right)}\right). \tag{22}$$

Since the rates are defined on the positive real line, we propose μ_1 and μ_2 independently from the gamma distribution. To select the parameter values in the proposal distribution, we use the observed mean and variance of s for each educational level (in the anticipatory cases) as guidance.

According to (12), the distribution of s is implicitly an exponential distribution, which implies that the mean of s is inversely related to the rate μ. The parameters in the proposal distribution are, therefore, obtained by matching the observed mean and variance of

s computed from the data with the corresponding moments of an inverse-gamma distribution. The observed mean and variance are 4.28 and 11.55 respectively, for the secondary level, and 4.19 and 7.12, for post-secondary level. Hence, the proposal distribution g is set to independent Gamma(4, 11) and Gamma(5, 15) for μ_1 and μ_2, respectively.

The final step requires evaluating the integral (19) which, for our model, is intractable. In addition, the autocorrelations of the effect parameters induced by the random walk prior process (15) may hinder the convergence of the sampler. To overcome these issues, Andrieu *et al.* [1] suggest using a conditional sequential Monte Carlo kernel (commonly known as particle filter) to approximate the conditional posterior in Equation (18). Particle filters approximate $p(\boldsymbol{\theta}_j|\mathbf{t}_{1:j}, \mathbf{Z}_{1:j})$ empirically by a discrete distribution defined on a finite set of points $\{\boldsymbol{\theta}_j^{(h)}\}_{h=1}^H$, referred to as particles, with probability masses $\{w_j^{(h)}\}_{h=1}^H$ commonly known as importance weights. With this weighted sample of particles, the integral (19) can be approximated as

$$p(\boldsymbol{\theta}_j|\mathbf{t}_{1:j-1}, \mathbf{Z}_{1:j-1}) \approx \sum_{h=1}^H p(\boldsymbol{\theta}_j|\theta_{j-1}^h) w_{j-1}^h.$$

Particle filter algorithms provide a framework of computing the importance weights recursively as new data from intervals I_j are observed, and they enable efficient and computationally fast inference in dynamic models. A comparison of the particle filter and the Markov Chain Monte Carlo (MCMC) algorithms applied to survival dynamic models is provided in [20].

The conditional sequential Monte Carlo kernel of Andrieu *et al.* [1] runs the particle filter with the condition that one reference path $\boldsymbol{\theta}_{1:j}^{ref}$ is *a priori* set deterministically and after a complete run of the particle filter, one path $\boldsymbol{\theta}_{1:J}^m$ is selected from the sample of particles $\{\boldsymbol{\theta}_{1:J}^{(h)}\}_{h=1}^H$. However, setting the reference path deterministically may lead to a slow mixing of the sampler because particle filters are prone to degeneracy in the importance weights. This problem occurs when a few particles have significantly high importance weights but the rest have importance weights close to zero.

To overcome this problem, Lindsten *et al.* [19] suggest randomizing the reference path through an extra resampling step referred to as ancestor sampling. The ancestor sampling procedure allows fast mixing of the sampler and does not require many particles in the underlying particle filter step [19]. We, therefore, follow the particle Gibbs with ancestor sampling (PGAS) of Lindsten *et al.* [19] and apply the particle filter algorithm of Munezero [20]. This algorithm was designed specifically for survival data and proven to be computationally fast and efficient. In our illustrative example, we achieved fast convergence by setting the number of particles H to 150. See supplemental material for details on convergence diagnostics and posterior predictive checks.

4. Analysis of effect of education on divorce risk

4.1. Preliminary analyses using Cox proportional hazards (PH) model

To begin with, we fitted a standard Cox proportional hazards (PH) model [3] of the form

$$\lambda(t) = \lambda_0(t) \exp(\beta x), \tag{23}$$

Table 2. Relative risks of divorce by educational level (from Cox PH model).

Education	All (n=1312)	Non-Anticip. (n=1067)	Anticip. (n=245)
Primary	0.897 (p=0.502)	1.050 (p=0.074)	-
Secondary	1	1	1
Post-secon.	1.194 (p=0.314)	1.596 (p=0.068)	0.693 (p=0.160)

to the entire data as well as separately to non-anticipatory and anticipatory cases (those on the left and right of the main diagonal in Figure 1, respectively), with results shown in Table 2. Thus using those with secondary level as a baseline (reference) category, we note, for the entire sample, that those with primary level education have about 10% lower risk while those with post-secondary level have about 20% higher risk of divorce. In other words, divorce risk increases as educational level increases though the *p*-values indicate that the results are not significant at significance levels of about 5%.

For those who completed their reported educational level before marriage (the non-anticipatory cases to the left of the diagonal in Figure 1) the corresponding excess risks of divorce are 5% for those with primary level education and about 60% for those with post-secondary education. Further, the results are not significant at significance levels of about 5%.

In contrast, the results for the anticipatory cases are in the opposite direction – those with post-secondary education have about 30% lower risk of divorce than those with secondary education. Again, the corresponding p-value shows the difference is not statistically significant at significance levels of about 5%.

These preliminary results give an early warning that educational gradients in divorce risks differ between the anticipatory and non-anticipatory cases which, in turn, prompts to our proposed adjusted model.

4.2. Results from the adjusted model and comparison with previous findings

Our proposed model requires defining the J intervals partitioning the marriage duration. Common practice sets the interval limits at each event time [6]. However, doing this would result in $J = 151$ intervals which requires huge computational efforts. To alleviate this, Munezero [20] proposes setting the intervals at event times in such a way that all intervals contain the same number of events. Inference about the optimal number of events per interval, E, can be done using any predictive information criteria. Following [20], we use the Watanabe–Akaike information criterion, WAIC (see [8] for the definition and details about WAIC). We obtain the following WAIC values: 2209.28, 2205.49, 2206.47 and 2208.15, respectively, for $E = 5$, $E = 10$, $E = 15$ and $E = 20$. These WAIC values suggest to use $E = 10$ which leads to a partition of time into $J = 15$ intervals.

In Figure 2, we present the estimated probabilities of transition for different educational levels. It turns out that a respondent (among anticipatory cases) who reported post-secondary educational level is most likely to have had secondary level at his marriage date, if he spent at least 2–4 years of studies after marriage. However, if he spent at least 4 years studying, it is more likely that he had completed two educational levels; that is he had primary level at his marriage date. For those who reported secondary educational level, the probability that they had primary educational level at their marriage date becomes
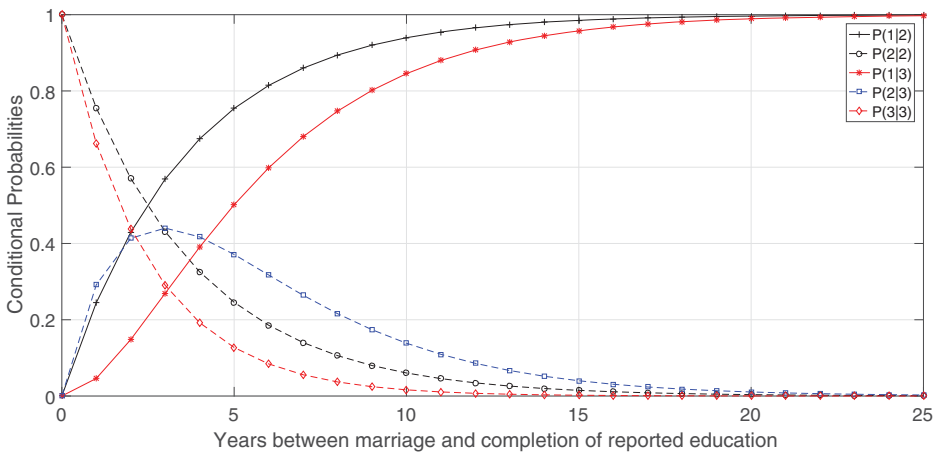
Figure 2. Estimated conditional probabilities of educational levels at marriage time for given educational levels at survey time

considerably high after 2.5 years of study after marriage. The estimated mean times to complete secondary and post-secondary education are $\frac{1}{\mu_1} = 3.56$ years and $\frac{1}{\mu_2} = 2.43$ years, respectively.

From the estimated transition rates, we can also compute $\widehat{\frac{\mu_2}{\mu_1+\mu_2}} = 0.40$ and $\widehat{\frac{\mu_1}{\mu_1+\mu_2}} = 0.60$ which are estimates of the proportions (among the anticipatory cases) with secondary and post-secondary level educations, respectively. These figures can be compared with the corresponding entries in Table 1 ($94/245 = 0.38$ and $151/245 = 0.62$, respectively). We use these quantities in our posterior predictive checks, see the supplemental material.

Figure 3 shows relative risks of divorce for those with primary level education (relative to those with secondary level education) by marriage durations and across the three models (reduced model, anticipatory model and adjusted model). The corresponding relative risks for those with post-secondary education are shown in Figure 4. We note from the figures that the relative risks are underestimated in the anticipatory model (red line) compared to the reduced model and adjusted model (black and blue, respectively) in most of the marriage durations. We also note that the degree of over estimation varies across marriage durations and that after about 11 years in marriage the underestimation turns to overestimation.

Figures 5–7 show relative risks of divorce for those with primary and post-secondary education (relative to those with secondary education) in the Reduced Model, Anticipatory Model and Adjusted Model, respectively. Comparing Figures 6 and 7 we note that those with post-secondary education have higher risk of divorce than the baseline level for most of the marriage durations in the adjusted model (Figure 7) but that such super-risk is underestimated (in some intervals to the extent of being sub-risk) in the anticipatory model (Figure 6). Even those with primary-level education seem to have higher risks of divorce than the baseline over most of the marriage durations but, again, these are underestimated in the anticipatory model (Figure 6) compared to the adjusted model (Figure 7). Figure 5 shows that the educational gradient of divorce is much higher in the reduced model (where the anticipatory cases are discarded) and this is consistent with what we found in the preliminary analyses in Section 4.1.
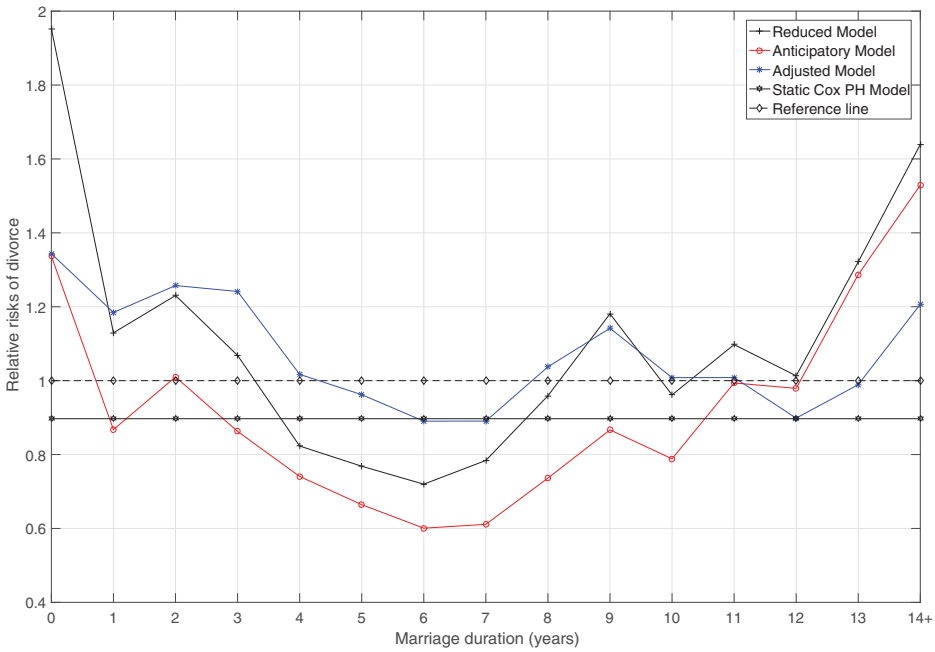
Figure 3. Relative risks of divorce for primary education relative to secondary education across the three models.
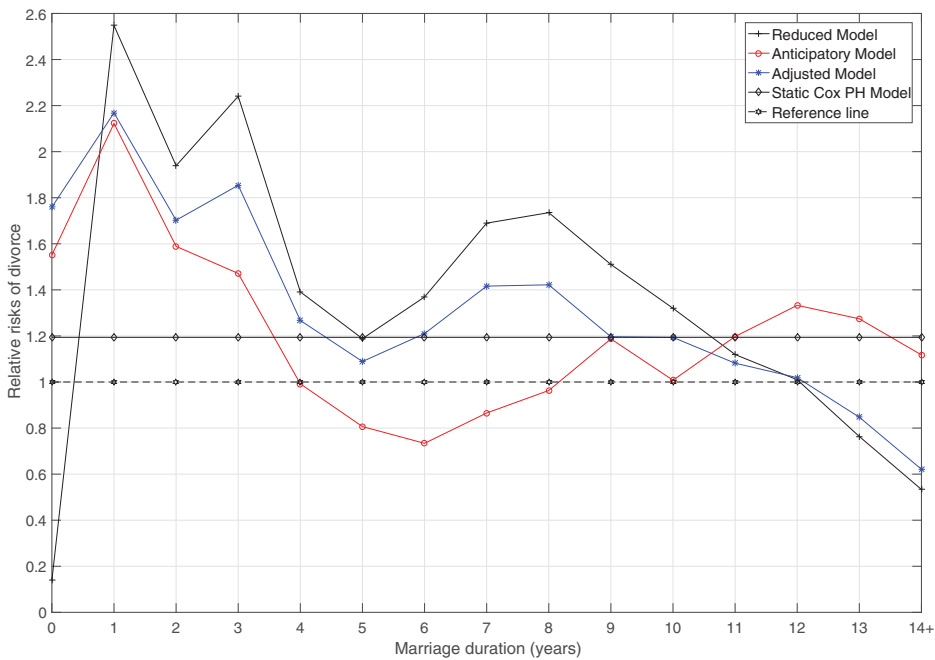

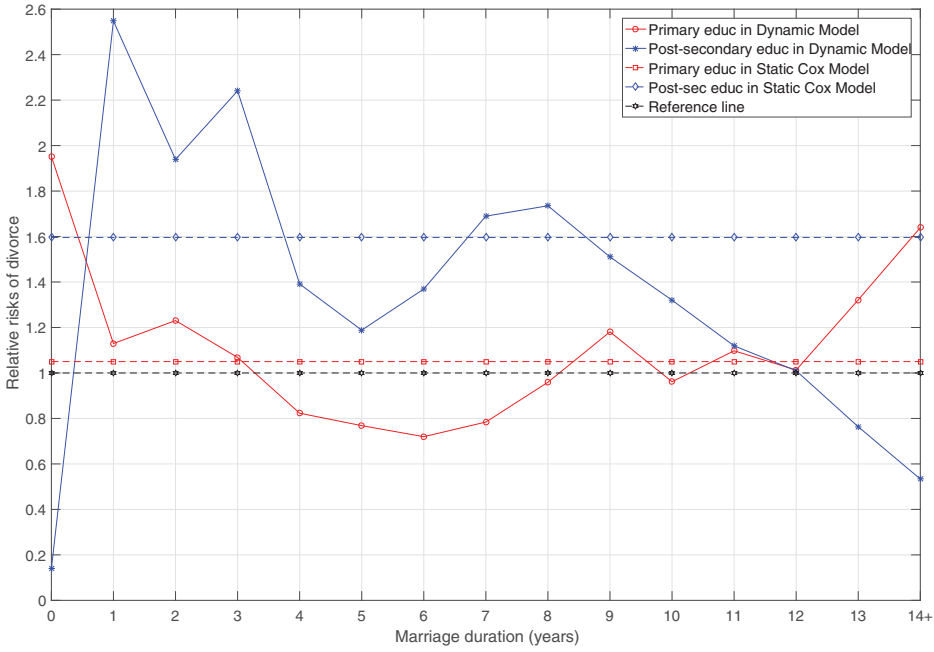**Figure 4.** Relative risks of divorce for post-secondary education relative to secondary education across the three models.

Figure 5. Relative risks of divorce for primary and post-secondary education relative to secondary education in the Reduced Model.
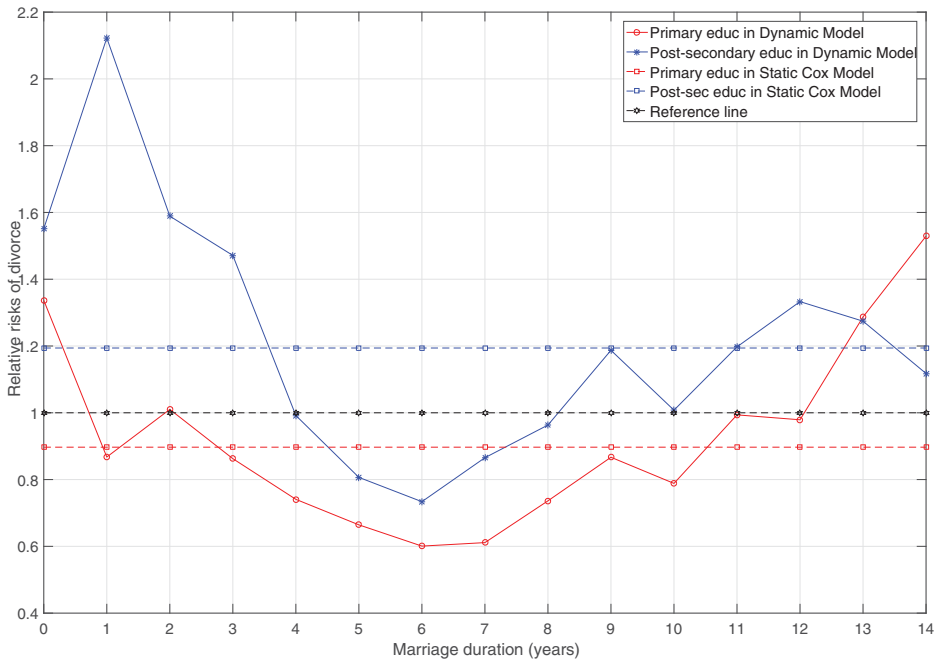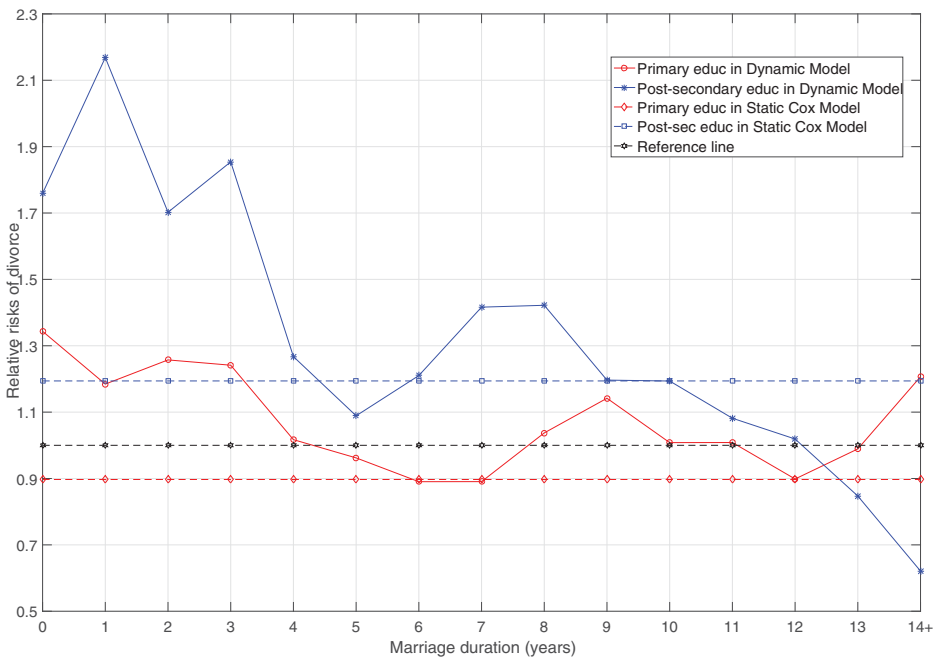

Figure 6. Relative risks of divorce for primary and post-secondary education (relative to secondary education) in the Anticipatory Model.

Figure 7. Relative risks of divorce for primary and post-secondary education (relative to secondary education) in the Adjusted Model.

Figure 8 presents the percentage under/over-estimation of relative risks of divorce in the anticipatory model (compared to the adjusted model). We note here that the relative risks of divorce for both educational levels (primary and post-secondary) are underestimated in the anticipatory model for marriage durations until 10 years whereafter they turn to overestimation.

In Table 3, we summarize our results and compare them with those from previous studies – Ghilagaber and Koskinen (G & K [9]) and Ghilagaber and Larsson (G & L [10]). Our present results shown in the last column are obtained by averaging the corresponding relative risks over the $J = 15$ time intervals partitioning the marriage duration.

We note that our results in the adjusted model (last column third panel) are closer to those of Ghilagaber and Koskinen [9] for those with primary level education. For those with post-secondary education, our results are more close to those of Ghilagaber and Larsson [10].

Note, however, that the three approaches are not comparable in all aspects and we will focus on our own results in the discussion below. While the three studies analyse the same data set, the two previous studies assume constant relative risks of divorce across marriage duration while the present study allows relative risks to vary over marriage duration. Further, the previous studies consider a Markovian process evolving forward in time (from birth) for times to complete the various educational levels. In the present study, we model educational levels using a reversed (backward) Markov chain and allow the transition probabilities between educational levels to depend on the time spent on education after marriage.
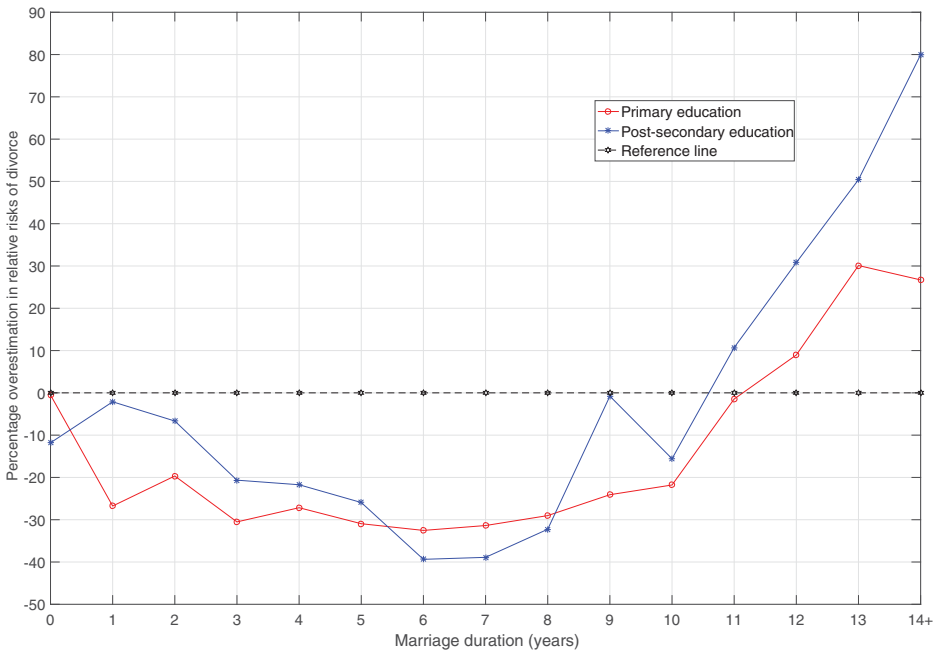
Figure 8. Percentage under-estimation (or over-estimation) of relative risks of divorce in the anticipatory model compared to the adjusted model.

Table 3. Estimated educational gradients of divorce-risks (and 95% confidence intervals in brackets) in the reduced, anticipatory and adjusted models from three studies. Results in the third column refer to those of Ghilagaber and Koskinen (2009), results in the fourth column refer to those of Ghilagaber and Larsson (2019), while the results in the last column are our findings in the present study.

Model	Education	G & K [9]	G & L [10]	Present work
Reduced	Primary	1.04 (0.74,1.51)	1.03 (0.72,1.49)	1.07 (0.45,2.48)
Model	Second (ref)	1	1	1
	Post-Sec	1.65 (1.38,1.85)	1.62 (1.34,1.81)	1.16 (0.78,1.78)
Anticip.	Primary	0.89 (0.65,1.23)	0.89 (0.64,1.20)	0.89 (0.42,1.91)
Model	Second (ref)	1	1	1
	Post-Sec	1.21 (1.14,1.26)	1.19 (1.12,1.22)	1.17 (1.03,1.26)
Adjusted	Primary	1.02 (0.77,1.41)	0.95 (0.69,1.28)	1.06 (0.47,2.43)
Model	Second (ref)	1	1	1
	Post-Sec	1.09 (0.98,1.18)	1.28(1.17,1.33)	1.26 (0.59,3.59)
Estimated	Primary	14.80	16.40	-
years to	Secondary	6.5	3.65	3.56
complete	Post-Sec	6.3	1.90	2.43

As measures of uncertainty and 'indirect' indicators of goodness-of-fit, we also present, in Table 3, 95% confidence/credible intervals and estimated number of years to complete the respective educational levels. Thus the expected number of years to complete primary level educational is 14.8 years (counted from birth) according to Ghilagaber and Koskinen [9] and 14.4 years according to Ghilagaber and Larsson [10]. We don't have a corresponding estimate in the current study because, as mentioned before, the formulation in the present study treats primary level education as absorbing state since all individuals who reported to have primary-level education at the survey time have completed it before

they married (and, hence, there is no anticipatory case among those with primary level education).

5. Discussion

We addressed an important issue in inference with observational data where values of a covariate refer to what is achieved by the survey time but the covariate is used as regressor in models that relate it to behaviour that took place long before the survey. This causes problem because the value of the current-date (anticipatory) covariate does not follow the temporal order of events. We attempted to tackle the problem by proposing a dynamic Bayesian approach allowing to model the event of interest and the anticipatory covariate jointly, and the effect parameters to vary over time.

We illustrated our problem by modelling effects of educational level attained at survey time on the risk of divorce among Swedish men. The results showed that ignoring the anticipatory nature of the education variable led to the underestimation of educational gradients on divorce for marriage durations until about 10 years. After 10 years, the effect becomes overestimation of the relative risks. This may be associated with differential risks in divorce between those who have completed their reported educational level short after marriage and those who did so long after marriage and closer to the survey date.

The overall effect of ignoring the anticipatory nature of the education variable led to spurious significance of the relative risk of divorce for those with post-secondary education (relative to those with secondary education) and to much lower risk for those with primary education without affecting its significance. We also demonstrate that the relative risks are not constant over the marriage durations – something that is ignored while using standard methods like the proportional hazards models.

There are some open questions that can be addressed in future investigations. For instance, it is not clear whether the anticipatory nature of our education variable (the fact that the unobserved educational level comes before the observed one) makes the problem different from classical misclassification problems. We believe that our present study does not suffer from differential measurement error because we were interested in the effect of educational level attained by the time of marriage. A possible suggestion for future investigation is, thus, to examine the prevalence, impact and adjustment of differential measurement error where the likelihood of having anticipatory education (completing it after divorce) is correlated with divorce risk. This can arise if, for instance, the experience of divorce increases or decreases the propensity to continue education.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Parfait Munezero 🔵 http://orcid.org/0000-0002-3902-3846
Gebrenegus Ghilagaber 🔵 http://orcid.org/0000-0002-2910-8432

References

[1] C. Andrieu, A. Doucet, and R. Holenstein, *Particle Markov Chain Monte Carlo methods*, J. R. Stat. Soc. Ser. B. Stat. Methodol 72 (2010), pp. 269–342.

[2] W. Arulampalam and S. Bhalotra, *Sibling death clustering in India: state dependence versus unobserved heterogeneity*, J. R. Stat. Soc. Ser. A. Stat. Soc. 169 (2006), pp. 829–848.

[3] D.R. Cox, *Regression models and life-tables*, J. R. Stat. Soc. Ser. B. Stat. Methodol. 34 (1972), pp. 187–202.

[4] L. Fahrmeir and T. Kneib, *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*, Oxford University Press, Oxford, 2011.

[5] C.L. Faucett, N. Schenker, and R.M. Elashoff, *Analysis of censored survival data with intermittently observed time-dependent binary covariates*, J. Am. Statist. Assoc. 93 (1998), pp. 427–437.

[6] D. Gamerman, *Dynamic Bayesian models for survival data*, J. R. Stat. Soc. Ser. C. Appl. Stat. 40 (1991), pp. 63–79.

[7] D. Gamerman, *Markov Chain Monte Carlo for dynamic generalised linear models*, Biometrika 85 (1998), pp. 215–227.

[8] A. Gelman, J. Hwang, and A. Vehtari, *Understanding predictive information criteria for Bayesian models*, Stat. Comput. 24 (2014), pp. 997–1016.

[9] G. Ghilagaber and J. Koskinen, *Bayesian adjustment of anticipatory covariates in analyzing retrospective data*, Math. Popul. Stud. 16 (2009), pp. 105–130.

[10] G. Ghilagaber and R. Larsson. Maximum likelihood adjustment of anticipatory covariates in the analysis of retrospective data. Department of Statistics, Stockholm University; 2019. (Research Report 2019:1).

[11] D. Gross, J.F. Shortle, J.M. Thompson, and C.M. Harris, *Fundamentals of Queueing Theory*, 4th ed., John Wiley & Sons, Hoboken, 2008.

[12] K. Hemming and J. Shaw, *A parametric dynamic survival model applied to breast cancer survival times*, J. R. Stat. Soc. Ser. C. Appl. Stat. 51 (2002), pp. 421–435.

[13] A. Hennerfeind, A. Brezger, and L. Fahrmeir, *Geoadditive survival models*, J. Am. Statist. Assoc. 101 (2006), pp. 1065–1075.

[14] J.M. Hoem, *The harmfulness or harmlessness of using an anticipatory regressor: how dangerous is it to use education achieved as of 1990 in the analysis of divorce risks in earlier years?*, Finn. Yearb. Popul. Res 0 (1996), pp. 34–43.

[15] J.M. Hoem and M. Kreyenfeld, *Anticipatory analysis and its alternatives in life-course research. Part 1: the role of education in the study of first childbearing*, Demogr. Res. 15 (2006), pp. 461–484.

[16] J.M. Hoem and M. Kreyenfeld, *Anticipatory analysis and its alternatives in life-course research. Part 2: two interacting processes*, Demogr. Res. 15 (2006), pp. 485–498.

[17] L. Johansson. Family formation among Swedish men and their plans for children. Sweden, Örebro: Statistics; 1990. (Demographic Reports 1990:2. In Swedish)

[18] Ø. Kravdal, *An illustration of the problems caused by incomplete education histories in fertility analyses*, Demogr. Res. 3 (2004), pp. 135–154.

[19] F. Lindsten, M.I. Jordan, and T.B. Schön, *Particle Gibbs with ancestor sampling*, J. Mach. Learn. Res. 15 (2014), pp. 2145–2184.

[20] P. Munezero. *Efficient particle smoothing for Bayesian inference in dynamic survival models*, preprint (2018). Available at arXiv:1806.07048

[21] D.J. Sargent, *A flexible approach to time-varying coefficients in the Cox regression setting*, Lifetime Data Anal. 3 (1997), pp. 13–25.

[22] H. van Houwelingen and H. Putter, *Dynamic Prediction in Clinical Survival Analysis*, Boca Raton , FL, CRC Press, 2011.

[23] H. Wagner, *Bayesian estimation and stochastic model specification search for dynamic survival models*, Stat. Comput. 21 (2011), pp. 231–246.

[24] M. West, P.J. Harrison, and H.S. Migon, *Dynamic generalized linear models and Bayesian forecasting*, J. Am. Statist. Assoc. 80 (1985), pp. 73–83.