Dynamic Bayesian Adjustment of Educational Gradients in Divorce Risks: Disentangling Causation and Misclassification*

Parfait Munezero† and Gebrenegus Ghilagaber‡

1 Introduction: Background and Aim

Anticipatory covariates (or current-date variables) are variables whose values refer to what is attained by the date of the survey (interview); but are used to explain behavior in life course which took place before the survey. Highest educational level and social class at survey time are typical examples of anticipatory variables. 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. In subsequent analyses of such retrospective survey data, these variables are used as regressors in modeling the rate/intensity of some event such as marriage or divorce - events which took place long before the survey. This causes a time inconsistency because data that pertain to the date of the survey become less and less informative the further the date of the survey is from the date of the event.

Suppose the highest level of education reported at the time of the survey is used as a covariate in modeling the risk of divorce (which occurred before the survey). Educational progress is likely to occur over time (for instance between entry into marriage and the interview date). It is then natural to ask:

*Contribution to the 2019 Annual Meeting of the Population Association of America (PAA), JW Marriott, Austin, TX, USA, 10 - 13 April 2019.

†Department of Statistics, Stockholm University, Sweden. Parfait.Munezero@stat.su.se

‡Presenting author: Dept. of Statistics, Stockholm Univ., Sweden. Gebre@stat.su.se

- to what extent can any changes in divorce-patterns across educational levels be attributed to changes in the distribution of respondents across the various educational levels;

- and to what extent do they reflect real (causal) differences in behavior of different levels of education?

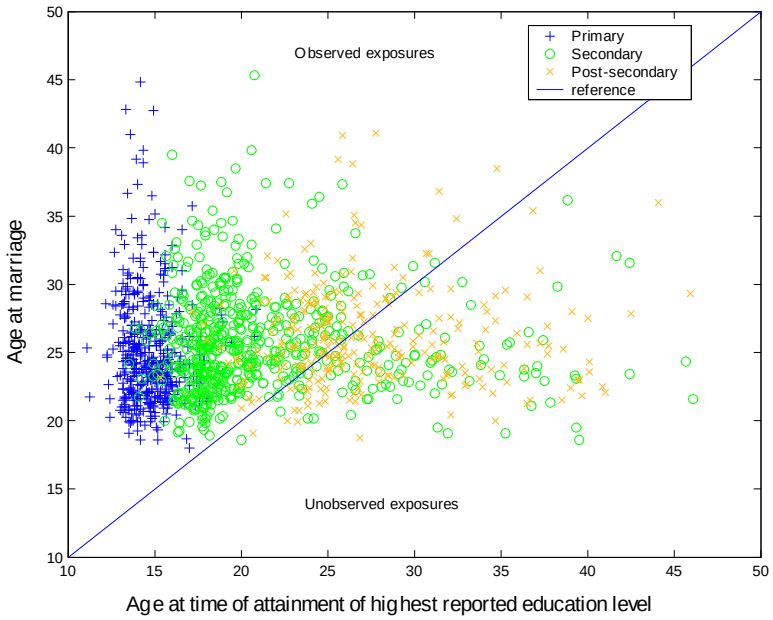
Clearly, ignoring such anticipatory nature of the covariate (education) will potentially produces biased estimates of educational gradients in divorce-risks..

The aim of the present work is to address this important issue and adjust for the anticipatory nature of education in modeling its effect on the risk of divorce. More precisely, we want to use respondents' educational level achieved by the date of marriage as a covariate in modeling the risk of divorce. Such information restors the temporal order of events and, hence, can be interpreted as causal effect in modeling the risk of divorce. Unfortunately, such information (educational level at the date of marriage) is not available as the only information we have at hand is the highest educational level achieved by the date of interview.

To solve this paradox we model the risk of divorce dynamically and let the effect of educational level on the risk of divorce vary over time which, in this, case is the duration of marriage. Note that this is different from time-varying covariates where the value of educational level is allowed to vary over time. Instead, dynamic modeling allows the effects (coefficients) to vary between the date of marriage and the date of divorce or interview, whichever comes first.

2 Data

Data for illustration of our proposed approach above is an extract from a larger sample and consist of 1312 ever- married Swedish men who were either divorced before the survey date (event) or still married at the time of interview (censored). These men have been exposed to the risk of divorce at the date of marriage (which was long before the survey date). Distribution of the data ages at marriage and at completion of the highest reported educational level are shown in the figure below.



In the hypothetical case where all respondents completed their reported highest educational level at the date of marriage, then all points will fall on the main diagonal. Note here that if all respondents completed. This is not the case, however, for our data and we note that 245 of the 1312 respondents (18. 67 %) fall to the right of the diagonal - meaning they have completed their reported highest educational level after they have married and, hence, their educational level is anticipatory.

More interesting is that a good proportion of these anticipatory cases have reported to have post-secondary education (shown in grey color in the figure) while some others have reported to have secondary-level education (shown in green). All of those who reported to have only primary-level education (those in blue in the figure) have completed their education before they married.

The question at hand is, then, as to what to do with these 245 respondent with anticipatory educational level (completed after they have married) as we want to deal with situations where education is completed before entry to marriage (to keep the temporal order of events). We have three options:

1. One straightforward option is, of course, to delete these respondents from the analyses and use the reduced sample with 1312 - 245 = 1067 respondents. We shall subsequently label this option as the 'Reduced Model'.

2. Another option is to 'turn a blind eye' and pretend it is harmless to consider those to the right hand of the diagonal as if they actually were to the left. This approach, which is the common practice in the literature, will be labelled 'Anticipatory Model'.

3. A third approach, which is the main contribution of this study, is to devise a method to adjust for the anticipatory nature of the variable. In effect, this implies computing a plausible value of educational level at the time marriage for those respondents who have completed the highest reported education after they have married. We shall label this approach the 'Adjusted Model'

3 Models and Methods

The basic model a dynamic extension of the discrete time (piecewise exponential) Cox Proportional Hazards model where the coefficients are allowed to vary over time:

$$\lambda\left(t|Z\right)=z)=\lambda_0\left(t\right)\exp\left[\widehat{\beta\left(t\right)}*z\right]$$

where $\lambda_0\left(t\right)$ is the baseline hazard at time t, $\widehat{\beta\left(t\right)}$ is the effect coefficient at time t (to be estimated from the model), and z is a realized value of the covariate Z (here educational level).

A piecewise constant hazard specification of $\lambda_0\left(t\right)$ implies that the time until event or censoring (whichever comes first) in interval I_j is exponentially distributed with density function, $f_j\left(t_{ij}|\lambda_{ij},z\right)=\lambda_{ij}\exp\left[-\lambda_{ij}*t_{ij}\right]$ where $\lambda_{ij}=\exp\left(\beta_{1j}+\beta_{2j}z_{2j}+\beta_{3j}z_{3j}\right)$. Here, β_{1j} refers to the baseline educational level, level 1 (say primary level) while β_{2j} and β_{3j} refer to coefficients of educational level 2 (say secondary) and educational level 3 (say post-secondary), respectively. The joint likelihood of the time to event, T, and the latent educational level, Z, is then given by

$$f\left(t,z|\beta_{1:J},x\right)=f\left(t,z|\beta_{1:J}\right)*P\left(Z=z|X=x\right)=\prod_{j=1}^{I}f_j\left(t_j,z|\beta_j,z\right)*P\left(Z=z|X=x\right),$$

where $P\left(Z=z|X=x\right)$ corresponds to the missing educational level at marriage, z, for those who have reported educational x at the time of survey.

The above joint likelihood is then combined with reasonable priors for the parameters of interest to form a posterior distribution from which we sample using conditional particle filter with ancestor sampling algorithm.

4 Preliminary Results

Relative risks of divorce across marriage durations (in years) are plotted in the figure below. The plot in the left presents relative risks for those who reported to have secondary level education (relative to the baseline group with primary education) while that on the right presents relative risks of divorce for those who reported to have post-secondary education. The three curves in each plot refer to the three models (greed for reduced model, red for anticipatory model, and blue for adjusted model). Thus, the plots show that ignoring the anticipatory nature of education and, hence, using anticipatory model (as is the case in most contemporary analyses) leads to over-estimation of the relative risks of divorce. The degree of over-estimation is stronger for those who reported to have secondary-level education. This is in accordance with expectation as those who reported to have secondary-level are more likely (than those who reported post-secondary) to have had primary level education (which is the baseline level) at the time of marriage. Further, we note that the over-estimation varies over marriage durations for those with secondary-level education while there is only slight overestimation at the early stages of marriage for those with post-secondary education.