EXOGENOUS TIME-VARYING COVARIATES IN DOUBLE ADDITIVE CURE SURVIVAL MODEL WITH APPLICATION TO FERTILITY

Philippe Lambert, Michaela Kreyenfeld

LIDAM Discussion Paper ISBA
2023 / 06

UCLouvain

LIDAM
ISBA | Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication.html

Exogenous time-varying covariates in double additive cure survival model with application to fertility

Philippe Lambert*†and Michaela Kreyenfeld‡

January 31, 2023

Abstract

Extended cure survival models enable to separate covariates that affect the probability of an event (or *long-term* survival) from those only affecting the event *timing* (or *short-term* survival). We propose to generalize the bounded cumulative hazard model to handle additive terms for time-varying (exogenous) covariates jointly impacting long- and short-term survival. The selection of the penalty parameters is a challenge in that framework. A fast algorithm based on Laplace approximations in Bayesian P-spline models is proposed. The methodology is motivated by fertility studies where women's characteristics such as the employment status and the income (to cite a few) can vary in a non-trivial and frequent way during the individual follow-up. The method is furthermore illustrated by drawing on register data from the German Pension Fund which enabled us to study how women's time-varying earnings relate to first birth transitions.

Keywords: Additive model ; Bounded hazard ; Cure survival model ; Fertility study ; Laplace approximation ; P-splines ; Time-varying covariates.

1 Introduction

Proportional hazards models are used extensively to analyze time-to-event data and their association to covariates. They enable to summarize group differences using risk ratios assumed constant over time. Cure survival models (Boag, 1949; Berkson and Gage, 1952) explicitly acknowledge that a proportion of the studied population will never experience the event of interest whatever the duration of the follow-up. This can be revealed or confirmed with the inspection of the estimated survival functions (such as Kaplan-Meier curves) found to reach a plateau at a non-zero level for large values of the follow-up time. We will focus here on the *promotion time* (cure) survival model, also named the *bounded cumulative hazard model* (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen

*Institut de Mathématique, Université de Liège, Belgium. Email: p.lambert@uliege.be
†Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), Université catholique de Louvain, Belgium.
‡Hertie School Berlin, Germany.

et al., 1999). Let $\mathbf{v} = (\mathbf{z}, \mathbf{x})$ where $\mathbf{z}$ denote a $p-$vector of categorical covariates, and $\mathbf{x}$ a $J-$vector of quantitative covariates (with $\mathbf{0}$ generically used to refer to their reference values). If $S_p(t|\mathbf{v})$ is the conditional survival function for subjects (including cured individuals) sharing these characteristics, then

$$S_p(t|\mathbf{v}) = \exp\{-\vartheta(\mathbf{v})F(t)\} \tag{1}$$

where $t > 0$, $\vartheta(\mathbf{v}) > 0$ and $F(t)$ is a cumulative distribution function such that $F(0) = 0$ and $F(T) = 1$ with T denoting the minimal survival time after which a subject can be declared cured. The proportion of cured subjects in the sub-population defined by $\mathbf{v}$ is

$$\pi(\mathbf{v}) = S_p(T|\mathbf{v}) = \exp\{-\vartheta(\mathbf{v})F(T))\} = \exp\{-\vartheta(\mathbf{v})\} > 0.$$

Let $f(t) = dF(t)/dt$. When $\vartheta(\mathbf{v}) = \exp\{\eta_\vartheta(\mathbf{v})\}$ with $\eta_\vartheta(0) = \beta_0$, (1) corresponds to a proportional hazards (PH) model with baseline hazard $\mathrm{e}^{\beta_0} f(t)$ and a cumulative hazard bounded by $\vartheta(\mathbf{v})$. The dynamics in the hazard function, governed by $f(t)$, is not affected by covariates: it ensures the constant hazard ratio characterizing the PH model. This is a crucial assumption that is not always properly assessed with potential consequences on the quality of the conclusion derived from the research. Even when the estimated survival curves for the compared groups do not cross and the parallelism of the logarithm of the underlying cumulative hazards is not challenged during a follow-up interrupted by right censoring, the PH hypothesis might be violated further out in time. Indeed, the survival probabilities could increasingly diverge or become similar beyond the largest observation time, or even converge in the longer term. In the latter case, a lower risk reported in the treatment group from right-censored data would only indicate a delayed event rather than a long-term treatment gain. It motivated the *extended promotion time* model (Bremhorst and Lambert, 2016; Bremhorst et al., 2016)

$$S_p(t|\mathbf{v}, \tilde{\mathbf{v}}) = \exp\{-\vartheta(\mathbf{v})F(t|\tilde{\mathbf{v}})\}. \tag{2}$$

with a dynamics in the (population) hazard function changing with the covariates in $\tilde{\mathbf{v}}$. An accelerated failure time (AFT) or a proportional hazards (PH) model could be considered further to describe the dependence of F on $\tilde{\mathbf{v}}$.

The inclusion of time-varying covariates (TVCs) is challenging in that framework. It was studied in the mixture cure model formulation by Dirick et al. (2019) with TVCs restricted to the conditional survival model for non-cured subjects with, therefore, only constant covariates entering the logistic regression submodel for the cure probability. A more general formulation was proposed by Lambert and Bremhorst (2020) in the framework of the extended promotion time model with categorical TVCs affecting not only the event timing for non-cured subjects, but also entering the regression model for the cure probability. However, the follow-up duration after each change of covariate had to be long enough for identification purposes in the cure probability submodel.

In the current paper, a reformulation of the extended promotion time model allowing an unlimited number of changes in categorical or continuous TVCs is proposed and studied. Additive terms for nonlinear effects of (constant or time-varying) quantitative covariates can also be considered jointly in the long- and short-term survival submodels.

The plan of the paper is as follows. In Section 2, we propose a detailed recall of the extended promotion time model. The inclusion of additive terms using P-splines will also be discussed. The methodological core of the paper is in Section 3 with a novel proposal for the inclusion of categorical or quantitative time-varying covariates in a cure survival model. Algorithms to explore the joint posterior of the model parameters and to compute their posterior mode (MAP) are described in Section 3.2. A strategy for selecting penalty parameters tuning the smoothness of the unknown functionals in the long- and short-term survival sub-models is proposed with a simple to implement and fast converging algorithm. The merits of this proposal are evaluated by means of an extensive simulation study in Section 4. The methodology is illustrated in Section 5 with the analysis of pension register data and of the association between women's time-varying earnings and fertility transitions in Germany. We conclude the paper with a discussion in Section 6.

2 The extended promotion time model

The use of covariates to alter the dynamics in the nonparametric baseline hazard of the promotion time model was first explored by Bremhorst and Lambert (2016), see (2), with a log-linear model for $\vartheta(\mathbf{v})$ and PH model for $F(t|\tilde{\mathbf{v}})$. More specifically, consider the following formulation for the latter expression, $F(t|\tilde{\mathbf{v}}) = 1 - S_0(t)^{\exp(\eta_F(\tilde{\mathbf{v}}))}$ where $S_0(t) = 1 - F(t|\mathbf{0})$ is a baseline survival function and $\eta_F(\cdot)$ is a (possibly non-linear) function of the covariates with an identification constraint, for example $\eta_F(\mathbf{0}) = 0$. Then, the population cumulative hazard and hazard functions associated to (2) become, respectively,

$$H_p(t|\mathbf{v}, \tilde{\mathbf{v}}) = \vartheta(\mathbf{v})F(t|\tilde{\mathbf{v}}) = \exp\{\eta_\vartheta(\mathbf{v})\}\left(1 - S_0(t)^{\exp(\eta_F(\tilde{\mathbf{v}}))}\right) \ ; \tag{3}$$

$$h_p(t|\mathbf{v}, \tilde{\mathbf{v}}) = \vartheta(\mathbf{v})f(t|\tilde{\mathbf{v}}) = \mathrm{e}^{\eta_\vartheta(\mathbf{v})+\eta_F(\tilde{\mathbf{v}})}f_0(t)S_0(t)^{\exp(\eta_F(\tilde{\mathbf{v}}))-1} \ , \tag{4}$$

where $f_0(t) = -dS_0(t)/dt$. Identification issues can be solved provided that the follow-up is sufficiently long, even in the challenging case where some covariates are common to $\mathbf{v}$ and $\tilde{\mathbf{v}}$, see Lambert and Bremhorst (2019) for more details. For fixed given values $\tilde{\mathbf{v}}$ of the short-term survival covariates, (4) defines a proportional hazards model with a cured fraction $\exp(-\vartheta(\mathbf{v}))$ and non time-varying hazard ratios for contrasts corresponding to different values of $\mathbf{v}$ as $h_p(t|\mathbf{v}_2, \tilde{\mathbf{v}})/h_p(t|\mathbf{v}_1, \tilde{\mathbf{v}}) = \vartheta(\mathbf{v}_2)/\vartheta(\mathbf{v}_1)$. This is not true anymore when hazards are compared for different values of $\tilde{\mathbf{v}}$, as $h_p(t|\mathbf{v}, \tilde{\mathbf{v}}_2)/h_p(t|\mathbf{v}, \tilde{\mathbf{v}}_1) = \tilde{\vartheta}(\tilde{\mathbf{v}}_2)/\tilde{\vartheta}(\tilde{\mathbf{v}}_1)S_0(t)^{\tilde{\vartheta}(\tilde{\mathbf{v}}_2)-\tilde{\vartheta}(\tilde{\mathbf{v}}_1)}$ (with $\tilde{\vartheta}(\tilde{\mathbf{v}}) = \mathrm{e}^{\eta_F(\tilde{\mathbf{v}})}$) changes over time.

Bremhorst and Lambert (2016) considered a linear combination of B-splines to specify $h_0(t) = f_0(t)/S_0(t)$. Here, a flexible form based on P-splines (Eilers and Marx, 1996) is preferred for $f_0(\cdot)$,

$$f_0(t) = \frac{\exp\left(\sum_k b_k(t)\phi_k\right)}{\int_0^T \exp\left(\sum_k b_k(u)\phi_k\right) du}\mathbb{1}_{\{0\leq t\leq T\}} \tag{5}$$

where $\{b_k(\cdot) : k = 1, \ldots, K\}$ denotes a large B-splines basis associated to equidistant knots on $(0, T)$ and $\boldsymbol{\phi} = (\phi_k)_{k=1}^K$ is a vector of spline parameters with $\phi_{\lfloor K/2 \rfloor} = 0$ (for identification purposes). It is directly connected to the reference population hazard, $h_p(t|\mathbf{0}, \mathbf{0}) = \mathrm{e}^{\beta_0} f_0(t)$, with β_0 governing the total risk exposure and $f_0(t)$ its distribution over time. Smoothness will be forced on $f_0(\cdot)$ by penalizing changes in the spline coefficients, see Section 3.2. The possible nonlinear effects of continuous covariates (such as age or earnings on the probability of pregnancy and its timing in a fertility study) will be modelled using additive forms. Assume that n independent units are observed with data $\mathcal{D} = \cup_{i=1}^n \mathcal{D}_i$ where $\mathcal{D}_i = \{t_i, \delta_i, \mathbf{v}_i, \tilde{\mathbf{v}}_i\}$, δ_i is the event indicator for the follow-up time t_i and $\mathbf{v}_i, \tilde{\mathbf{v}}_i$ the long- and short-term survival covariates for unit i. We add flexibility to the extended promotion time model by considering nonlinear forms to describe the effects of quantitative covariates on $\eta_\vartheta(\mathbf{v})$ and $\eta_F(\tilde{\mathbf{v}})$,

$$\big(\eta_\vartheta(\mathbf{v}_i)\big)_{i=1}^n = \left(\beta_0 + \sum_{k=1}^p \beta_k z_{ik} + \sum_{j=1}^J f_j(x_{ij})\right)_{i=1}^n = \mathbf{Z}\boldsymbol{\beta} + \sum_{j=1}^J \mathbf{f}_j \ , \tag{6}$$

$$\big(\eta_F(\tilde{\mathbf{v}}_i)\big)_{i=1}^n = \left(\sum_{k=1}^{\tilde{p}} \gamma_k \tilde{z}_{ik} + \sum_{j=1}^{\tilde{J}} \tilde{f}_j(\tilde{x}_{ij})\right)_{i=1}^n = \tilde{\mathbf{Z}}\boldsymbol{\gamma} + \sum_{j=1}^{\tilde{J}} \tilde{\mathbf{f}}_j \ , \tag{7}$$

where $f_j(\cdot)$ and $\tilde{f}_j(\cdot)$ denote smooth additive terms quantifying the effect of the associated quantitative covariate on long- and short-term survival, respectively, $\mathbf{f}_j = \big(f_j(x_{ij})\big)_{i=1}^n$ and $\tilde{\mathbf{f}}_j = \big(\tilde{f}_j(\tilde{x}_{ij})\big)_{i=1}^n$ their values over units stacked in vectors, $\mathbf{Z}$ the $n \times (1 + p)$ design matrix with a column of 1's for the intercept and one column per additional categorical covariate, similarly for the $n \times \tilde{p}$ design matrix $\tilde{\mathbf{Z}}$ (but without the column of 1's given the absence of an intercept). Now consider a basis of $(L + 1)$ cubic B-splines $\{s^*_{j\ell}(\cdot)\}_{\ell=1}^{L+1}$ associated to equally spaced knots on the range $(x_j^{\min}, x_j^{\max})$ of values for x_j. They are recentered for identification purposes in the additive model using $s_{j\ell}(\cdot) = s^*_{j\ell}(\cdot) - \frac{1}{x_j^{\max} - x_j^{\min}} \int_{x_j^{\min}}^{x_j^{\max}} s^*_{j\ell}(u)du$ $(\ell = 1, \ldots, L)$. Similarly for the covariates associated to short-term survival, yielding recentered B-splines denoted by $\tilde{s}_{j\ell}(\cdot)$. Then, the additive terms in the conditional long-term and short-term survival sub-models can be approximated using linear combinations of such (recentered) B-splines, $\mathbf{f}_j = \left(\sum_{\ell=1}^L s_{j\ell}(x_{ij})\theta_{\ell j}\right)_{i=1}^n = \mathbf{S}_j\boldsymbol{\theta}_j$, $\tilde{\mathbf{f}}_j = \left(\sum_{\ell=1}^L \tilde{s}_{j\ell}(\tilde{x}_{ij})\tilde{\theta}_{\ell j}\right)_{i=1}^n = \tilde{\mathbf{S}}_j\tilde{\boldsymbol{\theta}}_j$, where $[\mathbf{S}_j]_{i\ell} = s_{j\ell}(x_{ij})$, $[\tilde{\mathbf{S}}_j]_{i\ell} = \tilde{s}_{j\ell}(\tilde{x}_{ij})$, $\big(\boldsymbol{\theta}_j\big)_\ell = \theta_{\ell j}$ and $\big(\tilde{\boldsymbol{\theta}}_j\big)_\ell = \tilde{\theta}_{\ell j}$. Hence, using vectorial notations, the expressions for the conditional long-term and short-term linear predictors in (6) and (7) can be rewritten as $\big(\eta_{\vartheta i} = \eta_\vartheta(\mathbf{v}_i)\big)_{i=1}^n = \boldsymbol{\mathcal{X}}\boldsymbol{\psi}$, $\big(\eta_{Fi} = \eta_F(\tilde{\mathbf{v}}_i)\big)_{i=1}^n = \tilde{\boldsymbol{\mathcal{X}}}\tilde{\boldsymbol{\psi}}$ with design matrices $\boldsymbol{\mathcal{X}} = [\mathbf{Z}, \mathbf{S}_1, \ldots, \mathbf{S}_J] = [\mathbf{Z}, \boldsymbol{\mathcal{S}}] \in \mathrm{I\!R}^{n \times q}$, $\tilde{\boldsymbol{\mathcal{X}}} = [\tilde{\mathbf{Z}}, \tilde{\mathbf{S}}_1, \ldots, \tilde{\mathbf{S}}_J] = [\tilde{\mathbf{Z}}, \tilde{\boldsymbol{\mathcal{S}}}] \in \mathrm{I\!R}^{n \times \tilde{q}}$; matrices of spline parameters (with one column per additive term) $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J]$ in $\mathrm{I\!R}^{L \times J}$, $\tilde{\boldsymbol{\Theta}} = [\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_{\tilde{J}}]$ in $\mathrm{I\!R}^{L \times \tilde{J}}$; vectors of (stacked) regression parameters $\boldsymbol{\psi} = \big(\boldsymbol{\beta}, \mathrm{vec}(\boldsymbol{\Theta})\big)$ in $\mathrm{I\!R}^q$, $\tilde{\boldsymbol{\psi}} = \big(\boldsymbol{\gamma}, \mathrm{vec}(\tilde{\boldsymbol{\Theta}})\big)$ in $\mathrm{I\!R}^{\tilde{q}}$, where $q = (1 + p + JL)$ and $q = (\tilde{p} + \tilde{J}L)$. Note that (6) corresponds to an additive log-log model for the cure probability, or equivalently, a complementary log-log model for the long-term event probability.

3 Cure model with time-varying covariates

3.1 Model specification

Assume now that the covariates are exogenous and can change values over time. Our proposal is to model the hazard rate at the population level using

$$\begin{aligned} h_p(t|\mathbf{v}(t), \tilde{\mathbf{v}}(t)) &= \vartheta(\mathbf{v}(t)) f(t|\tilde{\mathbf{v}}(t)) \\ &= \mathrm{e}^{\eta_\vartheta(\mathbf{v}(t)) + \eta_F(\tilde{\mathbf{v}}(t))} f_0(t) S_0(t)^{\exp(\eta_F(\tilde{\mathbf{v}}(t))-1}\ , \end{aligned} \tag{8}$$

yielding a (promotion time) cure survival model with time-varying covariates, shortly named the `TVcure` model. The associated cumulative hazard function can be obtained numerically using integration. In the special case where covariates are constant, we recover the expressions for the population hazard and cumulative hazard functions in Bremhorst and Lambert (2016), see (3) and (4) in Section 2, with the parameter interpretation already discussed. In the general case, the linear predictors η_ϑ and η_F not only change over units, but also potentially over time. Therefore, the associated design matrices also depend on time with $\big(\eta_\vartheta(\mathbf{v}_i(t))\big)_{i=1}^n = \boldsymbol{\mathcal{X}}_t\boldsymbol{\psi}$, $\big(\eta_F(\tilde{\mathbf{v}}_i(t))\big)_{i=1}^n = \tilde{\boldsymbol{\mathcal{X}}}_t\tilde{\boldsymbol{\psi}}$.

Further assume that data for each unit can be reported in a regular manner over time (measured in dt units of time), such as with the monthly report ($dt = 1$ month) of a woman status and her covariate values from age 20 ($t = 0$) till the event time (e.g. her first pregnancy) or the end of her follow-up period (if she is childless by that time). Then, the data for the ith unit would take the following form, $\mathcal{D}_i = \{(d_{it}, \mathbf{v}_i(t), \tilde{\mathbf{v}}_i(t)) : t = 1, \ldots, t_i\}$, where d_{it} is the event indicator identically equal to 0 for all t, except perhaps the last value d_{it_i} equal to one if $\delta_i = 1$ when an event is observed within $(t_i - dt, t_i)$, and zero otherwise.

3.2 Inference

Assuming that the covariates remain constant within a time unit dt, the conditional distribution of d_{it} for the ith subject still at risk at time $t - dt$ is approximately Poisson (Lindsey, 1995) with mean $\mu_{it} = h_p(t|\mathbf{v}(t), \tilde{\mathbf{v}}(t))\, dt$. The log-likelihood contribution for that subject is

$$\ell_i(\boldsymbol{\phi}, \boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}|\mathcal{D}_i) = -\sum_{t=1}^{t_i} \mu_{it} + d_{it_i} \log \mu_{it_i},$$

with the dependence of μ_{ij} on the three vectors of parameters made explicit in Eqs. (5) to (8). Smoothness priors for the spline parameters complete the model description to counterbalance the flexibility brought by the large B-spline bases in the specifications of

$f_0(t)$ and of the additive terms in (6) and (7) (Fahrmeir and Lang, 2001),

$$
\begin{aligned}
p(\boldsymbol{\phi}|\tau_0) &\propto \exp\left(-\frac{1}{2}\,\boldsymbol{\phi}^\top(\tau_0\mathbf{P}_0)\,\boldsymbol{\phi}\right)\;, \\
p(\boldsymbol{\theta}_j|\lambda_j) &\propto \exp\left(-\frac{1}{2}\,\boldsymbol{\theta}_j^\top(\lambda_j\mathbf{P})\,\boldsymbol{\theta}_j\right)\quad(j=1,\ldots,J)\;, \\
p(\tilde{\boldsymbol{\theta}}_j|\tilde{\lambda}_j) &\propto \exp\left(-\frac{1}{2}\,\tilde{\boldsymbol{\theta}}_j^\top(\tilde{\lambda}_j\tilde{\mathbf{P}})\,\tilde{\boldsymbol{\theta}}_j\right)\quad(j=1,\ldots,\tilde{J}).
\end{aligned}
\tag{9}
$$

with Gamma priors for the penalty parameters, $\tau_0, \lambda_j, \tilde{\lambda}_j \sim \mathcal{G}\left(1, d=10^{-4}\right)$ (Lang and Brezger, 2004), or mixture of Gammas (Jullion and Lambert, 2007) with, as special cases, half-Cauchy priors for the square-root of these parameters (Lambert and Bremhorst, 2019). Assuming joint Normal priors for the parameters associated to the other covariates $\mathbf{z}$ and $\tilde{\mathbf{z}}$, $\boldsymbol{\beta} \sim \mathcal{N}\left(\breve{b}, \mathbf{Q}^{-1}\right)$, $\boldsymbol{\gamma} \sim \mathcal{N}\left(\breve{\mathbf{g}}, \tilde{\mathbf{Q}}^{-1}\right)$, the joint priors for the regression and spline parameters in $\boldsymbol{\psi}$ and $\tilde{\boldsymbol{\psi}}$ induce Gaussian Markov random fields (GMRF) (Rue and Held, 2005) as they can be written as

$$
\begin{aligned}
p(\boldsymbol{\psi}|\boldsymbol{\lambda}) &\propto \exp\left(-\frac{1}{2}\,(\boldsymbol{\psi}-\mathbf{b})^\top\mathbf{K}_\lambda(\boldsymbol{\psi}-\mathbf{b})\right)\;; \\
p(\tilde{\boldsymbol{\psi}}|\tilde{\boldsymbol{\lambda}}) &\propto \exp\left(-\frac{1}{2}\,(\tilde{\boldsymbol{\psi}}-\mathbf{g})^\top\tilde{\mathbf{K}}_{\tilde{\lambda}}(\tilde{\boldsymbol{\psi}}-\mathbf{g})\right),
\end{aligned}
\tag{10}
$$

where $\mathbf{b} = (\breve{\mathbf{b}}^\top, \mathbf{0}_{JL}^\top)^\top$, $\mathbf{K}_\lambda = \mathrm{diag}\big(\mathbf{Q}, \boldsymbol{\mathcal{P}}_\lambda\big)$, $\boldsymbol{\mathcal{P}}_\lambda = \boldsymbol{\Lambda} \otimes \mathbf{P}$, $[\boldsymbol{\Lambda}]_{jj'} = \delta_{jj'}\lambda_j$, $\mathbf{g} = (\breve{\mathbf{g}}^\top, \mathbf{0}_{\tilde{J}L}^\top)^\top$, $\tilde{\mathbf{K}}_{\tilde{\lambda}} = \mathrm{diag}\big(\tilde{\mathbf{Q}}, \tilde{\boldsymbol{\mathcal{P}}}_{\tilde{\lambda}}\big)$, $\tilde{\boldsymbol{\mathcal{P}}}_\lambda = \tilde{\boldsymbol{\Lambda}} \otimes \tilde{\mathbf{P}}$ and $[\tilde{\boldsymbol{\Lambda}}]_{jj'} = \delta_{jj'}\tilde{\lambda}_j$. Let $\boldsymbol{\lambda} = (\lambda_j)_{j=1}^J$, $\tilde{\boldsymbol{\lambda}} = (\tilde{\lambda}_j)_{j=1}^{\tilde{J}}$ with joint priors $p(\boldsymbol{\lambda}) \propto \prod_j p(\lambda_j)$, $p(\tilde{\boldsymbol{\lambda}}) \propto \prod_j p(\tilde{\lambda}_j)$. If $\ell = \sum_{i=1}^n \ell_i$ denotes the log-likelihood, then the joint posterior for the model parameters directly follows from Bayes's theorem,

$$
\begin{aligned}
&p(\boldsymbol{\phi}, \boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}, \tau_0, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) \propto \\
&\qquad \exp\{\ell(\boldsymbol{\phi}, \boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}|\mathcal{D})\}\, p(\boldsymbol{\phi}|\tau_0)p(\tau_0)\, p(\boldsymbol{\psi}|\boldsymbol{\lambda})p(\boldsymbol{\lambda})\, p(\tilde{\boldsymbol{\psi}}|\tilde{\boldsymbol{\lambda}})p(\tilde{\boldsymbol{\lambda}}).
\end{aligned}
$$

3.2.1 Conditional estimation of the regression parameters

The conditional posterior mode of the regression and spline parameters, $\boldsymbol{\zeta} = (\boldsymbol{\psi}^\top, \tilde{\boldsymbol{\psi}}^\top)^\top$ and $\boldsymbol{\phi}$, coincide with their conditional penalized maximum likelihood estimates (PMLE) optimizing

$$
\ell_p = \ell - \frac{\tau_0}{2}\,\boldsymbol{\phi}^\top\mathbf{P}_0\,\boldsymbol{\phi} - \frac{1}{2}\,(\boldsymbol{\psi}-\mathbf{b})^\top\mathbf{K}_\lambda(\boldsymbol{\psi}-\mathbf{b}) - \frac{1}{2}\,(\tilde{\boldsymbol{\psi}}-\mathbf{g})^\top\tilde{\mathbf{K}}_{\tilde{\lambda}}(\tilde{\boldsymbol{\psi}}-\mathbf{g})\,.
$$

They can be obtained iteratively using the Newton-Raphson (N-R) algorithm) built upon explicit forms for their respective gradient and precision matrix, see Appendix A. Practically, given values for $\boldsymbol{\phi}$ and the penalty parameters $\boldsymbol{\tau} = (\boldsymbol{\lambda}^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$, repeat the following substitution till convergence:

$$
\boldsymbol{\zeta} \longleftarrow \boldsymbol{\zeta} - \mathcal{H}_{\tau}^{-1}\mathbf{U}_{\tau}^{\zeta}\;. \tag{11}
$$

Estimates for the spline parameters $\boldsymbol{\phi}$ defining $f_0(t)$ in (5) are obtained in a similar manner, iteratively and conditionally on the penalty parameter τ_0,

$$\boldsymbol{\phi}_{-k} \longleftarrow \boldsymbol{\phi}_{-k} - \left((\boldsymbol{H}^{\phi\phi}_{\tau_0})_{-k,-k}\right)^{-1} (\mathbf{U}^{\phi}_{\tau_0})_{-k} \ , \tag{12}$$

with, for identification purposes, one of the vector components arbitrarily set to zero, $\phi_k = 0$.

3.2.2 Selection of the penalty parameters

The marginal posterior for the penalty parameters $\boldsymbol{\tau} = (\boldsymbol{\lambda}^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$ tuning the smoothness the additive terms can be obtained using the following identity (with an implicit conditioning on $\boldsymbol{\phi}$ and τ_0),

$$p(\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) = p(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) / p(\boldsymbol{\psi}, \tilde{\boldsymbol{\psi}}|\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}, \mathcal{D}) \ ,$$

with a Laplace's approximation substituted to the conditional posterior of the spline parameters in the denominator, see Lambert (2021) for a similar strategy in nonparametric double additive location-scale models. Evaluating that expression at the conditional posterior modes yields the following approximation to the marginal posterior of the penalty parameters,

$$p(\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) \dot\propto p(\hat{\boldsymbol{\psi}}_\lambda, \hat{\tilde{\boldsymbol{\psi}}}_{\tilde{\lambda}}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) \left|\Sigma_\tau^{-1}\right|^{-1/2} \ , \tag{13}$$

where the blocks in the precision matrix

$$\Sigma_\tau^{-1} = -\mathcal{H}_\tau = \begin{bmatrix} -\mathbf{H}^{\psi\psi}_\lambda & -\mathbf{H}^{\psi\tilde{\psi}} \\ -\mathbf{H}^{\tilde{\psi}\psi} & -\mathbf{H}^{\tilde{\psi}\tilde{\psi}}_{\tilde{\lambda}} \end{bmatrix} \ , \tag{14}$$

have explicit forms, see Appendix A. Maximizing that marginal posterior enables to select the penalty parameters. Remembering that the normalizing constants of the priors for $\boldsymbol{\psi}$ and $\tilde{\boldsymbol{\psi}}$ in (10) depend on the penalty parameters, the log of (13) can be written (up to an additive constant) as

$$\begin{aligned}\log p(\boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}|\mathcal{D}) &\doteq \ell_p(\hat{\boldsymbol{\phi}}_{\tau_0}, \hat{\boldsymbol{\zeta}}_\tau|\tau, \mathcal{D}) + \frac{1}{2}\left\{\log|\mathbf{K}_\lambda|^+ + \log|\tilde{\mathbf{K}}_{\tilde{\lambda}}|^+ - \log|-\mathcal{H}_\tau|\right\} \\ &= \ell_p(\hat{\boldsymbol{\phi}}_{\tau_0}, \hat{\boldsymbol{\zeta}}_\tau|\tau, \mathcal{D}) + \frac{\rho(\mathbf{P})}{2}\sum_{j=1}^{J}\log\lambda_j + \frac{\rho(\tilde{\mathbf{P}})}{2}\sum_{j=1}^{\tilde{J}}\log\tilde{\lambda}_j - \frac{1}{2}\log|-\mathcal{H}_\tau| \ ,\end{aligned} \tag{15}$$

where $|\mathbf{A}|^+$ denotes the product of the non-zero eigenvalues of a semi-positive definite matrix $\mathbf{A}$. Let $\Sigma_\tau(\boldsymbol{\theta}_j)$ be the $L \times L$ submatrix in Σ_τ corresponding to the sub-vector $\boldsymbol{\theta}_j$ in $(\boldsymbol{\psi}^\top, \tilde{\boldsymbol{\psi}}^\top)^\top$. Given that

$$\frac{\partial \ell_p(\hat{\boldsymbol{\phi}}_{\tau_0}, \hat{\boldsymbol{\zeta}}_\tau|\tau, \mathcal{D})}{\partial\lambda_j} = \frac{\partial \ell_p(\hat{\boldsymbol{\phi}}_{\tau_0}, \hat{\boldsymbol{\zeta}}_\tau|\tau, \mathcal{D})}{\partial\hat{\boldsymbol{\psi}}_\lambda^\top}\frac{\partial\hat{\boldsymbol{\psi}}_\lambda}{\partial\lambda_j} - \frac{1}{2}\hat{\boldsymbol{\theta}}_{j\lambda}^\top\mathbf{P}\hat{\boldsymbol{\theta}}_{j\lambda} \quad (1 \le j \le J)$$

with the first factor in that expression equal to zero, and remembering that $\partial \log |A_s|/\partial s = \mathrm{Tr}\left(A_s^{-1}\partial A_s/\partial s\right)$ for a positive definite matrix A_s, one has

$$\frac{\partial \log p(\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}|\mathcal{D})}{\partial \lambda_j} = \frac{1}{2}\left\{\frac{\rho(\mathbf{P})}{\lambda_j} - \hat{\boldsymbol{\theta}}_{j\lambda}^{\top}\mathbf{P}\hat{\boldsymbol{\theta}}_{j\lambda} - \mathrm{Tr}\left(\Sigma_\tau(\boldsymbol{\theta}_j)\mathbf{P}\right)\right\}. \tag{16}$$

A similar expression (with $1 \le j \le \tilde{J}$) can be obtained for

$$\frac{\partial \log p(\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}|\mathcal{D})}{\partial \tilde{\lambda}_j} = \frac{1}{2}\left\{\frac{\rho(\tilde{\mathbf{P}})}{\tilde{\lambda}_j} - \hat{\tilde{\boldsymbol{\theta}}}_{j\tilde{\lambda}}^{\top}\tilde{\mathbf{P}}\hat{\tilde{\boldsymbol{\theta}}}_{j\tilde{\lambda}} - \mathrm{Tr}\left(\Sigma_\tau(\tilde{\boldsymbol{\theta}}_j)\tilde{\mathbf{P}}\right)\right\}. \tag{17}$$

The MAP estimate for $\boldsymbol{\lambda}$ and $\tilde{\boldsymbol{\lambda}}$ are the solutions of (16) and (17) set to zero for all j. This can be done using the fixed point method with the following substitutions iterated till convergence:

$$\begin{aligned}
\lambda_j^{-1} &\longleftarrow \frac{\hat{\boldsymbol{\theta}}_{j\lambda}^{\top}\mathbf{P}\hat{\boldsymbol{\theta}}_{j\lambda} + \mathrm{Tr}\left(\Sigma_\tau(\boldsymbol{\theta}_j)\mathbf{P}\right)}{\rho(\mathbf{P})} = \frac{\mathbb{E}(\boldsymbol{\theta}_j^{\top}\mathbf{P}\boldsymbol{\theta}_j|\boldsymbol{\lambda},\mathcal{D})}{\rho(\mathbf{P})} \qquad (1 \le j \le J)\\
\tilde{\lambda}_j^{-1} &\longleftarrow \frac{\hat{\tilde{\boldsymbol{\theta}}}_{j\tilde{\lambda}}^{\top}\tilde{\mathbf{P}}\hat{\tilde{\boldsymbol{\theta}}}_{j\tilde{\lambda}} + \mathrm{Tr}\left(\Sigma_\tau(\tilde{\boldsymbol{\theta}}_j)\tilde{\mathbf{P}}\right)}{\rho(\tilde{\mathbf{P}})} = \frac{\mathbb{E}(\tilde{\boldsymbol{\theta}}_j^{\top}\tilde{\mathbf{P}}\tilde{\boldsymbol{\theta}}_j|\tilde{\boldsymbol{\lambda}},\mathcal{D})}{\rho(\tilde{\mathbf{P}})} \qquad (1 \le j \le \tilde{J})\,.
\end{aligned} \tag{18}$$

The connection to conditional expectations in (18) results from the preceding Laplace approximations to $(\boldsymbol{\theta}|\boldsymbol{\lambda},\mathcal{D})$ and $(\tilde{\boldsymbol{\theta}}|\tilde{\boldsymbol{\lambda}},\mathcal{D})$. The combination of (11) and (18) leads to Algorithm 1 for the selection of $\boldsymbol{\tau}$ and the estimation of $\boldsymbol{\zeta}$ (for a given value of $\boldsymbol{\phi}$).

The selection of τ_0 tuning the smoothness of $f_0(t)$ proceeds in a similar way. Conditionally on the regression parameters, the maximization of $p(\tau_0|\mathcal{D})$ (approximated using the same type of arguments as with $p(\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}|\mathcal{D})$) can be made using the following substitution repeated till convergence:

$$\tau_0^{-1} \longleftarrow \frac{\hat{\boldsymbol{\phi}}_{\tau_0}^{\top}\mathbf{P}_0\hat{\boldsymbol{\phi}}_{\tau_0} + \mathrm{Tr}\left((-\mathbf{H}_{\tau_0}^{\phi\phi})^{-1}\mathbf{P}_0\right)}{\rho(\mathbf{P}_0)} = \frac{\mathbb{E}(\boldsymbol{\phi}^{\top}\mathbf{P}_0\boldsymbol{\phi}|\tau_0,\mathcal{D})}{\rho(\mathbf{P}_0)} \tag{19}$$

The combination of (12) and (19) leads to Algorithm 2 for the selection of τ_0 and the estimation of $\boldsymbol{\phi}$ (for a given value of $\boldsymbol{\zeta}$).

3.2.3 Global estimation algorithm

The double additive TVcure model with time-varying covariates can be fitted using Algorithm 3. It alternates (till convergence) the selection and estimation of parameters $\{\boldsymbol{\tau},\boldsymbol{\zeta}\}$ in the regression submodels, with that of the parameters $\{\tau_0,\boldsymbol{\phi}\}$ specifying the baseline short-term survival dynamics (for reference values of the categorical covariates and additive terms set to zero). Possible initial values to initiate the algorithm are 0 for all the spline and regression parameters in $\boldsymbol{\zeta}$ and $\boldsymbol{\phi}$, and moderately large values (100, say) for all the penalty parameters in $\boldsymbol{\tau}$ and τ_0. The procedure was implemented using pure R code in a package named `tvcure` maintained by the author. Convergence is fast and only takes a couple of seconds using a basic laptop computer.

Algorithm 1: Selection of $\boldsymbol{\tau}$ and estimation of $\boldsymbol{\zeta}$ (for given $\boldsymbol{\phi}$)

Input: Spline parameters $\boldsymbol{\phi}$ from the short-term survival submodel, data $\mathcal{D} = \cup_{i=1}^{n} \{(d_{it}, \mathbf{v}_i(t), \tilde{\mathbf{v}}_i(t)) : t = 1, \ldots, t_i\}$.

Output: Selected $\boldsymbol{\tau} = (\boldsymbol{\lambda}^\intercal, \tilde{\boldsymbol{\lambda}}^\intercal)^\intercal$ and estimated $\hat{\boldsymbol{\zeta}}_\tau$ (given $\boldsymbol{\phi}$).

Note: $\boldsymbol{\zeta} = (\boldsymbol{\psi}^\intercal, \tilde{\boldsymbol{\psi}}^\intercal)^\intercal$ with $\boldsymbol{\psi} = \big(\boldsymbol{\beta}, \mathrm{vec}(\boldsymbol{\Theta})\big)$, $\tilde{\boldsymbol{\psi}} = \left(\boldsymbol{\gamma}^\intercal, \mathrm{vec}(\tilde{\boldsymbol{\Theta}})^\intercal\right)^\intercal$

repeat
- **repeat**
  - Evaluate $\mathbf{U}_\tau^\zeta$ and $\mathcal{H}_\tau$ using (21), (22) & (23).
  - Update $\boldsymbol{\zeta} \longleftarrow \boldsymbol{\zeta} - \mathcal{H}_\tau^{-1} \mathbf{U}_\tau^\zeta$
- **until** $||\mathbf{U}_\tau^\zeta||_\infty < \epsilon$;
- **repeat**
  - $\lambda_j^{-1} \longleftarrow \left\{\boldsymbol{\theta}_j^\intercal \mathbf{P} \boldsymbol{\theta}_j + \mathrm{Tr}\left(\Sigma_\tau(\boldsymbol{\theta}_j)\mathbf{P}\right)\right\} / \rho(\mathbf{P}) \quad (1 \leq j \leq J)$
  - $\tilde{\lambda}_j^{-1} \longleftarrow \left\{\tilde{\boldsymbol{\theta}}_j^\intercal \tilde{\mathbf{P}} \tilde{\boldsymbol{\theta}}_j + \mathrm{Tr}\left(\Sigma_\tau(\tilde{\boldsymbol{\theta}}_j)\tilde{\mathbf{P}}\right)\right\} / \rho(\tilde{\mathbf{P}}) \quad (1 \leq j \leq \tilde{J})$
  - Update $\Sigma_\tau = (-\mathcal{H}_\tau)^{-1}$ using (21), (22) & (23).
- **until** *convergence*;

until *convergence*;

Algorithm 2: Selection of τ_0 and estimation of $\boldsymbol{\phi}$ (for given $\boldsymbol{\zeta}$)

Input: Regression and spline parameters $\boldsymbol{\zeta}$, data $\mathcal{D} = \cup_{i=1}^{n} \{(d_{it}, \mathbf{v}_i(t), \tilde{\mathbf{v}}_i(t)) : t = 1, \ldots, t_i\}$.

Output: Selected τ_0 and estimated $\hat{\boldsymbol{\phi}}_{\tau_0}$ (given $\boldsymbol{\zeta}$).

repeat
- **repeat**
  - Evaluate $\mathbf{U}_{\tau_0}^\phi$ and $\mathbf{H}_{\tau_0}^{\phi\phi}$ using (20)
  - Update $\boldsymbol{\phi}_{-k} \longleftarrow \boldsymbol{\phi}_{-k} - \left((\boldsymbol{H}_{\tau_0}^{\phi\phi})_{-k,-k}\right)^{-1} (\mathbf{U}_{\tau_0}^\phi)_{-k}$ with $\phi_k = 0$.
- **until** $||\mathbf{U}_{\tau_0}^\phi||_\infty < \epsilon$;
- **repeat**
  - Update $\tau_0^{-1} \longleftarrow \left\{\boldsymbol{\phi}^\intercal \mathbf{P}_0 \boldsymbol{\phi} + \mathrm{Tr}\left((-\mathbf{H}_{\tau_0}^{\phi\phi})^{-1}\mathbf{P}_0\right)\right\} / \rho(\mathbf{P}_0)$
  - Update $\mathbf{H}_{\tau_0}^{\phi\phi}$ using (20).
- **until** *convergence*;

until *convergence*;

Algorithm 3: Fitting the double additive cure model with time-varying covariates (TVcure)

Goal: Fit of the TVcure model described in Section 3.

Input: Data $\mathcal{D} = \cup_{i=1}^{n} \{(d_{it}, \mathbf{v}_i(t), \tilde{\mathbf{v}}_i(t)) : t = 1, \ldots, t_i\}$ and a TVcure model specification.

Output: Estimates for $\boldsymbol{\zeta} = (\boldsymbol{\psi}^\top, \tilde{\boldsymbol{\psi}}^\top)^\top$ and $\boldsymbol{\phi}$ for the selected penalty parameters $\boldsymbol{\tau} = (\boldsymbol{\lambda}^\top, \tilde{\boldsymbol{\lambda}}^\top)^\top$ and τ_0.

repeat

- Select τ_0 and estimate $\boldsymbol{\phi}$ (given $\boldsymbol{\zeta}$ and $\mathcal{D}$) using Algorithm 2.
- Select $\boldsymbol{\tau}$ and estimate $\boldsymbol{\zeta}$ (given $\boldsymbol{\phi}$ and $\mathcal{D}$) using Algorithm 1.

until *convergence*;

4 Simulation study

A simulation study was setup to evaluate the ability of the algorithms described in Section 3 to estimate the different ingredients of the TVcure model from right-censored data, including the additive terms and the reference cumulative hazard function $\mathrm{e}^{\beta_0} F_0(t)$. With the application from Section 5 in mind, $S = 500$ datasets of size $n = 500$ or 1500 were simulated using the data generating mechanism corresponding to the extended promotion time model with population hazard function (4), where $F_0(t) = 1 - S_0(t)$ with $S_0(t)$ given by the survival function of a Weibull with shape parameter 2.65 and scale parameter 133. The regression parameters in (6) and (7) were taken to be $\beta_0 = 0$, $\beta_1 = -.1$, $\beta_2 = .15$ and $\gamma_1 = .1$, $\gamma_2 = .2$, with independent Bernoulli or Normally distributed covariates, $z_1, z_3 \sim .5\,\mathrm{Bern}(.5)$, $z_2, z_4 \sim \mathcal{N}(0,1)$. The following additive terms were considered,

$$f_1(x_1) = -1.14 + 2.4x_1 - .88x_1^2 \;;\; f_2(x_2) = -.3\cos(2\pi x_2)$$
$$\tilde{f}_1(x_1) = .15 - .5\cos(\pi(x_1 - .75)) \;;\; \tilde{f}_2(x_3) = .6(x_3 - .5)$$

with $x_1 \sim \mathrm{Uniform}(0, 1.5)$ shared by the long- and short-term survival submodels, while $x_2, x_3 \sim \mathrm{Uniform}(0,1)$ are independent covariates specific to each of the preceding submodels. Such data can be generated using the biological motivation of the promotion time model given by Yakovlev and Tsodikov (1996) with $(n_i|\mathbf{v}_i) \sim \mathrm{Pois}(\vartheta(\mathbf{v}_i))$, $y_i = +\infty$ if $n_i = 0$ (in which case unit i is 'cured') and $y_i = \min\{\breve{y}_{im} : m = 1, \ldots, n_i\}$ otherwise, where $(\breve{y}_{im}|\tilde{\mathbf{v}})$ are independently and identically distributed random variables with c.d.f. $F(\cdot|\tilde{\mathbf{v}})$. Independent censoring times were generated using a Uniform on $(120, 299)$ (Scenario 1) or $(60, 299)$ (Scenario 2), yielding an observed (large) right-censoring rate close to 44% or 51%, including a marginal (unknown) cure rate of 40%. If y_i denotes the generated non-censored response for unit i ($i = 1, \ldots, n$), then the observed response is (t_i, δ_i) where $t_i = \min\{y_i, c_i\}$ and $\delta_i = I(y_i < c_i)$ is the event indicator. These data are transformed in a person-month format in a second step, see Section 3.2, with $dt = 1$ (month) and event indicator d_{it} for a unit i still at risk after a follow-up of t months ($t = 1, \ldots, 299$).

Table 1: Simulation study (Scenario 2): estimation of the additive terms in the long- and short-term submodels. Mean absolute bias, Root mean integrated squared error (RMISE) and mean effective coverage of 95% pointwise credible intervals.

Censoring	n		$f_1(x)$	$f_2(x)$	$\tilde{f}_1(x)$	$\tilde{f}_2(x)$
Scenario 1	500	MA-Bias	0.097	0.087	0.131	0.068
		RMISE	0.127	0.112	0.174	0.097
		Coverage	81.2	90.7	87.3	96.0
	1500	MA-Bias	0.053	0.053	0.070	0.042
		RMISE	0.071	0.069	0.092	0.057
		Coverage	93.1	93.3	94.5	96.0
Scenario 2	500	MA-Bias	0.107	0.091	0.110	0.073
		RMISE	0.139	0.118	0.186	0.105
		Coverage	80.2	90.5	86.6	96.2
	1500	MA-Bias	0.057	0.055	0.075	0.040
		RMISE	0.077	0.072	0.010	0.056
		Coverage	93.0	93.4	94.8	96.7

The estimated additive terms for the $S = 500$ datasets and their average values over these S replicates can be found in Fig. 1 for the least favorable setting (Scenario 2 when $n = 500$) and with $K = 10$ (penalized) B-splines associated to equidistant knots spanning the observed range for the associated covariate. The shape of the additive terms is quite well estimated despite the moderate sample size, the large proportion of right-censored data and the sophistication of the model. Summary information on the quality of the estimation of the additive terms are reported in Table 1. Not surprisingly, the reconstruction improves with the sample size and with decreasing right censoring rates. The mean absolute bias and the root mean integrated squared error (RMISE) decrease with sample size, proportionally to $n^{-1/2}$ for the RMISE. The effective (mean) coverages of pointwise 95% credible intervals for the additive terms are close to their nominal value. Similar conclusions can be reached for the estimation of the regression parameters, see Table 2. Finally, the estimates of the standardized cumulative hazard function $F_0(t)$ for the S simulated datasets of size $n = 500$ can be found in Fig. 2 with also very satisfactory results.

5 Application: women's earnings and fertility in Germany

In this section we illustrate how the method works with real data. For this purpose, we draw on register data from Germany and study how women's earnings relate to first birth behavior. This topic is an ideal test case to illustrate the method that was developed in this paper. First, Germany is a country with one of the highest shares of ever childless women in Europe (Kreyenfeld and Konietzka, 2017). Thus, there is always a high 'cure fraction' in each cohort. Second, Germany has enacted several major family policy

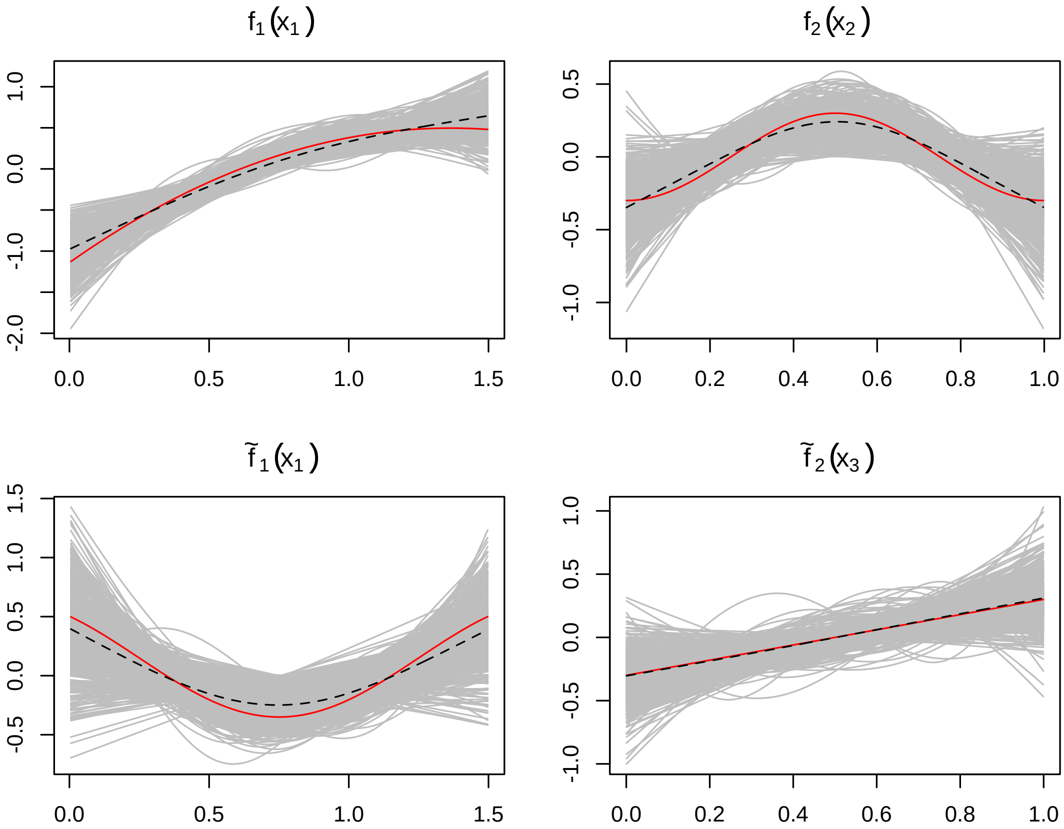


Figure 1: Simulation study ($n = 500$ - Scenario 2): Estimated additive terms (in grey) for each of the $S = 500$ datasets and their average values (dashed curves) over the S replicates with the solid curves (in red) corresponding to the 'true' additive functions.

reforms in recent years. A very important reform has been the parental leave benefit reform, enacted in 2007. Compared to prior regulations, it sets stronger incentives than before to establish in the labour market before having children. Thus, one might expect the association between women's earnings and fertility to be stronger for younger than for older cohorts.

The data of interest is a random sample from the German Pension registers of years 2017 and 2019 (Data extract SUFVSKT 2017/SUFVSKT 2019). The German pension registers cover roughly 90% of the resident population in Germany. Certain professions (farmers, lawyers) and civil servants are not included in the data. For the subsequent analysis, we have limited the investigation to West German women of the birth cohorts 1950-74. East Germany is eliminated from this study, as fertility patterns in the two parts of the country were rather different, in particular before reunification. We focus furthermore on women who did not have any children yet at age 20. Thus, teenage fertility is not part of this investigation. The sample size comprises 15,248 women and

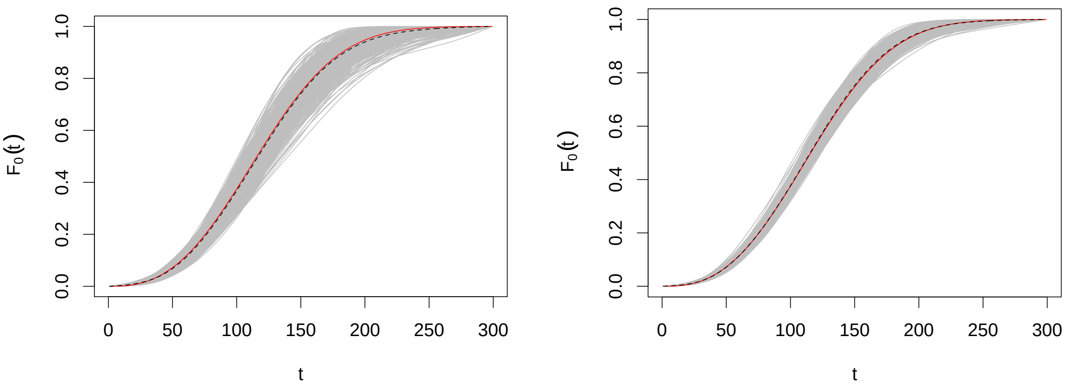


Figure 2: Simulation study (Scenario 2): Estimated cumulative baseline hazard function $F_0(t)$ for each of the $S = 500$ datasets and their average values (dashed curves) over the S replicates with the solid curves (in red) corresponding to the 'true' $F_0(t)$ (Left panel: $n = 500$; Right panel: $n = 1500$).

11,019 first births.

The outcome of interest, the starting month of the pregnancy, was calculated as the birth date of the first child minus 9 months. The follow-up considered for each woman started at the age of 20 until at most 45 with a possible interruption at the first pregnancy or due to a loss of follow-up (i.e. right-censoring) with, in the latter case, an uncertainty on the 'cure' (i.e. childless) final status of the person. Thus, in this fertility context, a woman will be considered 'cured' if she doesn't have a child by age 45. The dataset also contains complete monthly employment and earning histories of women. Earnings are stored as earning points, where one earning point represents the average annual earnings in a given year. Note that we only observe earnings during regular employment. Regular employment is employment that results into pension credits to the German pension fund. Women who do not receive any earnings from regular employment enter the analysis with 'zero' earnings. These women may be studying, unemployed or out of the labour market for other reasons. For the main parts of the investigation, we use earnings as a continuous covariate. A set of time-varying 'employment variables' control for whether a woman is studying, employed, registered unemployed or not in regular employment for other reasons. We also control for birth cohort where we distinguish between cohorts born 1950-54, 1955-59, 1960-64, 1965-69, 1970-74. Table 3 includes descriptive statistics by groups of birth cohorts. The women described here as 'right-censored' had their follow-up interrupted before their first potential pregnancy by age 45: thus, there is uncertainty about their maternal status at age 45. The table suggests that the proportion of West German women childless by age 45 is growing, starting from about 21% for the 1950-55 cohort to 28% for the 1970-74 cohort. Uncertainty remains, however, given the decreasing percentage of complete follow-ups.

For robustness checks, we also conducted an additional analysis where we grouped earnings into four categories $(0, .33]$ ('Low'), $(.33, .66]$ ('Medium'), $(.66, 1.0)$ ('High') and

Table 2: Simulation study (Scenario 2): estimation of regression parameters in the long- and short-term submodels. Bias, RMSE and effective coverage of 95% credible intervals.

			β_0	β_1	β_2	γ_1	γ_2
Censoring	n		0.000	-0.100	0.150	0.100	0.200
Scenario 1	500	Bias	0.007	-0.006	0.010	0.007	0.005
		RMSE	0.102	0.065	0.064	0.083	0.078
		Coverage	89.4	92.8	93.8	81.6	95.8
	1500	Bias	-0.008	0.000	0.001	0.000	0.003
		RMSE	0.053	0.035	0.037	0.045	0.045
		Coverage	93.2	94.4	93.8	85.4	96.0
Scenario 2	500	Bias	0.007	-0.008	0.006	0.016	0.002
		RMSE	0.110	0.067	0.068	0.089	0.085
		Coverage	90.4	94.4	94.2	83.8	95.4
	1500	Bias	-0.011	0.002	0.000	0.002	0.005
		RMSE	0.061	0.038	0.037	0.046	0.051
		Coverage	91.6	94.4	95.4	85.4	93.8

Table 3: Summary statistics on the follow-up of West German women's cohorts and the number of months between their 20th birthday and their first pregnancy (Source: SUFVSKT 2018).

		Person	Mother by Age 45		
Cohort	n	months	Yes	No	Right-cens.
1950-54	2423	277668	1900 (78.4%)	519 (21.4%)	4 (0.2%)
1955-59	2388	313692	1762 (73.8%)	613 (25.7%)	13 (0.5%)
1960-64	3029	431144	2186 (72.2%)	797 (26.3%)	46 (1.5%)
1965-69	3385	511119	2410 (71.2%)	887 (26.2%)	88 (2.6%)
1970-74	4023	655489	2761 (68.6%)	1130 (28.1%)	132 (3.3%)

(1.00, 1.5] ('Top'). As earnings are measured in earning points, the cut-point represent women who earn less than 33% of average earnings, those between 33-66%, those between 66% and average earnings. The last category comprises women who earn more than average earnings. Figure 3 plots the distribution of the categorized outcome variable. As expected, a large fraction of the younger women (ages 20–24) are in education. Noteworthy is the relatively high share of women who do not receive any earnings from regular employment. Further, only a small share of 25% earns more than average at age 30 and older.

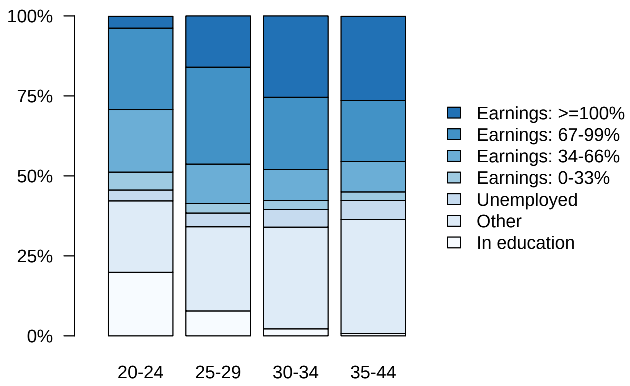


Figure 3: Earnings and employment status of nulliparous women, cohorts 1950-74, West Germany.

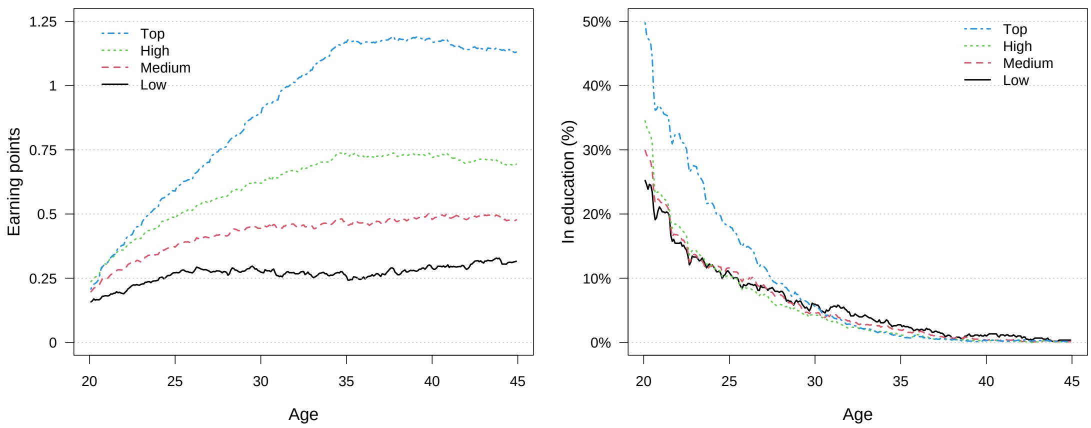


Figure 4: Age-earning trajectories of nulliparous women by highest earnings at age 35-44, cohorts 1950-74, West Germany.

It is clear that the earning trajectories of women greatly differ, depending on the career track. In order to illustrate the differences in women's earning profiles, Figure 4 plots the average earnings (left panel) and the share of women in education (right panel)

Table 4: TVcure models fitted to the West German women's cohort data: `Status` refers to the categorical covariate indicating the employment status and `s(Earnings)` to an additive earnings effect on long-term (Quantum) and/or short-term (Timing) survival.

		Model						
Submodel	Covariates	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$	$\mathcal{M}_7$
Quantum	`Status`	✓	✓	✓	✓	✓	–	–
	`s(Earnings)`	✓	✓	–	–	–	–	–
Timing	`Status`	✓	✓	✓	✓	–	✓	–
	`s(Earnings)`	✓	–	✓	–	–	–	–

by their highest earnings at age 35-44. The figure shows that the age-earning profiles of the women who eventually reach high earnings are flat at early ages, but fairly steep at advanced ages. Differences in earnings at later ages can largely be attributed to differences in educational participation (right panel). This aspect is of relevance, when prototypical trajectories for monthly earnings and employment status will be used below to illustrate the model results.

Analyses based on the TVcure model described in Section 3 were made separately for the groups of birth cohorts mentioned above. Seven TVcure models were fitted to the data, see Table 4 for the included covariates in the additive submodels for long-term (Quantum) and short-term (Timing) survival. For example, model $\mathcal{M}_1$ assumes joint effects of the employment status and earnings both on event quantum and timing. The model fit measured by the deviance, $D = 2\sum_{i=1}^{n}\sum_{j=1}^{n_i}\{d_{ij}\log(d_{ij}/\mu_{ij}) - (d_{ij} - \mu_{ij})\}$, and the model complexity quantified by the effective degrees of freedom (EDF) given by the trace of $\mathbf{H}_{\tau}^{-1}\mathbf{H}_0$ (Hastie and Tibshirani, 1990), can be combined to obtain the Akaike Information Criterion, AIC $= D + 2\,\mathrm{EDF}$. A model selection relying on the AIC, see Table 5, suggests a significant effect of employment status (working, studying, unemployed or other reasons for not being in the labour market) on the probability to have a first child and on the timing of the pregnancy for all cohorts. There is also a statistically significant association between female earnings and first birth timing and quantum, with the exception of the 1960-64-cohorts. For the latter cohorts, earnings do not affect fertility tempo. Figure 5 plots the estimates for model $\mathcal{M}_1$ separately for each cohort. The first row displays the first birth pattern of a woman with half the average gross earnings. It shows clearly how the fertility schedule has shifted across birth cohorts. First birth has been postponed, at the same time the distribution has become wider, suggesting greater heterogeneity in the age at first parenthood. The estimates of the additive terms for earnings in the quantum and timing are displayed in the 2nd and 3rd columns of Fig. 5. These results suggest that employed West German women in the 1950's who had low earnings were more likely to become mothers at a younger age than those who earned more. The association flips across cohorts, with the 1960-64 cohort playing a pivotal role. For the recent cohorts, the association is now positive, with low earnings reducing the chances of having a first child. The patterns for the

Table 5: Deviance, effective degrees of freedom (EDF) and Akaike Information Criterion (AIC) for the TVcure models in Table 4

	Cohort 1950-54			Cohort 1955-59		
Model	Deviance	EDF	AIC	Deviance	EDF	AIC
$\mathcal{M}_1$	18055.72	14.2	18084.19	17460.17	14.8	17489.84
$\mathcal{M}_2$	18069.87	11.0	18091.81	17469.91	11.1	17492.07
$\mathcal{M}_3$	18082.72	10.9	18104.45	17491.76	11.0	17513.82
$\mathcal{M}_4$	18125.19	7.0	18139.19	17518.86	7.0	17532.86
$\mathcal{M}_5$	18134.80	4.0	18142.80	17528.57	4.0	17536.57
$\mathcal{M}_6$	18145.99	4.0	18153.99	17581.49	4.0	17589.49
$\mathcal{M}_7$	18209.75	1.0	18211.75	17644.05	1.0	17646.05

	Cohort 1960-64			Cohort 1965-69		
Model	Deviance	EDF	AIC	Deviance	EDF	AIC
$\mathcal{M}_1$	22247.38	15.4	22278.24	25007.62	15.7	25038.94
$\mathcal{M}_2$	22250.92	11.4	22273.80	25019.54	11.7	25042.88
$\mathcal{M}_3$	22263.40	11.3	22285.94	25017.12	11.3	25039.80
$\mathcal{M}_4$	22274.21	7.0	22288.21	25034.64	7.0	25048.64
$\mathcal{M}_5$	22288.15	4.0	22296.15	25045.19	4.0	25053.19
$\mathcal{M}_6$	22347.23	4.0	22355.23	25093.10	4.0	25101.10
$\mathcal{M}_7$	22491.47	1.0	22493.47	25288.45	1.0	25290.45

	Cohort 1970-74		
Model	Deviance	EDF	AIC
$\mathcal{M}_1$	29357.81	16.4	29390.52
$\mathcal{M}_2$	29374.80	11.8	29398.33
$\mathcal{M}_3$	29387.04	11.5	29410.10
$\mathcal{M}_4$	29404.60	7.0	29418.60
$\mathcal{M}_5$	29436.07	4.0	29444.07
$\mathcal{M}_6$	29468.11	4.0	29476.11
$\mathcal{M}_7$	29660.88	1.0	29662.88

younger cohorts (1965-69 and 1970-74) are very similar, with an increasing influence of earnings on the decision to have a first child, with some delay for the better-off.

An approach based on TVcure models with a categorized version of monthly earnings, `EarnCat` (for classification, see above), was also explored for robustness checks. Four TVCure models with or without `EarnCat` in the quantum or the timing submodels were fitted separately for each cohort of women, with deviance, EDF and AIC also computed. Deviances (not reported here to save space) are, unsurprisingly, larger than those obtained with the continuous version of earnings given the loss of information resulting from categorization. However, qualitative conclusions are coherent with significant joint effects of `EarnCat` on the probability to have a first child and on the timing of the pregnancy for all cohorts. The analysis shows that the economic prerequisite for having children have shifted for the recent cohorts. While periods of low earnings increased "fertility quantum" for the older cohorts, it is rather vice versa for the younger cohorts.

So far, the analysis has focused on the association between earnings and first birth behavior. We have ignored that some women may have low incomes at an early age, while they may earn high wages later in life. To account for that, we estimated the conditional probability of being pregnant for four 'prototypical' earning trajectories using the fitted TVCure models. We have selected the following four scenarios, its descriptive name referring to the highest earning value at age 35-44 (see also the top left graph of Fig. 6):

1. 'Low': a woman with low earning trajectory studying until age 20 (solid line);
2. 'Medium': a woman with medium earning trajectory after studying until age 22 (dashed line);
3. 'High': a woman who completes her education at age 25 and then moves to the high earning trajectory (dotted line);
4. 'Top': A woman who completes her education at age 27 and moves to the top earning trajectory (dashed-dotted line).

We have displayed the estimates by birth cohort and earning group (top and bottom panels of Fig. 6, respectively). It shows how the association of women's earning and first birth progression has shifted across birth cohorts. While there were large fertility differences by women's earning profiles for the older cohorts, patterns have become more similar across time. We have also flipped the figure around and displayed the developments across cohorts within earning groups (lower panel of Fig. 6). While the fertility schedule is relatively stable across cohorts for high and top earning trajectories, first birth is increasingly postponed for low- and medium-income women.

6 Discussion

The proposed methodology extends cure survival models by enabling the inclusion of time-varying covariates in the quantum and timing additive submodels. They are not restricted to the conditional survival model for non-cured subjects as in Dirick et al.

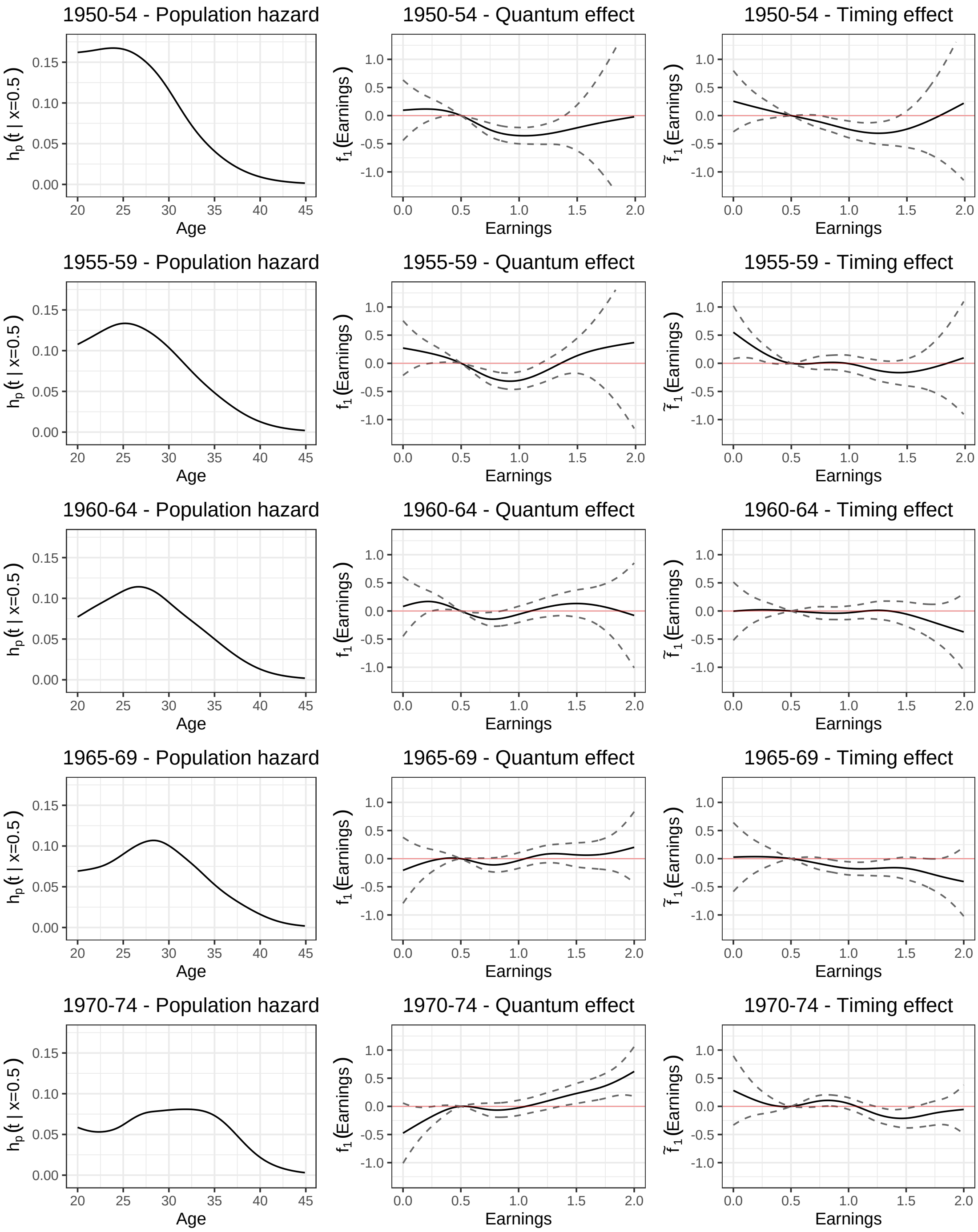
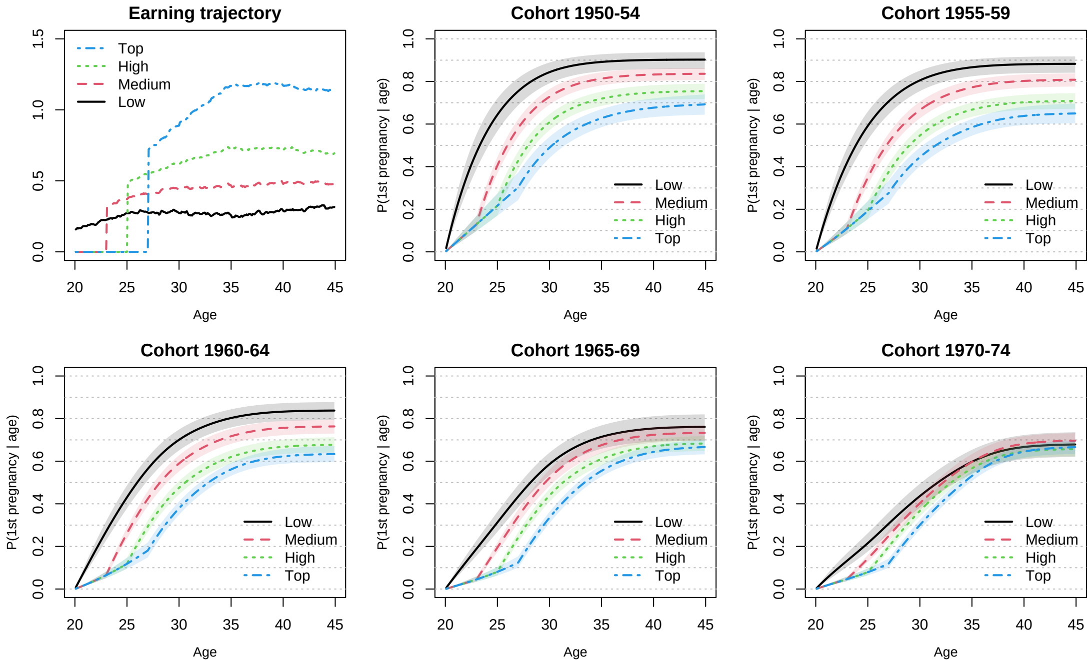


Figure 5: Estimated reference hazard $e^{\beta_0} f_0(t)$ and earnings effects on the quantum and timing of a first pregnancy with earnings reference value set at one-half of its annual average value in model $\mathcal{M}_1$.

Differences by earning trajectory within cohort groups



Differences by cohorts within earning trajectory groups

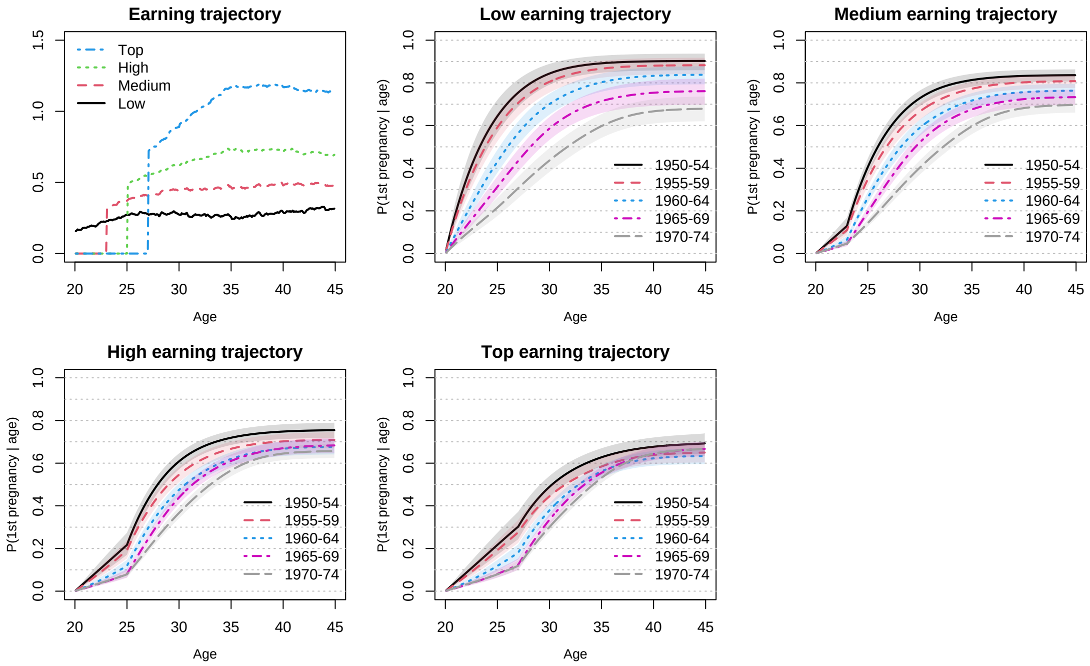


Figure 6: Estimated probability of a first pregnancy of West German women by prototypical earning profiles and birth cohort.

(2019), nor are the number of changes in covariate values limited as in Lambert and Bremhorst (2020) where only a small number of exclusively categorical covariate updates were allowed. Our new specification enables the inclusion of time-varying quantitative explanatory variables such as frequently changing monthly earnings in the fertility application. Smooth nonlinear forms can also be assumed for these effects with values for the penalty parameter $\boldsymbol{\tau}$ associated to P-splines automatically selected using Laplace-based approximations to the marginal posterior distributions of $\boldsymbol{\tau}$.

In a last step of the investigation, we used register data for Germany to illustrate how the method worked with real data. We examined whether the effect of women's earnings on first birth behavior had changed across cohorts in Germany. Data and research question seemed an ideal test case to showcase how the method unfolds in practice. There is a high share of women who remain childless in Germany. Thus, there is a sizeable cure fraction for each cohort. Further, Germany has enacted major policy reforms in the last decades. Particularly the parental leave benefit reform, enacted in 2007, set stronger incentives than before to postpone childbearing until one had reached "sufficient" earnings. It was expected that it would lower childlessness among the highly educated and career-oriented women in Germany. Our investigation, that separated timing and quantum, provides important and social policy relevant evidence on the matter. We indeed find that high earnings used to increase childlessness in the old cohorts, while we no longer find such a relationship for the younger cohorts. The analysis shows that the association between women's earnings and first birth has changed across birth cohorts. While low earnings used to accelerate fertility timing and reduce childlessness among the older cohorts, we do not find the same patterns anymore for the recent cohorts. Indeed, levels of childlessness are very similar now, regardless of earnings. Insufficient female earnings seem to increasingly delay childbearing in Germany. This pattern may be attributed to the policy reform and the introduction of the earnings-related parental leave benefit which set strong incentives to postpone parenthood until one had gathered earnings that resulted into adequate parental leave benefits.

Acknowledgments

The first author acknowledges the support of the ARC project IMAL (grant 20/25-107) financed by the Wallonia-Brussels Federation and granted by the Académie universitaire Louvain.

References

Berkson, J. and R. Gage (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association 47*, 501–515.

Boag, J. (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society. Series B (Methodological) 11*(1), 11–53.

Bremhorst, V., M. Kreyenfeld, and P. Lambert (2016). Fertility progression in Germany: An analysis using flexible nonparametric cure survival models. *Demographic Research 35*, 505–534.

Bremhorst, V. and P. Lambert (2016). Flexible estimation in cure survival models using Bayesian P-splines. *Computational Statistics and Data Analysis 93*, 270–284.

Chen, M.-H., J. Ibrahim, and D. Sinha (1999). A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association 94*(447), 909–919.

Dirick, L., T. Bellotti, G. Claeskens, and B. Baesens (2019). Macro-economic factors in credit risk calculations: including time-varying covariates in mixture cure models. *Journal of Business and Economic Statistics 37*, 40–53.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science 11*, 89–102.

Fahrmeir, L. and S. Lang (2001). Bayesian Inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society, Series C (Applied Statistics) 50*, 201–220.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized Additive Models.* London: Chapman & Hall.

Jullion, A. and P. Lambert (2007). Robust specification of the roughness penalty prior distribution in spatially adaptive Bayesian P-splines models. *Computational Statistics and Data Analysis 51*(5), 2542–2558.

Kreyenfeld, M. and D. Konietzka (2017). Analyzing childlessness. In M. Kreyenfeld and D. Konietzka (Eds.), *Childlessness in Europe: Contexts, Causes, and Consequences*, pp. 3–15. Springer International Publishing.

Lambert, P. (2021). Fast Bayesian inference using Laplace approximations in nonparametric double additive location-scale models with right- and interval-censored data. *Computational Statistics and Data Analysis 161*, 107250.

Lambert, P. and V. Bremhorst (2019). Estimation and identification issues in the promotion time cure model when the same covariates influence long- and short-term survival. *Biometrical Journal 61*(2), 275–289.

Lambert, P. and V. Bremhorst (2020). Inclusion of time-varying covariates in cure survival models with an application in fertility studies. *J. R. Statist. Soc. A 183*, 333–354.

Lang, S. and A. Brezger (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics 13*, 183–212.

Lindsey, J. K. (1995). Fitting parametric counting processes by using log-linear models. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 44*, 201–212.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC.

Tsodikov, A. (1998). A proportional hazard model taking account of long-term survivors. *Biometrics 54*, 1508–1516.

Yakovlev, A. and A. Tsodikov (1996). *Stochastic Models for Tumor of Latency and Their Biostatistical Applications*. World Scientific Publishing Singapore.

Appendices

A Closed form expressions for $\mathbf{U}_\lambda$ and $\mathbf{H}_\lambda$

Let $b_{tk} = b_k(t)$, $\pi_t = f_0(t)\,dt = \exp\left(\sum_{k=1}^K b_{tk}\phi_k\right)/\sum_{s=1}^T \exp\left(\sum_{k=1}^K b_{sk}\phi_k\right)$, $h_{it} = h_p(t|\mathbf{v}_i(t), \tilde{\mathbf{v}}_i(t))$, $F_0(t) = \sum_{s\leq t}\pi_t$ and $S_0(t) = 1 - F_0(t)$. For the spline parameters involved in $f_0(t)$, the conditional score and precision matrix are

$$\begin{aligned}
\mathbf{U}^{\phi}_{\tau_0} &= \frac{\partial \ell_p}{\partial \boldsymbol{\phi}} = \sum_{i=1}^{n}\sum_{t=1}^{t_i}(d_{it} - \mu_{it})\frac{\partial \log h_{it}}{\partial \boldsymbol{\phi}} - \tau_0 \mathbf{P}\boldsymbol{\phi} \\
-\mathbf{H}^{\phi\phi}_{\tau_0} &= \sum_{i=1}^{n}\sum_{t=1}^{t_i}\mu_{it}\frac{\partial \log h_{it}}{\partial \boldsymbol{\phi}}\frac{\partial \log h_{it}}{\partial \boldsymbol{\phi}^{\mathsf{T}}} + \tau_0\mathbf{P}
\end{aligned} \tag{20}$$

with

$$\begin{aligned}
\frac{\partial \log h_{it}}{\partial \phi_k} &= \frac{\partial \log f_0(t)}{\phi_k} + (\mathrm{e}^{\eta_F(\tilde{\mathbf{v}}_i(t))} - 1)\frac{\partial \log S_0(t)}{\phi_k} \\
\frac{\partial \log f_0(t)}{\phi_k} &= \breve{b}_{tk} = b_{tk} - \sum_s \pi_s b_{sk} \\
\frac{\partial \log S_0(t)}{\phi_k} &= -\frac{1}{S_0(t)}\sum_{s\leq t}\pi_s \breve{b}_{sk}\ .
\end{aligned}$$

For the regression and spline parameters defining long-term survival, the conditional score and precision matrix are

$$\begin{aligned}
\mathbf{U}^{\psi}_{\lambda} &= \frac{\partial \ell_p}{\partial \boldsymbol{\psi}} = \sum_{i=1}^{n}\sum_{t=1}^{t_i}(d_{it} - \mu_{it})\frac{\partial \log h_{it}}{\partial \boldsymbol{\psi}} - \mathbf{K}_\lambda(\boldsymbol{\psi} - \mathbf{b}) \\
-\mathbf{H}^{\psi\psi}_{\lambda} &= \sum_{i=1}^{n}\sum_{t=1}^{t_i}\mu_{it}\frac{\partial \log h_{it}}{\partial \boldsymbol{\psi}}\frac{\partial \log h_{it}}{\partial \boldsymbol{\psi}^{\mathsf{T}}} + \mathbf{K}_\lambda
\end{aligned} \tag{21}$$

with $\partial \log h_{it} / \partial \psi_k = (\boldsymbol{\mathcal{X}}_t)_{ik}$. For the regression and spline parameters defining short-term survival, the conditional score and precision matrix are

$$\begin{aligned}
\mathbf{U}_{\tilde{\lambda}}^{\tilde{\psi}} &= \frac{\partial \ell_p}{\partial \tilde{\boldsymbol{\psi}}} = \sum_{i=1}^{n} \sum_{t=1}^{t_i} (d_{it} - \mu_{it}) \frac{\partial \log h_{it}}{\partial \tilde{\boldsymbol{\psi}}} - \tilde{\mathbf{K}}_{\tilde{\lambda}} (\tilde{\boldsymbol{\psi}} - \mathbf{g}) \\
-\mathbf{H}_{\tilde{\lambda}}^{\tilde{\psi}\tilde{\psi}} &= \sum_{i=1}^{n} \sum_{t=1}^{t_i} \mu_{it} \frac{\partial \log h_{it}}{\partial \tilde{\boldsymbol{\psi}}} \frac{\partial \log h_{it}}{\partial \tilde{\boldsymbol{\psi}}^{\intercal}} + \tilde{\mathbf{K}}_{\tilde{\lambda}}
\end{aligned} \tag{22}$$

with $\partial \log h_{it} / \partial \tilde{\psi}_k = (\tilde{\boldsymbol{\mathcal{X}}}_t)_{ik} \big(1 + \mathrm{e}^{\eta_F(\tilde{\mathbf{v}}(t))} \log S_0(t)\big)$. Closed forms for cross-derivatives (independent of the penalty parameters) can also be obtained:

$$-\mathbf{H}^{\psi\tilde{\psi}} = \sum_{i=1}^{n} \sum_{t=1}^{t_i} \mu_{it} \frac{\partial \log h_{it}}{\partial \boldsymbol{\psi}} \frac{\partial \log h_{it}}{\partial \tilde{\boldsymbol{\psi}}^{\intercal}} \,.$$

Let $\boldsymbol{\zeta} = (\boldsymbol{\psi}^{\intercal}, \tilde{\boldsymbol{\psi}}^{\intercal})^{\intercal}$ and $\boldsymbol{\tau} = (\lambda^{\intercal}, \tilde{\lambda}^{\intercal})^{\intercal}$. Then, the score and (minus) the precision matrix for $\boldsymbol{\zeta}$ are

$$\mathbf{U}_{\tau}^{\zeta} = \begin{pmatrix} \mathbf{U}_{\lambda}^{\psi} \\ \mathbf{U}_{\tilde{\lambda}}^{\tilde{\psi}} \end{pmatrix} \quad ; \quad \mathcal{H}_{\tau} = \begin{bmatrix} \mathbf{H}_{\lambda}^{\psi\psi} & \mathbf{H}^{\psi\tilde{\psi}} \\ \mathbf{H}^{\tilde{\psi}\psi} & \mathbf{H}_{\tilde{\lambda}}^{\tilde{\psi}\tilde{\psi}} \end{bmatrix} \,. \tag{23}$$