Nonparametric double additive cure survival models: an application to the estimation of the nonlinear effect of age at first parenthood on fertility progression

Vincent Bremhorst [1], **Michaela Kreyenfeld** [2], and **Philippe Lambert** [3,1]

[1] Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Louvain-la-Neuve, Belgium

[2] Hertie School of Governance, Berlin, Germany

[3] Université de Liège, Faculté des sciences sociales, Méthodes quantitatives en sciences sociales, Liège, Belgium

---

Address for correspondence: Vincent Bremhorst, Université catholique de Louvain, Institut de Statistique, Biostatistique et Sciences Actuarielles, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

E-mail: vincent.bremhorst@uclouvain.be.

Phone: (+32) 010 47 43 14.

Fax: (+32) 010 47 30 32.

Abstract: This paper introduces double additive models to describe the effect of

continuous covariates in cure survival models, thereby relaxing the traditional linear assumption in the two regression parts. This class of models extends the classical event history models when an unknown proportion of the population under study will never have the event-of-interest. They are used on data from the German Socio-Economic Panel (GSOEP) to examine how age at first birth relates to the timing and quantum of fertility for given education levels of the respondents. It is shown that the conditional probability of having further children decreases with the mother's age at first birth. While the effect of age at first birth in the third birth's probability model is fairly linear, this is not the case for the second child with an accelerating decline detected for women that had their first kid beyond age 30.

Key words: Bayesian P-splines; Births, Cure survival models; Continuous covariates; Double additive models, Fertility studies

1 Introduction

The increase in the age at first parenthood is a key indicator of the second demographic transition (Lesthaeghe, 1995, 2010; Sobotka, 2008; Sobotka et al., 2011). Albeit that the onset and pace of this development varied greatly across countries and in spite of the fact that some European countries have reported that postponement is even about to come to a halt, late first-time childbearing is a pertinent and common feature of the fertility patterns in European countries (Frejka and Sardon, 2006; Goldstein et al., 2009).

On the macro level, the great challenge has been to adequately account for the tempo

effects that have distorted period fertility rates (Billari and Kohler, 2004; Sobotka, 2004). On the micro level, researchers have examined whether and how a late age at first birth affects parity progression, birth spacing and completed fertility (Marini and Hodsdon, 1981; Dommaraju, 2009; Bratti and Tatsiramos, 2012; Berrington et al., 2015). Both approaches have in common that they investigate whether and to which extent births can be 'recuperated' at later ages. In other words: Does the increase in the age at first birth simply shift the fertility schedule within the life course? Does it compress it into a shorter time period? Or does a late age at first childbearing eventually result in a decline in cohort fertility rates?

On the one hand, late age at first parenthood is assumed to result in lower completed fertility because of the rapid drop in fecundity with women's age (Billari et al., 2007). Albeit that the availability of assisted reproduction has fuelled a discussion on the biological limits of fertility, medical research usually points out that women's ability to conceive and bear children declines over time due to the 'depletion and ageing of the pool of oocytes stored in both ovaries during the fetal period' (Velde et al., 2012, p. 1179). The ability to conceive and bear children declines gradually, but this process is assumed to accelerate around age 35. For men, fecundity is not subject to the same mechanisms, but the ability to father a child is nevertheless assumed to decay over time due to the decline of sperm quality with age (Schmidt et al., 2012; Johnson et al., 2015).

On the other hand, a late onset of fertility may not necessarily lead to lower completed fertility, because couples anticipate their impaired possibilities to have children at later ages. Postponement may be irrelevant for completed fertility, because couples have the ability to influence the spacing of births and may, thus, 'squeeze' their children into a shorter time interval. This 'time squeeze' has been assumed to particularly

explain the fertility behaviour of work-oriented women who postpone childbearing to advance in their career and accelerate childbearing at later ages (Kreyenfeld, 2002; Gerster et al., 2007; Bartus et al., 2013). Furthermore, it was pointed out that the correlation between age at first birth and completed fertility was not a solid evidence of a causal relationship. Women and men who postpone first birth might do so because they had originally planned to have fewer children. Thus, couples self-select themselves into late childbearing, depending on their fertility preferences. Furthermore, genetic factors, such as health impairment, are important unobservable factors that affect the onset as well as completion of fertility (Kohler et al., 1999; Rodgers et al., 2008).

There is a significant body of empirical studies that has examined the association of age at first birth and subsequent fertility. However, the studies radically differ in terms of data and method employed. Schmidt et al. (2012) correlated the period TFR and the mean age at first birth for several European countries for 2007, but did not find a strong association between the two measures. Velde et al. (2012) examined the impact of fertility postponement on involuntary childlessness and total fertility in six European countries. Based on a simulation with macro-level data, they show that fertility would have been between 0.03 and 0.05 higher in 2008 if there had been no postponement since the 1980s. Andersson et al. (2009) rely on Scandinavian register data to show that early onset of childbearing leads to a higher number of children among women born between the 1930s and 1950s. The mean number of children is well above 2 children for women who were age 25 or younger when they had their first child. For women who postponed first birth beyond age 35, total fertility drops below 1.5 children per woman. The effect of age at first birth on total fertility is linear and strong. But it is also shown that patterns vary across birth cohorts and countries.

This paper adds to the discussion on the association between first birth and completed fertility by employing the promotion time model (Yakovlev and Tsodikov, 1996; Tsodikov, 1998; Chen et al., 1999) initially motivated to analyze cancer data. It belongs to the class of cure survival models extending classical event history models by acknowledging that an unknown proportion of the studied population will never have the event-of-interest. Bremhorst et al. (2016) motivated their use in fertility when studying the effect of the educational levels of a woman and of her partner on second and third parity progression. The motivation for the promotion in a fertility context works as follows. Assume that at the onset of the process (i.e. direclty after last birth), the woman has $N \sim \mathcal{P}(\theta)$ (Poisson distributed) possible decisive arguments to opt for an additional child. Let F be the proper c.d.f. of the (independent and identically distributed) latent times $Y_1, ..., Y_n$ necessary for any of these n arguments to initiate a new pregnancy. Then, one can show that the population survival and density functions of the event-time are

$$S_p(t|\theta, F) = \exp\left(-\theta F(t)\right) \; ; \; f_p(t|\theta, F) = \theta f(t) S_p(t|\theta), \tag{1.1}$$

where $f(t) = \frac{\partial F(t)}{\partial t}$ is the latent density.

In particular, the probability of never becoming pregnant again (or 'cured' in the jargon of the survival literature) is

$$P[N = 0] = \exp(-\theta) = \lim_{t \to +\infty} S_p(t|\theta, F). \tag{1.2}$$

The expressions of the survival and hazard functions of the susceptible population can be found, for example in Bremhorst and Lambert (2016, Section 2).

Independent baseline covariates, denoted by $\boldsymbol{x}$ (including a constant '1' to multiply the intercept) and $\boldsymbol{z}$ (without such a constant), may enter, for example, the model through a log-link on parameter θ and through a Cox model for $F(t)$, respectively,

yielding

$$\theta(x) = \exp(\eta_\theta(x)) \qquad \text{with} \qquad \eta_\theta(x) = \alpha^T x \tag{1.3}$$

$$F(t|z) = 1 - S_0(t)^{\exp(\eta_F(z))} \qquad \text{with} \qquad \eta_F(z) = \beta^T z. \tag{1.4}$$

Tsodikov (2002) proposed, in a frequentist framework, a nonparametric estimation of the baseline survival function $S_0(t)$. Within the Bayesian paradigm, Yin and Ibrahim (2005) assumed a piecewize exponential distribution, while Bremhorst and Lambert (2016) opted for a flexible specification of $S_0(t)$ using P-splines (Eilers and Marx, 1996, 2010). For recent papers using or extending the promotion time model, we refer the interested reader to Liu and Shen (2009); Kim et al. (2009); Lopes and Bolfarine (2012) and Li and Lee (2017).

In this work, instead of assuming a linear effect of the continuous covariates in the regression parts $\eta_\theta(\cdot)$ and $\eta_F(\cdot)$ in (1.3) and (1.4), double additive models based on Bayesian P-splines (Lang and Brezger, 2004) will be specified. Flexible modelling of continuous covariates were already considered in many different contexts, see Hastie and Tibshirani (1990) for an early reference and, to cite a few, Wood (2006, 2011) for an efficient software implementation and Lambert (2013) for their double use in a semi-parametric Bayesian framework.

The paper is organized as follows : Section 2 and Section 3 define the flexible additive models using Bayesian P-splines for the continuous covariates and for the logarithm of the baseline hazard function, respectively. Bayesian inference techniques, including the specification of the prior distributions and the description of the Metropolis-within-Gibbs algorithm to sample from the joint posterior are described in Section 4. A simulation study assessing the accuray of the new methodology is presented in Section 5 while Section 6 focusses on the analysis of the GSOEP datasets. A

discussion concludes the paper.

2 Flexible regression model for covariate effects

Let $\boldsymbol{x^b} = (x_1^b, \ldots, x_P^b)$ and $\boldsymbol{z^b} = (z_1^b, \ldots, z_Q^b)$ be the binary covariate vectors influencing, respectively, the probability of having the event and its timing for susceptible subjects. Denote by $\boldsymbol{x^c} = (x_1^c, \ldots, x_R^c)$ and $\boldsymbol{z^c} = (z_1^c, \ldots, z_S^c)$ the sets of continuous covariates assumed to take values in $(-1, 1)$ and having an impact on the probability of being cured and on the timing of the event for susceptible individuals, respectively. The assumption made on the support of the continuous covariates is not restrictive since any continuous covariate can be transformed to meet that requirement.

The regression models defined in (1.3) and (1.4) assume that the continuous covariates contribute linearly to the functions $\eta_\theta(\cdot)$ and $\eta_F(\cdot)$ involved in the conditional the probability of being cured and in the conditional distribution of the event time for susceptible subjects. These linearity assumptions can be relaxed by specifying double additive models for the effects of continuous covariates:

$$\eta_\theta(x^b, x^c) = \alpha_0 + \boldsymbol{\alpha}^T \boldsymbol{x^b} + \sum_{r=1}^{R} g_r^\theta(x_r^c), \tag{2.1}$$

$$\eta_F(z^b, z^c) = \boldsymbol{\beta}^T \boldsymbol{z^b} + \sum_{s=1}^{S} g_s^F(z_s^c) \tag{2.2}$$

The functions $g_r^\theta(.)$ and $g_s^F(.)$ are specified using a linear combination of a large number (say, L) of cubic B-splines:

$$g_r^\theta(x_{c_r}) = \sum_{l=1}^{L} \phi_{rl}^\theta \tilde{b}_l(x_{c_r}) \quad ; \quad g_s^F(z_{c_s}) = \sum_{l=1}^{L} \phi_{sl}^F \tilde{b}_l(z_{c_s}), \tag{2.3}$$

where $\{\tilde{b}_1(.), \ldots, \tilde{b}_l(.)\}$ denotes a cubic B-splines basis associated to a predefined number of equidistant knots on $[-1, 1]$.

To avoid identification issues with the intercept α_0 (cf. Eq. 2.1) and with the baseline distribution of the Cox model (defined in Eq. 2.2), each spline coefficients vector $(\boldsymbol{\phi_1^\theta}, \ldots, \boldsymbol{\phi_R^\theta})$ and $(\boldsymbol{\phi_1^F}, \ldots, \boldsymbol{\phi_S^F})$ is constrained to sum at 0.

To ensure smoothness and a linear behaviour in the limiting case, the likelihood is combined with a second order roughness penalty on finite differences of adjacent B-spline parameters (Eilers and Marx, 1996, 2010). For example, a second order penalty associated to the r^{th} continuous covariate having an impact on the probability of the event is written as $\tau_r^\theta \sum_l \left(\Delta^2 \phi_{rl}^\theta\right)^2 = \tau_r^\theta (\boldsymbol{\phi_r^\theta})^T (\boldsymbol{D}_2)^T \boldsymbol{D}_2 \boldsymbol{\phi_r^\theta}$, where τ_r^θ is the penalty parameter and $\boldsymbol{D_2}$ is the difference matrix defined as follows:

$$\boldsymbol{D_2} = \begin{bmatrix} 1 & -2 & 1 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 & -2 & 1 \end{bmatrix}.$$

3 Flexible specification of the baseline distribution

As suggested by Bremhorst and Lambert (2016) in the context of the promotion time model, the baseline distribution $S_0(t)$ (of the Cox model defined in (2.2)) can be specified in a flexible way through the logarithm of the baseline hazard function $\log\left(h_0(t)\right)$:

$$\log\left(h_0(t)\right) = \sum_{k=1}^{K} \phi_k b_k(t), \tag{3.1}$$

where $\{b_1(.), \ldots, b_K(.)\}$ denotes a cubic B-splines basis associated to a predefined number of equidistant knots on $[0, T_{\max}]$, where $T_{\max}$ is the upper bound of the follow-

up interval.

Remember that the latent cumulative distribution function $F(t)$ in (1.1) is assumed to be proper. When a non(semi)-parametric model is proposed for the latent distribution, that assumption can be forced using the zero tail constraint (Taylor, 1995; Zeng et al., 2006). Accordingly, Bremhorst and Lambert (2016) suggest to fix the last spline coefficient to a large enough value such that $S_0(t)$ smoothly decreases to 0 when approaching the maximum follow-up, thereby translating that a susceptible subject must have experienced the event of interest by that time. However, when the population survival function reaches its minimum value long before $T_{\max}$ (as can be revealed from the wide plateau in the right tail of the Kaplan Meier estimate of the population survival, when the maximum censoring time is much larger than the maximum observed failure time), the estimation of the spline parameters can be numerically instable since no information is available to estimate the spline parameters supported by knots located in that part of the follow-up interval. Therefore, in this paper, we suggest to set all the spline parameters associated to knots located over the maximum observed failure time (i.e. in the plateau of the Kaplan Meier estimate) to an arbitrary large value (such as 10).

As for the flexible additive models defined in (2.3), a roughness penalty on finite difference of adjacent B-spline coefficients is used to force smoothness. However, for the baseline distribution, a third order penalty is used to force a limiting quadratic behaviour of the logarithm of the baseline hazard for large penalty parameter values.

4 Bayesian inference

4.1 Likelihood

Let $\mathbb{D}_i = (t_i, \delta_i, \boldsymbol{x_i^b}, \boldsymbol{x_i^c}, \boldsymbol{z_i^b}, \boldsymbol{z_i^c})$ be the set of the observable data of the i^{th} individual under study, where t_i and δ_i denote the observed survival time and the event indicator (i.e. $\delta_i = 0$ if he or she is right censored). Since the data are right censored, the survival log-likelihood is given by

$$l(\boldsymbol{\Phi}|\mathbb{D}) = \sum_{i=1}^{I} \delta_i \log\left(h_p\left(t_i|\boldsymbol{\Phi}, \mathbb{D}_i\right)\right) + \log\left(S_p\left(t_i|\boldsymbol{\Phi}, \mathbb{D}_i\right)\right), \tag{4.1}$$

where $\boldsymbol{\Phi}$ is the set of the model specific parameters.

4.2 Bayesian model

The frequentist roughness penalty introduced by Eilers and Marx (1996, 2010) is translated in a Bayesian framework by Lang and Brezger (2004) into a multivariate normal prior distribution for the spline parameters:

$$\boldsymbol{\phi}|\tau \sim N_K\left(0, (\tau \boldsymbol{P_3})^{-1}\right); \tag{4.2}$$

$$\boldsymbol{\phi_r^\theta}|\tau_r^\theta \sim N_L\left(0, \left(\tau_r^\theta \boldsymbol{P_2}\right)^{-1}\right) \qquad \forall r = 1, \ldots, R; \tag{4.3}$$

$$\boldsymbol{\phi_s^F}|\tau_s^F \sim N_L\left(0, \left(\tau_s^F \boldsymbol{P_2}\right)^{-1}\right) \qquad \forall s = 1, \ldots, S, \tag{4.4}$$

where $\boldsymbol{P_d} = \boldsymbol{D_d}^T \boldsymbol{D_d} + \epsilon \boldsymbol{I}$ is a full rank matrix for some small quantity ϵ (10^{-6}, say). Lang and Brezger (2004) suggests a Gamma distribution $\mathbb{G}(a, b)$ with mean $\frac{a}{b}$ as a prior for the penalty parameters. However, it can be shown that the shape of the estimated curve could be significanlty influence by the choice of the values of the hyperparameters a and b. Therefore, in this work and as in Bremhorst and Lambert

(2016) and in Bremhorst et al. (2016), we suggest to use the robust roughness penalty prior distribution proposed by Jullion and Lambert (2007):

$$\tau|\lambda \sim \mathbb{G}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right) \quad ; \quad \lambda \sim \mathbb{G}\left(a, b\right); \tag{4.5}$$

$$\tau_r^\theta|\lambda_r^\theta \sim \mathbb{G}\left(\frac{\nu}{2}, \frac{\nu\lambda_r^\theta}{2}\right) \quad ; \quad \lambda_r^\theta \sim \mathbb{G}\left(a, b\right) \qquad \forall r = 1, \ldots, R; \tag{4.6}$$

$$\tau_s^F|\lambda_s^F \sim \mathbb{G}\left(\frac{\nu}{2}, \frac{\nu\lambda_s^F}{2}\right) \quad ; \quad \lambda_s^F \sim \mathbb{G}\left(a, b\right) \qquad \forall s = 1, \ldots, S. \tag{4.7}$$

Jullion and Lambert (2007) showed that if the value of the hyperparameters a and b are small enough (say, $a = b = 10^{-4}$), the posterior distribution of ν is close to a uniform and, therefore, the value of ν would not have any impact on the shape of the estimate.

Independent normal distributions with a large variance σ^2 are used as priors for the regression parameters:

$$(\alpha_0, \boldsymbol{\alpha}, \boldsymbol{\beta}) \sim N_{1+p+q}\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right) \tag{4.8}$$

4.3 Posterior sampling using MCMC

The logarithm of the joint posterior distribution is obtained, using Bayes' theorem, as the sum of (4.1), the log-likelihood, and the logarithm of the prior density functions defined in (4.2)-(4.8).

It can be shown that only the conditional posterior distribution of the penalty pa-

rameters belongs to a known family of distributions:

$$\begin{aligned}
\tau|\boldsymbol{\phi}, \lambda, \mathbb{D} &\sim \mathbb{G}\left(\frac{\nu + K}{2}, \frac{\nu\lambda + \boldsymbol{\phi}^T \boldsymbol{P_3} \boldsymbol{\phi}}{2}\right); \\
\lambda, \mathbb{D} &\sim \mathbb{G}\left(a + \frac{\nu}{2}, b + \frac{\nu\tau}{2}\right); \\
\tau_r^\theta|\boldsymbol{\phi_r^\theta}, \lambda_r^\theta, \mathbb{D} &\sim \mathbb{G}\left(\frac{\nu + L}{2}, \frac{\nu\lambda_r^\theta + \left(\boldsymbol{\phi_r^\theta}\right)^T \boldsymbol{P_2} \boldsymbol{\phi_r^\theta}}{2}\right) \quad \forall \quad r = 1, \ldots, R; \\
\lambda_r^\theta|\tau_r^\theta, \mathbb{D} &\sim \mathbb{G}\left(a + \frac{\nu}{2}, b + \frac{\nu\tau_r^\theta}{2}\right) \quad \forall \quad r = 1, \ldots, R; \\
\tau_s^F|\boldsymbol{\phi_r^F}, \lambda_s^F, \mathbb{D} &\sim \mathbb{G}\left(\frac{\nu + L}{2}, \frac{\nu\lambda_s^F + \left(\boldsymbol{\phi_s^F}\right)^T \boldsymbol{P_2} \boldsymbol{\phi_s^F}}{2}\right) \quad \forall \quad s = 1, \ldots, S; \\
\lambda_s^F|\tau_s^F, \mathbb{D} &\sim \mathbb{G}\left(a + \frac{\nu}{2}, b + \frac{\nu\tau_s^F}{2}\right) \quad \forall \quad s = 1, \ldots, S.
\end{aligned}$$

Therefore, a Metropolis-within-Gibbs algorithm will be used to sample from the joint posterior. When subsets of the model parameters are correlated, as with the P-spline models defined in (2.3) and (3.1), reparametrizing the joint posterior distribution using an adequate estimation of the posterior correlation structure of the model parameters might improve the mixing of the posterior chains (Lambert, 2007). Such an estimation can be achieved using a nonlinear optimizer, based e.g. on the Augmented Lagrangian method (see, for example Nocedal and Wright (2006, Chapter 17)), enabling to deal with linear constraints on the parameters.

To speed up the convergence of MCMC algorithm, the posterior mode of the joint posterior distribution, for fixed values of $\lambda, \lambda_1^\theta, \ldots, \lambda_R^\theta, \lambda_1^F, \ldots, \lambda_S^F$, is used as initial values of the MCMC algorithm. For the Metropolis steps and as recommended by Haario et al. (2001) and Atchadé and Rosenthal (2005), the standard deviation of the proposal distribution is updated during the burnin period to achieve the target acceptation rate (23% in a multivariate setting).

5 Simulation

Table 1: Simulation results for $S = 500$ replicates: coverage probabilities (in %) of the 95% credible intervals for each regression parameter.

		$n = 500$					$n = 1\,000$				
		Cure probability			Timing		Cure probability			Timing	
Cure	Censored	α_0	α_1	α_2	β_1	β_2	α_0	α_1	α_2	β_1	β_2
23%	28%	98.4	94.0	94.6	93.2	93.2	99.2	95.2	94.8	94.6	94.2
23%	50%	98.2	92.0	91.8	92.6	90.8	97.4	93.8	94.8	92.2	94.8
48%	53%	98.2	93.0	94.0	93.2	92.2	99.4	95.2	96.2	93.8	94.6
48%	69%	98.2	93.4	92.6	92.0	90.8	98.8	94.2	93.8	91.8	92.8

The numerical performances of the proposed methodology are illustrated through a simulation study. For all the settings described hereafter, $S = 500$ replicates of sample size $n = 500$ and $n = 1\,000$ were generated. The simulated datasets are analysed using the model presented in Sections 2 and 3 combined with the Bayesian inference techniques described in Section 4. The flexible additive models (cf. Eq. 2.3) are defined using a basis of $L = 10$ B-splines associated to equidistant knots on $(-1, 1)$ (for the relocated and rescaled continuous covariates) while a basis with $K = 15$ B-splines was taken on the follow-up interval to model (the log hazard of) the baseline distribution (cf. Eq. 3.1). The results presented in Tables 1 and 2 and in Figures 1-6 are based on one MCMC chain of length 75 000 including a burnin period of length 25 000. The z-scores of the Geweke convergence diagnostics (not reported) were used to assess the convergence of the Markov chains (Geweke, 1992).

Two percentages of cured individuals were considered: 23% and 48%. The percent-

ages of right censored subjects among the susceptible sub-population are controlled by two censoring distributions: a Weibull distribution with mean 17.86 and standard deviation 6.49 and an Exponential distribution with mean 18.18. Both are truncated at $T_{\max} = 25$, the upper bound of the follow-up, leading to 5% and more than 20% of right censored susceptible subjects, respectively. Since the baseline distribution is specified, in each setting, as a Weibull distribution with mean 8 and standard deviation 4.18, the sufficiently long follow-up assumption is respected. Indeed, less than 0.1% of the event occur after time $T_{\max} = 25$ under this distribution.

Let $W_1, W_2 \sim Bin(1, 0.5)$ and $W_3, W_4 \sim U(-1, 1)$ be the four independent covariates having simultaneously an impact on the probability of the event and on its timing for the susceptible sub-population. For each setting, one defines $\boldsymbol{x^b} = (W_1, W_2) = \boldsymbol{z^b}$, $(\beta_1, \beta_2) = (0.4, -0.4)$, $\boldsymbol{x^c} = (W_3, W_4) = \boldsymbol{z^c}$, $f_1^\theta(.) = g_2^F(.)$ and $f_2^\theta(.) = g_1^F(.)$, where $f_1^\theta(x) = \frac{x}{2}$ and $f_2^\theta(x) = \frac{\sin(2\pi x)}{1.5}$. The value of $\boldsymbol{\alpha}$ is tuned in each setting to reach the target percentage of cure subjects.

Figure 1 shows a negligible bias of the posterior medians (as estimators) for all regression parameters in each setting. The variability of the estimates increases with the percentage of right censored subjects among the susceptible sub-population. Not surprisingly, an increase of the proportion of right censored subjects is more influential on the precision of parameter estimates than a comparable increase in the percentage of cured units. The coverage probabilities of the 95% credible intervals for all the regression parameters (see Table 1) are close to their nominal value except for the intercept α_0 in the regression model for the cure probability where the coverage probabilities are to high.

Figures 2 and 5 illustrate the estimations of a linear effect of a continuous covariate on the probability of having the event and on its timing for the susceptible sub-

population. In each setting and in both parts of the model, the linear trend is captured without bias by the flexible additive models. As expected, the accuracy of the estimates increases with the sample size and decreases when the percentage of right censoring increases.

The estimation of a sine-like effect of a continuous covariate on the probability of being cured is pictured in Figure 3. In each considered scenario, the additive model was able to recover the nonlinear shape. However, when the percentage of right censoring is important, a small bias (becoming negligible when the sample size increases) appears in the area where the concavity of the curve changes, as illustrated on the fourth row of Figure 3). This issue is more pronounced when estimating the nonlinear effect on the timing of the event for the susceptible subjects, as shown in Figure 4. This can be explained by the decreasing information on the regression parameters in the Cox model for susceptible subjects when the percentage of right censored units increases. Regarding accuracy, the same conclusions as for the estimation of the linear effects hold.

Table 2 reports the coverage probabilities of the 95% simultaneous credible intervalsfor the additive terms in the two regression models and estimated using MCMC with the technique described in Held (2004). The estimated coverages are all close to their nominal values except for the nonlinear term $g_1^F(\cdot)$ involved in the regression model for the timing of the event for susceptible individuals, when information is very sparse (typically when the sample size and the percentage of events are simultaneously small), leading to an oversmoothed estimate. Table 2 also reveals that for replicates yielding a credible region that does not contain the true additive term, the average proportion of the (-1,1) covariate support leading to a coverage of the true value is quite satisfactory.

Finally, the same conclusions as for the other functions estimates hold for the estimation of the baseline distribution, pictured in Figure 6.

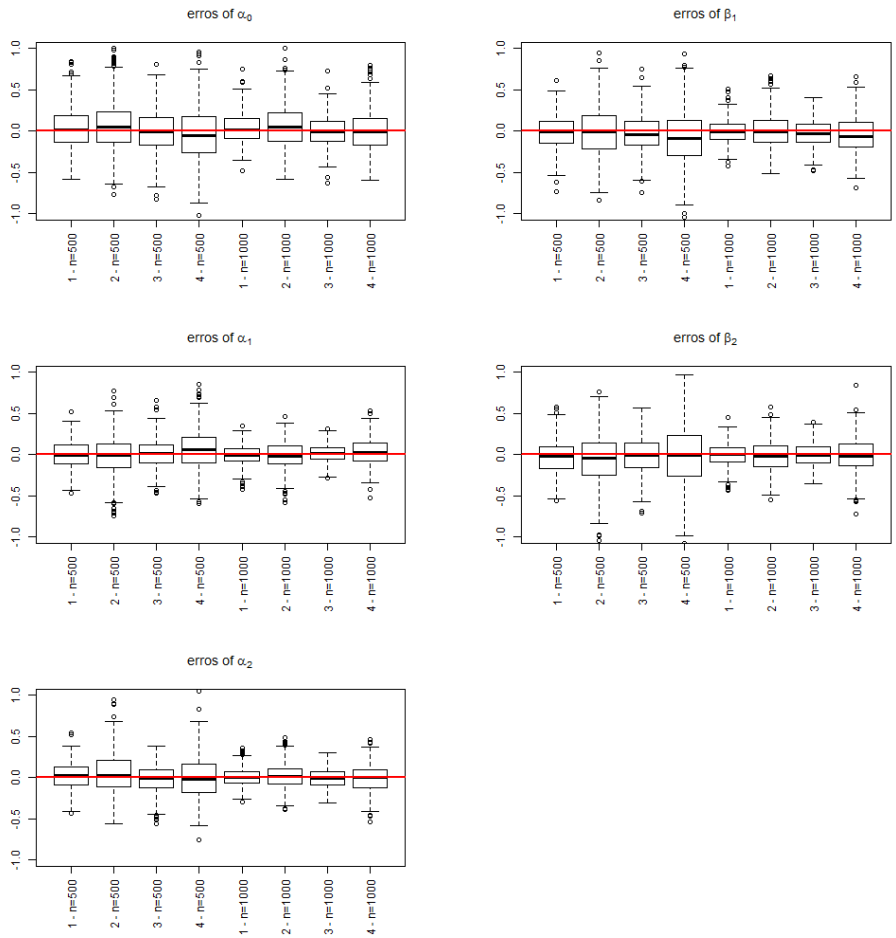

Figure 1: Simulation results for $S = 500$ replicates: boxplots of the errors of the regression parameter estimates (the posterior medians) for each considered scenario. The scenarios are numbered as follow : 1 = 23% of cured and 28% of right censoring; 2 = 23% of cured and 50% of right censoring; 3 = 48% of cured and 53% of right censoring; 4 = 48% of cured and 69% of right censoring

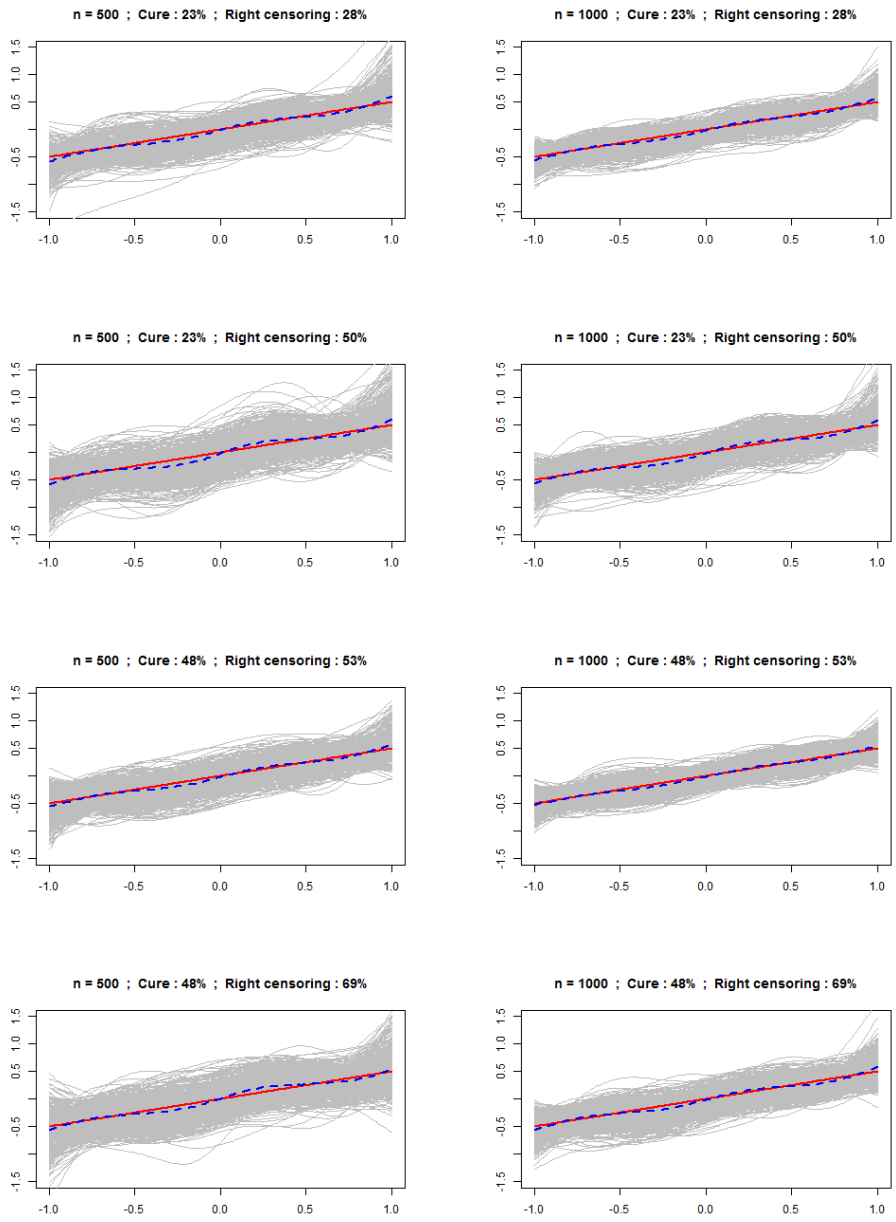

Figure 2: Simulation results: estimation of $g_1^\theta(x_1) = \frac{x_1}{2}$. $S = 500$ replicates (one gray curve per data set) with sample size $n = 500$ (left) or $n = 1\,000$ (right). Each row refers to a percentage of cured and right censored individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves.

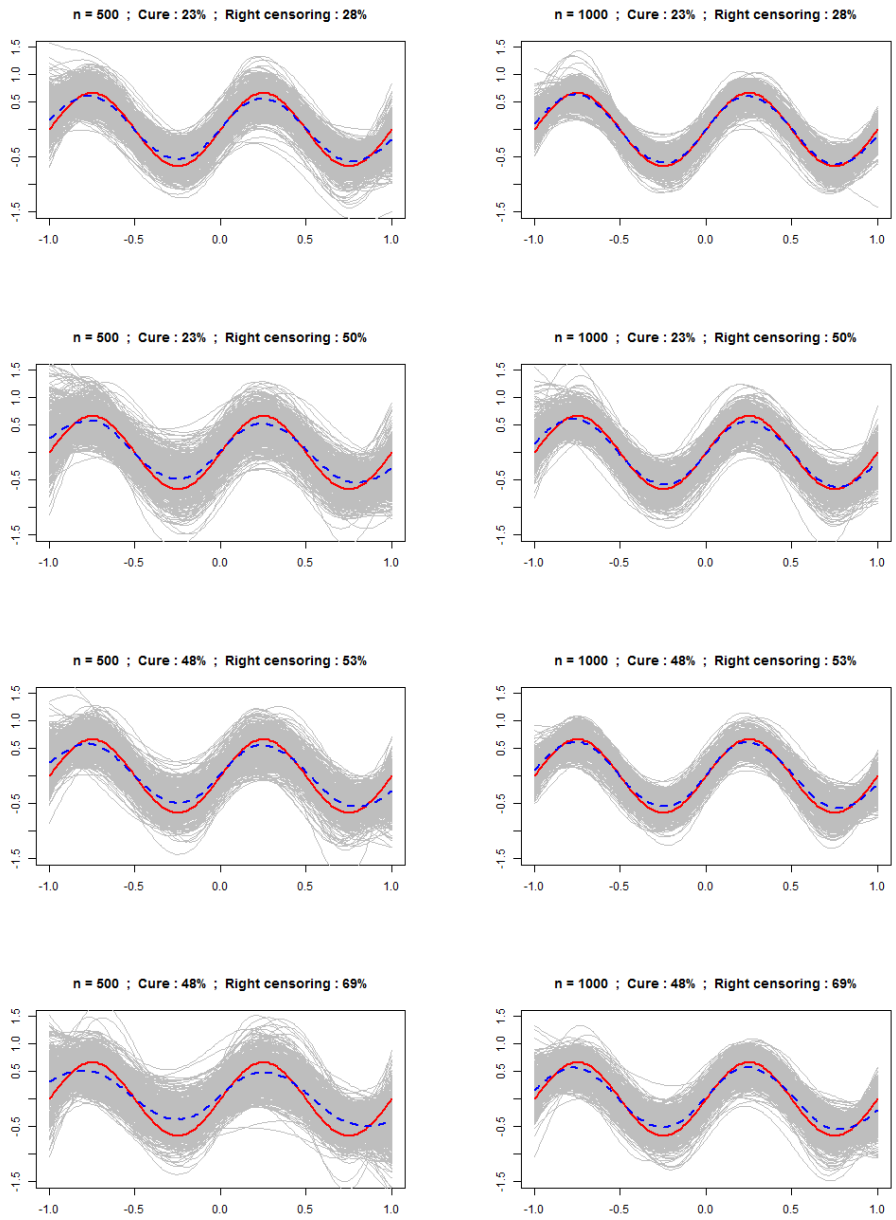

Figure 3: Simulation results: estimation of $g_2^{\theta}(x_2) = \frac{\sin(2\pi x_2)}{1.5}$. $S = 500$ replicates (one gray curve per data set) with sample size $n = 500$ (left) or $n = 1\,000$ (right). Each row refers to a percentage of cured and right censored individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves.

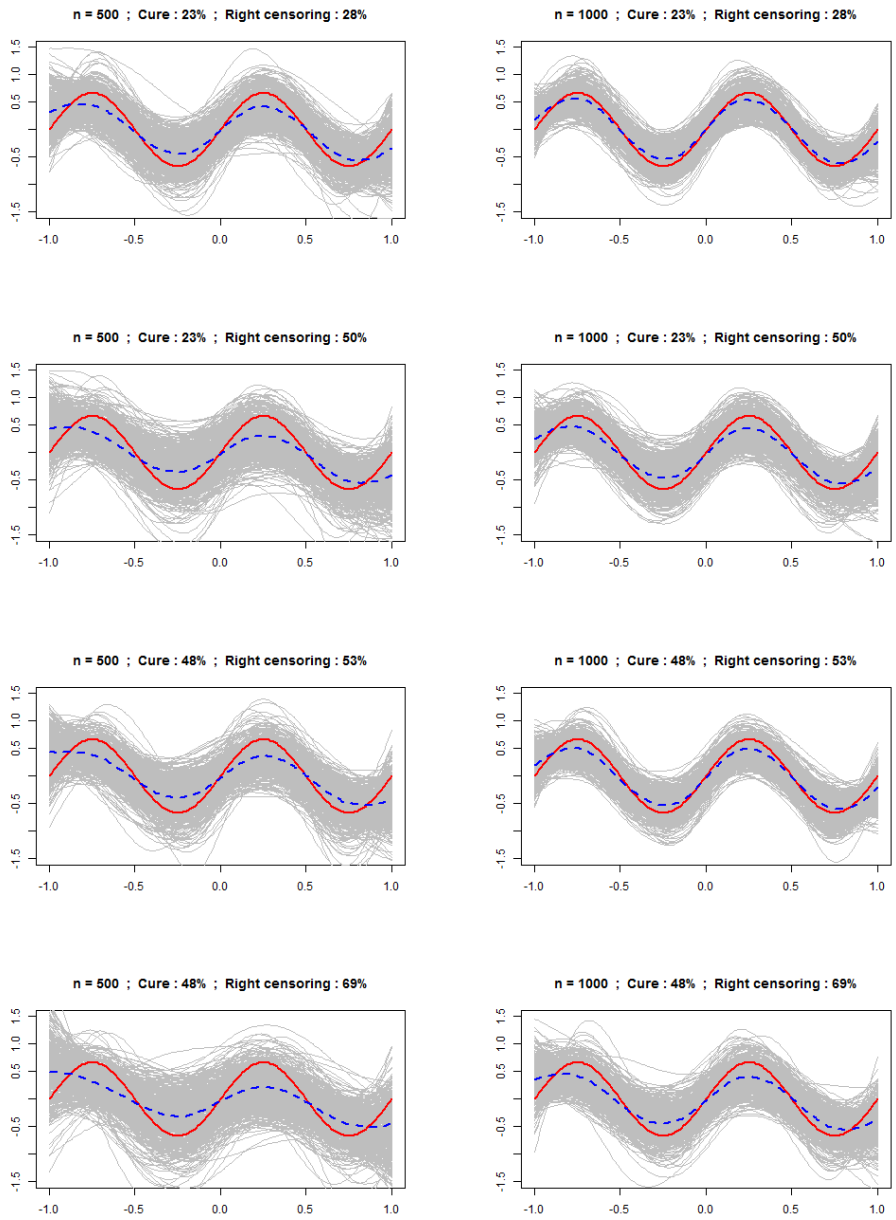

Figure 4: Simulation results: estimation of $g_1^F(z_1) = \frac{\sin(2\pi z_1)}{1.5}$. $S = 500$ replicates (one gray curve per data set) with sample size $n = 500$ (left) or $n = 1\,000$ (right). Each row refers to a percentage of cured and right censored individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves.

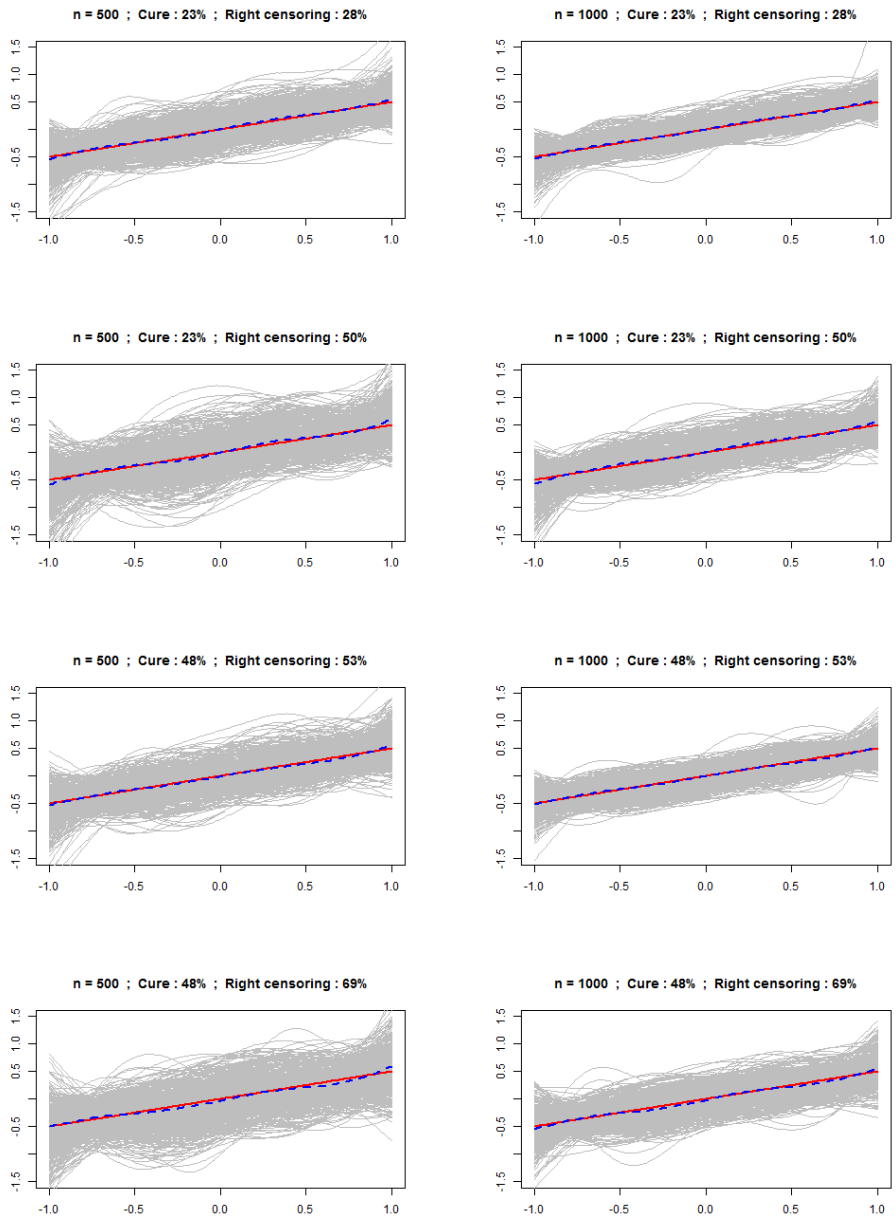

Figure 5: Simulation results: estimation of $g_2^F(z_2) = 0.5z_2$. $S = 500$ replicates (one gray curve per data set) with sample size $n = 500$ (left) or $n = 1\,000$ (right). Each row refers to a percentage of cured and right censored individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves.

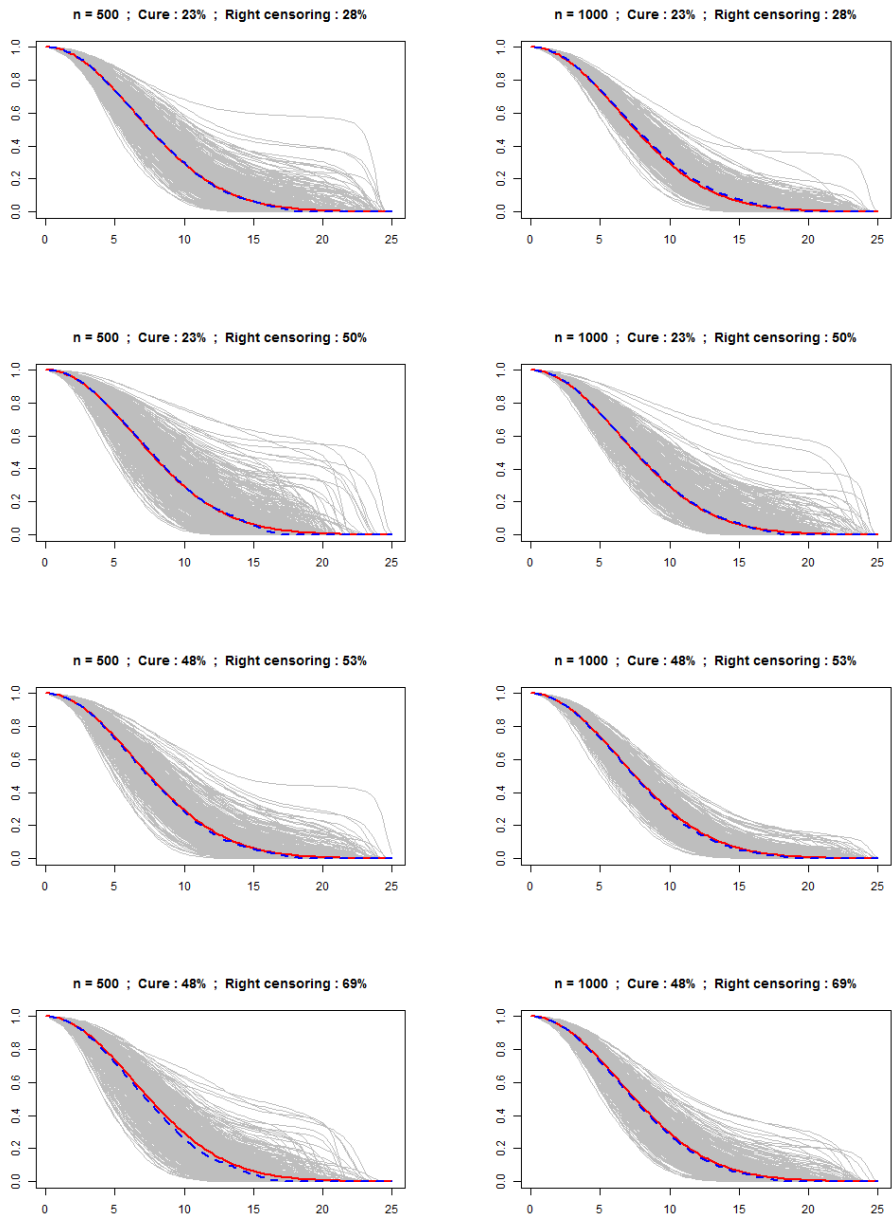

Figure 6: Simulation results: estimation of the baseline distribution $S_0(t)$. $S = 500$ replicates (one gray curve per data set) with sample size $n = 500$ (left) or $n = 1\,000$ (right). Each row refers to a percentage of cured and right censored individuals. The solid line corresponds to the true function and the dashed line is the pointwise median of the 500 estimated curves.

Table 2: Simulation results for $S = 500$ replicates: coverage probabilities (in %) of the 95% simultaneous credible regions. In brackets: replicates for which the credible region does not contain the true additive term: average proportion of the (-1,1) covariate support for which the true additive term belongs to the simulatenous credible region. The true functions are defined as $g_1^{\theta}(x_1^c) = 0.5x_1^c$; $g_1^F(z_1^c) = \sin(2\pi z_1^c)$; $g_2^{\theta}(x_2^c) = \sin(2\pi x_2^c))$; $g_2^F(z_2^c) = 0.5z_2^c$.

			Cure probability		Timing	
n	Cure	Censored	$g_1^{\theta}(x_1^c)$	$g_2^{\theta}(x_2^c)$	$g_1^F(z_1^c)$	$g_2^F(z_2^c)$
500	23%	28%	96.6 (92.54)	96.0 (95.22)	91.4 (93.22)	96.0 (91.47)
	23%	50%	96.8 (91.11)	93.6 (93.25)	87.4 (92.30)	95.6 (88.60)
	48%	53%	97.8 (94.21)	93.8 (93.52)	84.6 (93.69)	96.6 (89.11)
	48%	69%	97.4 (93.57)	91.0 (92.77)	78.6 (91.23)	96.2 (90.97)
1 000	23%	28%	98.4 (89.43)	97.8 (94.12)	96.8 (96.61)	97.6 (93.20)
	23%	50%	96.6 (95.08)	96.8 (94.50)	93.4 (95.46)	96.6 (91.28)
	48%	53%	98.8 (94.36)	99.2 (97.89)	97.4 (94.99)	96.8 (94.12)
	48%	69%	98.2 (92.32)	95.6 (96.07)	90.4 (92.64)	96.8 (91.85)

6 Application

Table 3: Descriptive statistics of the independent covariates for all mothers at risk of an additional birth.

Continuous variables (in years)		Second birth	Third birth
Age at 1^{st} birth (mother)	Mean	27.9	26.4
	Median	28.0	26.0
	std	4.8	4.7
Age at previous birth (partner)	Mean	30.6	32.7
	Median	30.0	32.0
	std	5.5	5.1
Birth interval $1^{st}/2^{nd}$ child	Mean	-	3.6
	Median	-	3.0
	std	-	2.2
Education levels		Frequency	
Mother	Low	340	393
	Medium	936	959
	High	231	238
Partner	Low	311	309
	Medium	901	945
	High	295	336

Data for this analysis comes from the German Socio-Economic Panel (GSOEP). The GSOEP is the longest standing household panel survey in Europe (Wagner et al., 2007). Interviews are conducted with all household members aged 17 and older (and

since recently also with younger respondents) and are repeated every year. The original sample included West German households and an oversample of foreigners. Since its initiation in 1984, the SOEP has been extended several times. One of the most significant extensions was the inclusion of an Eastern German subsample in the year 1990 when German reunification was ratified. For our analysis, we use data from wave 2014 (Release 31.1). We restrict the analysis to the period $1984 - 2013$. Our focus is the analysis of second and third birth transitions. Thus, we omit childless women from the investigation. Furthermore, we restrict the analysis to women aged $17 - 49$ who became under risk of having a second or third child after they entered the panel study. By doing so, we disregard left truncated observations. We have also decided to censor the cases 15 years after previous birth. Although our analysis mainly focusses on the mother, we also considered some characteristics of the co-residential partner. Omitted are all respondents who did not have a partner at last birth. We also omitted all cases with missing information on education or of any of the partners. Finally, we restricted the sample to women who had their preceding child in Western Germany. This restriction is attributed to the large East-West differences in fertility behaviour that prevailed after reunification. Also migrant women were omitted, as migrant fertility would require a separate investigation. Finally, we dropped cases from the so-called high income sample so that the analyses are not distorted due to the oversampling of high-income groups.

The dependent variables are the progression to the second and third child with time measured since the preceding birth. Transition to second and third births are studied separately. For both studies, the data are analysed using the extended promotion time model defined in (1.1) with (1.3) and (1.4) as expressions for the link functions $\eta_\theta(.)$ and $\eta_F(.)$ combined with the Bayesian inference methodology presented in Section 4.

The convergence of the resulting posterior chains (of length 150 000 with a burnin period of 50 000) were assessed by the trace plots and using convergence diagnostics tools such as those proposed by Geweke (1992). Our key independent variable is the age of the mother at first birth. We furthermore control for partner's age at the previous birth. For third birth, we also consider the duration between the first and second births. We control for the education levels of the mother and of her partner described as being either low (less than a vocational degree), medium (a vocational training degree) or high (a university or college degree). Education is handled as a time-constant covariate and was fixed at its value at the occasion of the last birth (see Table 3 for the descriptive statistics). The final sample comprises 1 507 one-child mothers and 1 590 two-child mothers. Among the one-child mothers, 57% had a second child by the end of the follow-up. Among the two-child mothers, 22.8% had a third child.

6.1 Second birth

In a first step, we investigate second birth behaviour. Our interest is mainly the effect of mother's age at first birth. The results from this covariate are presented on the first row of Figure 7. The left panel shows the estimated additive term $g^{\theta}(age)$ in (1.3) quantifying the effect of mothers's age at first birth on the probability of having a second child for given education levels and age of the partner at first birth. There is a clear negative nonlinear effect: the older a woman at first birth, the lower the probability of having a second child with an accelerating decline when the first birth occured after 30. This finding is compatible with medical studies that indicate a decline of fecundity over time and a rapid deterioration in the mid 30s. If the

interpretation was correct, one would assume that this effect is less pronounced for males. The results for partner's age at previous (first) birth, pictured on the bottom left graph of Figure 7, might partially support this claim. However, although second birth probabilities still decline, in particular if the partner was 40 or older at first birth, the sample size is too small to assert it with a 90% confidence level.

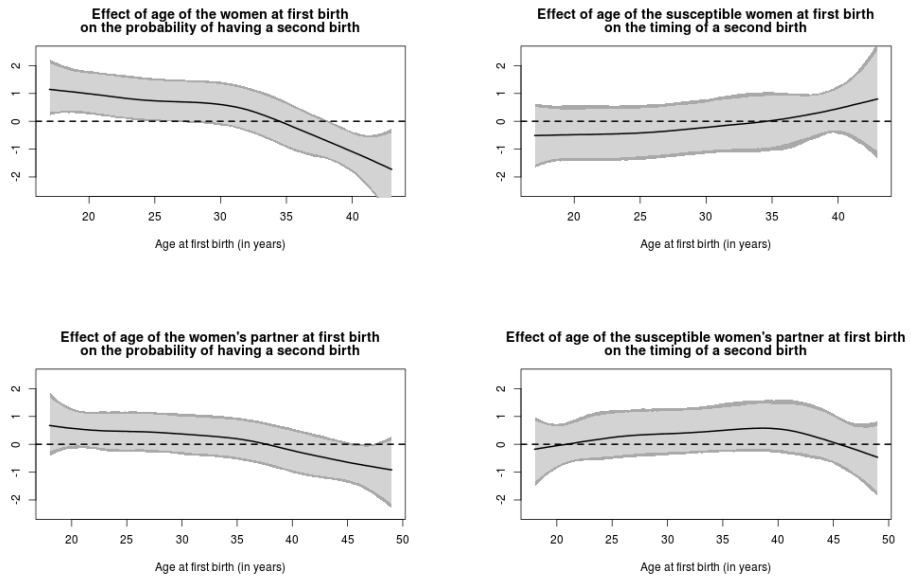

Figure 7: Estimation of the conditional effect the mother's age at first birth (row 1) and of the age at first birth of her partner (row 2) on the probability of having a second child (left) and on the timing of a second birth for the susceptible women (right). Dark (resp. light) grey region coincides with the 95% (resp. 90%) simultaneous credible region.

The results from the other covariates, namely the education levels of the two partners, are displayed in Table 4. The results corroborate earlier investigations with the German Socio-Economic Panel that focused on the effect of education on fertility in cure fraction models (Bremhorst et al., 2016). For given ages of the two partners,

highly educated women display a significantly higher probability of having a second child compared to medium or less educated women. In addition, having a highly educated partner significantly increases that probability compared to other couples. On the other hand, education does not have a significant effect on fertility timing (for given ages of the partners at first birth).

Table 4: Transition to second birth - Posterior median and 95% (HPD) credible interval of for the regression parameter associated to categorical covariates.

	Probability		Timing	
	Est	$HPD_{95\%}$	Est	$HPD_{95\%}$
Intercept	-0.696	[-1.478 ; -0.010]	-	-
Mother's education (ref. Middle)				
Low	-0.147	[-0.378 ; 0.069]	0.061	[-0.231 ; 0.354]
High	0.367	[0.047 ; 0.644]	-0.171	[-0.561 ; 0.198]
Partner's education (ref. Middle)				
Low	0.108	[-0.126 ; 0.335]	-0.233	[-0.524 ; 0.066]
High	0.406	[0.146 ; 0.653]	0.020	[-0.310 ; 0.336]

6.2 Third birth

In a second step, we examined the effect of mother's age at first birth on third birth behaviour (for a given time interval between the first two births, a given partner's age at second birth and given education levels of the partners). The results of the impact of women's age at first birth on third birth parity progression is illustrated on the first row of Figure 8. The top left graph shows a very strong (linear) negative effect

of age at first birth on the probability of having a third child. The top right graph surmises that the conditional effect of age at first birth on the timing of third birth for susceptible two-child women is positive, suggesting that women squeeze their further children into shorter birth intervals. In particular women who had their first child after age 30 tend to accelerate third birth transition. However, a larger number of births will be needed to claim it with at a 95% confience level. The bottom left graph of Figure 8 displays the conditional effect of the time elapsed between the first two births on parity progression. It suggests that the longer the time interval between the first two children, the smaller the probability of having a third child. Since no significant conditional effect was found for the partner's age at second birth on the probability or on the timing of having a third child, the estimated effects are not reported.

Table 5 reports the effect of women's and partner's education levels. It corroborates the findings in Bremhorst et al. (2016): While the mother's education level does not significantly affect the third birth probability, a U-shaped effect for the partner's education is suggested with medium educated partner significantly less likely to progress to a third child.

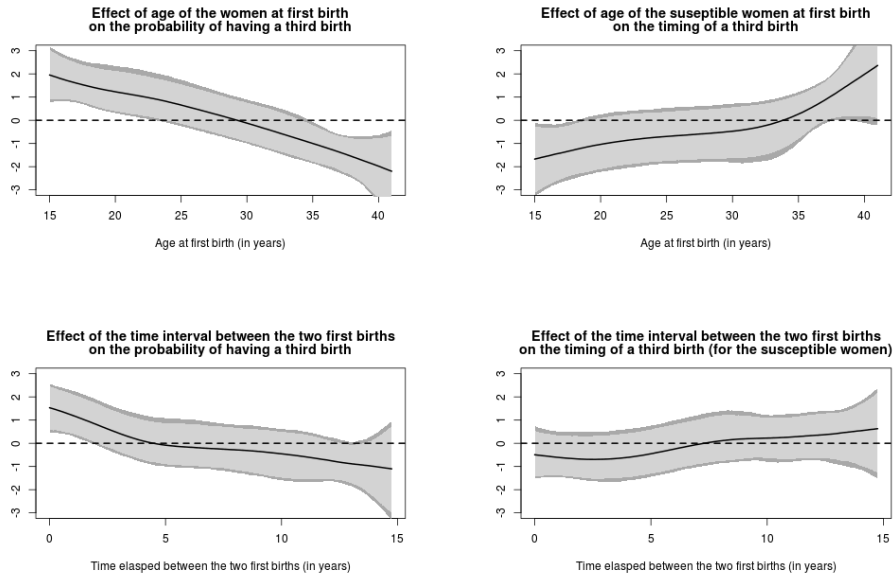

Figure 8: Estimation of the conditional effect of mother's age at first birth (row 1) and of the time elapsed between the first and the second births (row 2) on the probability of having a third child (left) and on the timing of a third birth for the susceptible women (right). Dark (resp. light) grey region coincides with the 95% (resp. 90%) simultaneous credible region.

Table 5: Transition to third birth - Posterior median and 95% (HPD) credible interval of for the regression parameter associated to categorical covariates.

	Probability		Timing	
	Est	$HPD_{95\%}$	Est	$HPD_{95\%}$
Intercept	-2.197	[-3.292 ; -1.173]	-	-
Mother's education (ref. Middle)				
Low	0.218	[-0.060 ; 0.501]	-0.279	[-0.647 ; 0.079]
High	0.290	[-0.131 ; 0.711]	-0.082	[-0.596 ; 0.435]
Partner's education (ref. Middle)				
Low	0.435	[0.152 ; 0.728]	-0.371	[-0.754 ; -0.006]
High	0.704	[0.363 ; 1.038]	0.056	[-0.373 ; 0.475]

7 Discussion

This paper has used cure survival models to estimate the effect of age at first parenthood on fertility progression in Germany. To reach that goal, we considered double additive models to specify in a flexible way the effect of continuous covariates in cure fraction models. With this tool we tried to detect non-linearities in the effect of mother's age at first birth on the progression to a second and third child. Our analysis shows that age at first birth has a diverging effect on timing and quantum. While a high age at first birth seems to accelerate parity progression (specially for third birth) for susceptible women, it reduces the overall chances of having further children. This important finding underscores the relevance of cure fraction models in fertility research. Standard event history models easily generate misleading results

because they conflate the effect of timing and quantum (see Bremhorst et al., 2016, Section 4.3 for more details). Cure survival models overcome this shortcoming, but have unfortunately not diffused much in social science research.

An important limitation of our study was the limited sample size. We had about 1 500 women at risk of second and another 1 500 at risk of third birth at our disposal. However, we often were unable to generate significant results. Due to the small sample size, we were neither able to study effect heterogeneities. In particular, we were unable to analyse the interaction effects of education and age at first birth. This is unfortunate because earlier studies have pointed out that the 'meaning' of age at birth depends on level of education (Hoem, 1996; Kreyenfeld, 2002; Bartus et al., 2013). While highly educated women regularly postpone childbirth in order to advance in their career, lowly educated women who postpone parenthood are a select group, possibly with impaired abilities to have children. Thus, one would assume that age at first birth has a stronger effect on the probability of having a second or third child for lowly than for highly educated women.

From a theoretical point of view, a possible extension would be to develop a statistical tool to test the linearity of the effects of continuous covariates. To reach that goal, non-linear perturbations could be added to the linear specification of the continuous covariates in (1.3) and (1.4). These nonlinear perturbations would be described using P-splines with a first order penalty and a zero mean to force a flat limiting behaviour of the correction for large values of the corresponding penalty parameter. An inspection of the simultaneous credible region for the correction and of the posterior of the associated penalty parameter could then be used to assess whether a linear approximation of the target additive term makes sense.

Acknowledgements

The authors acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract Projet d'Actions de Recherche Concertées (ARC) 11/16-039 of the Communauté française de Belgique, granted by the Académie universitaire Louvain. The authors also thanks the data provider: Socio-Economic Panel (SOEP) 1984-2014, international scientific use version 31.1, doi: 10.5684/soep.v31.1i.

References

Andersson, G., Knudsen, L. B., Neyer, G., Teschner, K., Ronsen, M., Lappegard, T., Skrede, K., and Vikat, A. (2009). Cohort fertility patterns in the nordic countries. *Demographic research*, **20**(14), 313–352.

Atchadé, Y. and Rosenthal, J. (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli*, **11**(5), 815–828.

Bartus, T., Murinkó, L., Szalma, I., and Szél, B. (2013). The effect of education on second births in hungary: A test of the time-squeeze, self-selection, and partner-effect hypotheses. *Demographic research*, **28**(1), 1–32.

Berrington, A., Stone, J., and Beaujouan, E. (2015). Educational differences in timing and quantum of childbearing in Britain: A study of cohorts born 1940-1969. *Demograaphic Research*, **33**(26), 733–764.

Billari, F. C. and Kohler, H.-P. (2004). Patterns of low and lowest-low fertility in Europe. *Population Studies*, **8**(2), 161–176.

Billari, F. C., Kohler, H.-P., Andersson, G., and Lundstrom, H. (2007). Approaching the limit: Long-term trends in late and very late fertility. *Population and Development Review*, **33**(1), 149–170.

Bratti, M. and Tatsiramos, K. (2012). The effect of delaying motherhood on the second childbirth in Europe. *Journal of Population Economics,*, **25**, 291–321.

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

Bremhorst, V., Kreyenfeld, M., and Lambert, P. (2016). Fertility progression in Germany : An analysis using flexible non parametric cure survival models. *Demographic research*, **35**(18), 505–534.

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

Dommaraju, P. (2009). Timing of first birth in India. *Genus*, **65**, 81–101.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical science*, **11**, 89–121.

Eilers, P. H. C. and Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**(6), 637–653.

Frejka, T. and Sardon, J.-P. (2006). First birth trends in developed countries: Persisting parenthood postponement. *Demographic Research*, **15**(6), 147–180.

Gerster, M., Keiding, N., Knudsen, L. B., and Strandberg-Larsen, K. (2007). Education and second birth rates in Denmark 1981-1994. *Demographic research*, **17**(8), 181–210.

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. Clarendon Press, Oxford, In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian statistics 4 edition.

Goldstein, J. R., Sobotka, T., and Jasilioniene, A. (2009). The end of 'lowest-low' fertility? *Population and Development Review*, **35**, 663–699.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive Metropolis algorithm. *Bernoulli*, **7**(2), 223–242.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. Chapman & Hall.

Held, L. (2004). Simultaneous posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics*, **13**, 20–35.

Hoem, B. (1996). The social meaning of the age at second birth for third-birth fertility: A methodological note on the need to sometimes respecify an intermediate cariable. *Yearbook of Population Research in Finland*, **33**, 333–339.

Johnson, S. L., Dunleavy, J., Gemmell, N. J., and Nakagawa, S. (2015). Consistent age-dependent declines in human semen quality: A systematic review and meta-analysis. *Ageing Research Reviews*, **19**, 22–33.

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

Kim, S., Xi, Y., and Chen, M. (2009). A new latent cure rate marker model for survival data. *annals of applied statistics*, **3**, 1124–1146.

Kohler, H.-P., Rodgers, J. L., and Christensen, K. (1999). Is fertility behavior in our genes? Findings from a Danish twin study. *Population and Development Review*, **25**(2), 253–288.

Kreyenfeld, M. (2002). Time squeeze, partner effect or self-selection? An investigation into the positive effect of women's education on second birth risks in West Germany. *Demographic research*, **7**(2), 15–48.

Lambert, P. (2007). Archimedean copula estimation using Bayesian splines smoothing techniques. *Computational Statistics and Data Analysis*, **51**, 6307–6320.

Lambert, P. (2013). Nonparametric additive location-scale models for interval censored data. *Statistics and Computing*, **23**, 75–90.

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

Lesthaeghe, R. (1995). *The Second Demographic Transition in Western Countries: An interpretation.* Clarendon Press, Oxford, In K. Oppenheim Mason & A.-M. Jensen, Gender and Family Change in Industrialized Countries edition.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and Development Review*, **36**(2), 211–251.

Li, L. and Lee, J. (2017). A latent promotion time cure rate model using dependent tail-free mixtures. *Journal of the royal statistical society. Serie A*. doi: http://dx.doi.org/10.1111/rssa.12226.

Liu, H. and Shen, Y. (2009). A semi-parametric accelerate failure time cure model. *Journal of the American Statistical Association*, **104**, 1168–1178.

Lopes, C. and Bolfarine, H. (2012). Random effects in promotion time cure rate models. *Computational Statistics and Data Analysis*, **56**, 75–87.

Marini, M. M. and Hodsdon, P. J. (1981). Effects of the timing of marriage and first birth of the spacing of subsequent births. *Demography*, **18**, 529–548.

Nocedal, J. and Wright, S. (2006). *Numerical optimization*. Springer Series in Operations Research and Financial Engineering.

Rodgers, J., Kohler, H., McGue, M., Behrman, J., Petersen, I., Bingley, P., and Christensen, K. (2008). Education and cognitive ability as direct, mediating, or spurious influences on female age at first birth: Behavior genetic models fit to Danish twin data. *American Journal of Sociology*, **114**, S202–S232.

Schmidt, L., Sobotka, T., Bentzen, J. G., and Nyboe Andersen, A. (2012). Demographic and medical consequences of the postponement of parenthood.? *Human Reproduction Update*, **28**, 29–43.

Sobotka, T. (2004). Is lowest-low fertility in Europe explained by the postponement of childbearing? *Population and Development Review*, **30**, 195–220.

Sobotka, T. (2008). Overview Chapter 6: The diverse faces of the second demographic transition in Europe? *Demographic Research*, **19**(8), 171–224.

Sobotka, T., Zeman, K., Lesthaeghe, R., and Frejka, T. (2011). Postponement and recuperation in cohort fertility: New analytical and projection methods and their application. *European Demographic Research Papers*, pages 1–86.

Taylor, J. (1995). Semi-parametric estimation in failure time mixture models. *Biometrics*, **51**, 899–907.

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

Tsodikov, A. (2002). Semi-parametric model of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. *Statistic in Medicine*, **21**, 895–920.

Velde, E. t., Habbema, D., Leridon, H., and Eijkemans, M. (2012). The effect of postponement of first motherhood on permanent involuntary childlessness and total fertility rate in six european countries since the 1970s. *Human Reproduction*.

Wagner, G. G., R., F. J., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP): Scope, evolution and enhancements. *Schmollers Jahrbuch*, **127**, 139–169.

Wood, S. (2006). *Generalized additive models: An introduction with R*. CRC Press.

Wood, S. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society:Series B (Statistical Methodology)*, **73**(1), 3–36.

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

Yin, G. and Ibrahim, J. (2005). Cure rate models : a unified approach. *Canadian Journal of Statistics*, **33**, 559–570.

Zeng, D., Yin, G., and Ibrahim, J. (2006). Semiparametric transformation models for survival data with a cure fraction. *Journal of the American Statistical Association*, **101**, 670–684.