DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

*DEMOGRAPHIC RESEARCH*

**VOLUME 47, ARTICLE 22, PAGES 647–694**
**PUBLISHED 28 OCTOBER 2022**
https://www.demographic-research.org/Volumes/Vol47/22/
DOI: 10.4054/DemRes.2022.47.22

*Research Article*

# Investigating the application of generalized additive models to discrete-time event history analysis for birth events

**Joanne Ellison**

**Ann Berrington**

**Erengul Dodd**

**Jonathan J. Forster**

# Contents

*Demographic Research*: Volume 47, Article 22
*Research Article*

# Investigating the application of generalized additive models to discrete-time event history analysis for birth events

**Joanne Ellison**[1]

**Ann Berrington**[2]

**Erengul Dodd**[3]

**Jonathan J. Forster**[4]

## Abstract

**BACKGROUND**
Discrete-time event history analysis (EHA) is the standard approach taken when modelling fertility histories collected in surveys, where the date of birth is often recorded imprecisely. This method is commonly used to investigate the factors associated with the time to a first or subsequent conception or birth. Although there is an emerging trend towards the smooth incorporation of continuous covariates in the broader literature, this is yet to be formally embraced in the context of birth events.

**OBJECTIVE**
We investigate the formal application of smooth methods implemented via generalized additive models (GAMs) to the analysis of fertility histories. We also determine whether and where GAMs offer a practical improvement over existing approaches.

**METHODS**
We fit parity-specific logistic GAMs to data from the UK Household Longitudinal Study, learning about the effects of age, period, time since last birth, educational qualification, and country of birth. First, we select the most parsimonious GAMs that fit the data sufficiently well. Then we compare them with corresponding models that use the existing methods of categorical, polynomial, and piecewise linear spline representations in terms of fit, complexity, and substantive insights gained.

[1] Department of Social Statistics and Demography, University of Southampton, UK.
Email: J.V.Ellison@soton.ac.uk.
[2] Department of Social Statistics and Demography, University of Southampton, UK.
[3] School of Mathematical Sciences, University of Southampton, UK.
[4] Department of Statistics, University of Warwick, Coventry, UK.

**RESULTS**
We find that smooth terms can offer considerable improvements in precision and efficiency, particularly for highly non-linear effects and interactions between continuous variables. Their flexibility enables the detection of important features that are missed or estimated imprecisely by comparator methods.

**CONTRIBUTION**
Our findings suggest that GAMs are a useful addition to the demographer's toolkit. They are highly relevant for motivating future methodological developments in EHA, both for birth events and more generally.

## 1. Introduction

Surveys provide a rich source of fertility data through collecting retrospective birth histories from individuals alongside personal, family, and neighbourhood characteristics. As the time of birth tends to be recorded imprecisely to the nearest month or year, methods that treat time as discrete are typically required (Steele 2005). To this end, discrete-time event history analysis (EHA) is the standard approach taken in the literature, commonly using a logistic setup (Steele 2005; Van Hook and Altman 2013). Modelling the log-hazard is also common (Fiori, Graham, and Feng 2014; Kravdal and Rindfuss 2008; Kulu and Washbrook 2014).

Research published in this area typically focuses on investigating the factors associated with entry into parenthood and progression to subsequent births. Covariates examined include:

- education (Begall and Mills 2013; Kravdal 2001; Piotrowski and Tong 2016; Zang 2019)
- parental fertility behaviour (Kim 2014; Morosow and Trappe 2018; Riise, Dommermuth, and Lyngstad 2016)
- region or residential context (Fiori, Graham, and Feng 2014; Hank 2002; Kulu 2013; White et al. 2008)
- receipt of welfare benefits (Erlandsson 2017)
- marital dissolution (Ivanova, Kalmijn, and Uunk 2014).

Methodological developments in the EHA of births include computing summary fertility measures (Retherford et al. 2010; Van Hook and Altman 2013), accounting for unobserved differences between women (Kravdal 2001), taking a Bayesian approach (Lewis and Raftery 1999; McDonald and Rosina 2001), and implementing sophisticated, data-driven methods to select parametric representations of covariates (Muggeo, Attanasio, and Porcu 2009; Raftery et al. 1996).

This paper presents a methodological examination of the different ways in which continuous time-varying and time-invariant covariates[5] can be incorporated into fertility analyses, focusing on the relevant “clock” variables (Raftery et al. 1996), i.e., age, period, cohort, and time since last birth. Following a review of the existing approaches in Section 2, in Section 3 we identify an important avenue for development, namely the use of smooth methods. Such methods allow more precise estimation of effects by making less restrictive assumptions and more effective use of the data. However, despite their increasing usage in the broader literature, they are yet to be formally adopted in the context of birth events.

To this end, the aims of this paper are twofold. First, we investigate the formal application of smooth methods to the analysis of birth event histories. Second, we quantify and visualise the impact of representing continuous covariates using smooth methods versus existing methods in terms of fit, complexity, and substantive insights gained, and assess how this varies with parity.

## 2. Review of literature regarding ways of representing continuous covariates

### 2.1 Categorical representations

The simplest approach is to discretise the continuous variable to create a categorical variable. This is termed a ‘piecewise constant’ representation, as the effect is the same within each interval but can differ between intervals. The method is very common in the literature. For example, 5-year categories are often used for age (Jefferies, Berrington, and Diamond 2000; Morosow and Trappe 2018). Justification for this includes insufficient data for stable single-year estimates and the frequent use of 5-year age groups in age-specific fertility rates (Van Hook and Altman 2013).

In terms of cohort, group widths of 5 years (Lewis and Raftery 1999; Muggeo, Attanasio, and Porcu 2009; Zang 2019), 10 years (Begall and Mills 2013; Kim 2017), and varying sizes (Piotrowski and Tong 2016; White et al. 2008) have been chosen. Similarly, 5-year groups (Kulu and Washbrook 2014; Raftery et al. 1996) and single-year categories (Erlandsson 2017; Torrisi 2020) have been used to represent period. Riise,

[5] We do not consider categorical variables because the only flexibility that they afford is the choice of the categories themselves.

Dommermuth, and Lyngstad (2016) specify 10-year groups relating to the period of the woman's first birth.

Various categorical representations of time since last birth (TSLB) can be found in the literature. Sometimes TSLB is truncated. For example, Retherford et al. (2010) and Erlandsson (2017) use single-year categories with maxima of 10 and 7 years respectively; Retherford et al. (2010) justify the omission of births occurring after 10 years by their rarity.[6] Often TSLB is top-coded. Ivanova, Kalmijn, and Uunk (2014), Jefferies, Berrington, and Diamond (2000), and Torrisi (2020) top-code TSLB at 7, 10, and 11 years respectively; the preceding values are grouped into several categories with widths of 1–4 years.

Torrisi (2020) states that the choice of categories (<2, 3–6, 7–10, 11+ years) was informed by the fast-changing nature of the underlying hazard early in the birth interval, compared to the slower declines exhibited later. The parameter estimates across these papers imply an increasing hazard peaking at 2–3 years after the previous birth, followed by a sustained decrease. This is consistent with Timæus and Moultrie (2008, 2020).

### 2.2 Polynomial representations

We henceforth consider approaches that treat the clocks as continuous, in order of increasing complexity. The assumption of a linear effect only requires one parameter, which is highly restrictive. Therefore, it is unsurprising that applications to age and TSLB, which are known to have complex, non-linear associations with fertility, are rare. Raftery et al. (1996) experiment with a linear term for cohort, while linear period effects are chosen by Kulu (2013), Muggeo, Attanasio, and Porcu (2009), and Lewis and Raftery (1999).

A quadratic representation allows the covariate effect to exhibit curvature. This is pertinent when estimating the effect of age, where the natural age-related decline in fecundability gives rise to a biological deadline, and the existence of social age deadlines and age norms limit childbearing even further (Billari et al. 2011; Van Bavel and Nitsche 2013). The ability to detect the resulting standard bell-shaped pattern of age (e.g., see Peristera and Kostaki (2007)) makes quadratic parameterisations appealing (e.g., see Kim 2014, Piotrowski and Tong 2016, White et al. 2008; and Zang 2019). Piotrowski and Tong (2016) and Zang (2019) additionally model TSLB quadratically, thereby capturing the peaking behaviour.

More complex polynomial specifications have also been proposed, such as the cubic age effect in Raftery et al. (1996): the presence of two turning points allows not only the

[6] A similar justification for a 10-year cut-off is used by Ní Bhrolcháin (1987) in the context of calculating period parity progression ratios (see footnote 23). Also see p.112 of Hinde (1998).

peak but also a change in the rate of decline at older ages to be detected. Despite continuous non-linear representations of cohort and period being uncommon in the literature, they could allow demographic phenomena such as baby booms and busts to be captured.

### 2.3 Spline representations

Although higher-degree polynomials offer greater flexibility than linear terms, their assumptions are still quite restrictive and they may lack parsimony. Piecewise linear splines approximate the effect with a series of line segments joined at knots, the number and locations of which are user-specified (Kulu 2013). In this way the required parameters are slopes for each interval, and an overall intercept (Kulu 2013); this may be less parsimonious than the corresponding categorical approach or polynomial specifications, but allows underlying patterns to be detected with greater precision.

In the literature, Begall and Mills (2013), Fiori, Graham, and Feng (2014), Kravdal and Rindfuss (2008), and Kulu and Washbrook (2014) use piecewise linear splines to estimate their age effects. In their models of progression to first birth, knots are at intervals of around 5 years, with those estimating age effects for higher parities choosing fewer knots; this reflects the prior knowledge that the age effect simplifies as parity increases. The models of first birth and higher-order births strongly support this assertion – a quadratic age term for entry into motherhood may be reasonable, but for higher parities the effect tends to change to a decline alone, making the choice questionable.

Papers utilising piecewise linear splines for the age effect also use them for TSLB: three or four knots are specified, typically at intervals of 2–3 years. Muggeo, Attanasio, and Porcu (2009) take the same approach in their model of progression to second birth, advancing the method by estimating the knot locations and selecting the number of knots within the modelling process. In all cases the estimates reflect the established form of the hazard (see Section 2.1).

### 2.4 Interactions

Interactions between a discretised variable and another categorical variable are often used (Erlandsson 2017; Morosow and Trappe 2018; Torrisi 2020). The more complex methods for representing age and TSLB also easily extend to interactions. For example, Kravdal (2001) and Zang (2019) allow their age effects to vary with educational attainment by respectively estimating piecewise linear spline and quadratic effects of age for each education category. The predicted hazard plots in the latter clearly demonstrate how this interaction captures the effect of postponement for more highly educated women. Rondinelli, Aassve, and Billari (2006) do something similar to Kravdal (2001) in their multi-spell random effect model, but with TSLB and wage level replacing age and education, and Retherford et al. (2010) modify their baseline hazard by additionally estimating quadratic functions of time for women with medium or high education levels, and those living in urban areas.

Muggeo, Attanasio, and Porcu (2009) again provide more advanced examples, including allowing the slopes and knots of their piecewise linear baseline hazard (see Section 2.3) to change linearly with cohort category. In conclusion, although sophisticated interactions are implemented in the literature, they tend to involve a single continuous variable with a non-linear representation. Interacting more than one would require advanced methodologies, but could allow intricate relationships such as changes in the effect of age over time to be captured precisely.

## 3. Towards a smooth framework for discrete-time EHA for birth events

Our review of the discrete-time EHA literature for birth events in Section 2 has identified varying levels of complexity in the incorporation of continuous clocks. Polynomial terms treat the variable as continuous but make restrictive global assumptions about the form of the effect; piecewise linear splines make the less restrictive local assumption of a constant rate of change within each interval, but do not exploit the inherent smoothness of the underlying effect. There is no borrowing of strength across the intervals – each slope is estimated independently, constrained only by the need to join at the knots. Furthermore, the number of parameters required is user-specified and cannot be easily informed by the data. In these ways the method does not make the most effective use of the information in the data.

To represent the effect as a smooth function, non-parametric methods are used. They do not assume a particular parametric form, only smoothness. Rather than being pre-specified, the complexity of the effect is data-driven and quantified by the effective degrees of freedom (EDF). This measure exists due to the enforced smoothness, meaning that parameters borrow strength from each other and are not estimated independently. Therefore, the EDF is smaller than the actual parameter count, and is a continuous measure. In summary, the flexibility of smooth methods allows underlying effects to be estimated more precisely, while their efficiency maintains parsimony.

Section 2 identifies examples in the context of birth events where such properties might be advantageous. These include curvature in the effects of age and TSLB, non-linearity in period and cohort effects, and the complexity of the age–time interaction.

Next, we summarise developments in the discrete-time EHA literature regarding non-parametric methods, which are yet to be embraced in the context of births.

Non-parametric methods have gradually become more common in the wider discrete-time EHA literature. Fahrmeir (1994), Fahrmeir and Knorr-Held (1997), and Fahrmeir and Wagenpfeil (1996) estimate smooth effects of duration and time-varying effects of other covariates. Studying the transition to motherhood, Galimberti (2002) achieves something similar by implementing and extending the work of Hastie and Tibshirani (1990, 1993). Fahrmeir and colleagues argue that parametric forms such as piecewise polynomials run the risk of missing "unexpected patterns" in these effects, and propose non-parametric methods as an exploratory technique to inform parametric specifications (Fahrmeir and Knorr-Held 1997); such sentiments are echoed by Wu (2003). A consequence of this shift in focus towards the use of smooth methods in EHA is the greater appeal of established data-driven approaches in the statistical literature (Wu 2003).

One such approach is that of generalized additive models (GAMs), proposed by Hastie and Tibshirani (1986, 1990); their work is referenced by Fahrmeir and Knorr-Held (1997) and Wu (2003). GAMs provide a statistical framework in which the main effect of each covariate, as well as interaction effects, can be estimated as non-parametric smooth functions. This formalises the challenges of determining both the construction of the functions and their level of smoothness (Wood 2017). Such issues have been addressed extensively in the statistical literature, with a plethora of extensions and efficient computational methodologies developed over the last 30 years (Wood 2020). GAMs have been applied in fields as diverse as ecology, health, air quality, geomorphology, statistical learning, finance, and economics.[7] The models of Fahrmeir and Galimberti can be viewed as GAMs with a smooth main effect of duration and smooth interactions of duration with various categorical variables.

More recently in discrete-time EHA, GAMs have been formally applied in contexts such as unemployment duration (Berger and Schmid 2018), organisation survival (Rousselière 2019), and seat reservation (Shao, Kauermann, and Smith 2020). Only smooth main effects are considered; however, Bender, Groll, and Scheipl (2018) allow for interactions between duration and another continuous variable, where each can contribute smoothly or linearly. The authors exploit the theory of generalized additive mixed models (GAMMs), which incorporate random effects, to provide an alternative to the Cox model. Therefore, the application of GAMs exists but is not commonplace in the

[7] Readers who are familiar with the economics literature may note similarities with the literature on production and cost functions, namely the comparison between the simple log-linear Cobb-Douglas functional form and the more complex transcendental logarithmic or 'translog' form, which additionally allows interactions. GAMs have been applied extensively as an alternative to these existing methods, with their flexible, non-linear specification often found to be advantageous (e.g., see Reyes Santías, Cadarso-Suárez, and Rodríguez-Álvarez 2011 and Ferrara and Vidoli 2017).

discrete-time EHA literature. The extent of their usage appears limited, with opportunities to estimate smooth interactions between any pair of continuous variables remaining unexplored.

This paper presents an application of GAMs to the discrete-time EHA of births. To the best of our knowledge GAMs have not been used for this purpose, and therefore our work represents a novel extension in this field which is consistent with the trend exhibited in the broader literature. We fit logistic GAMs for progression to first, second, third, and fourth or higher-order births using data from the UK Household Longitudinal Study (UKHLS), examining the smooth effects of age, period, and TSLB. To determine whether and where the increased precision and efficiency of GAMs offers a practical improvement over existing approaches, we then compare our chosen GAMs with corresponding models fitted using standard methods, both quantitatively and graphically. The remainder of the paper proceeds as follows: in Sections 4 and 5 we describe the dataset and our modelling approach respectively, while in Sections 6 and 7 we present the results and discuss our findings.

## 4. Data

The UKHLS is a large-scale longitudinal survey of 30,169 UK households (Knies 2018) which requests birth histories from respondents aged 16 and over (Institute for Social and Economic Research (ISER) 2020). The survey also collects information on family background and topics such as education and country of birth (ISER 2020), providing the opportunity to investigate the dependence of birth events on a range of covariates.

### 4.1 Analysis sample

We use data from Wave 1 (University of Essex 2017) collected in 2009–2011 (ISER 2020), as it removes the need to deal with the selective attrition present in subsequent waves (Lynn and Borkowska 2018). Our sample of 22,020 women (457,125 person-years) includes women who completed a valid fertility history, resided in England or Wales when interviewed, and lived at least one person-year in the UK during the observation period (January 1950 – November 2008). Appendix A gives further details about the sample selection process.

### 4.2 Birth events

We arrange our dataset into "person-period" records (Steele 2005). Although we could take our periods to be months, for computational efficiency we choose to work on the scale of years (of age) and so consider person-year records in our analysis. We take the reproductive age range to be 15–44, giving a maximum of 30 records per woman. The dependent variable in our analyses is a binary response: it takes the value 1 if the woman had a birth event during that year of age, and is 0 otherwise. The person-year record then also contains the values of the clocks and any additional covariates. We model progression to each birth order separately, which means that women can potentially contribute to more than one model. We discuss the implications of this modelling decision in Section 7.

### 4.3 Additional covariates

The focus of this paper is primarily methodological, in that it aims to demonstrate how GAMs can be used to perform discrete-time EHA of birth events, and to determine for which variables and parities the use of smooth terms outperforms existing methods. We are not attempting to provide a comprehensive analysis of the possible determinants of fertility in England and Wales. Therefore, in addition to the clocks, we consider just two important characteristics in our analysis: educational attainment and country of birth. Although these variables will both have categorical representations in our analyses, theory suggests that they are likely to interact with age and would therefore impact the estimation of smooth terms.

#### 4.3.1 Educational attainment

There is a large body of literature concerning the predictive nature of educational enrolment and attainment on fertility. Childbearing is less likely when enrolled in education for many reasons, including the practical challenge, financial implications, and social norms (Blossfeld and Huinink 1991; Lappegård and Rønsen 2005). For the most-educated women, this postponement is likely to continue while they focus on their career progression and professional goals (Liefbroer and Corijn 1999; Ní Bhrolcháin and Beaujouan 2012).

The distinction between enrolment and attainment is also reflected in the discrete-time EHA literature. Attainment is usually represented through a categorical variable which can be time-constant, typically the highest level at the time of interview (Muggeo, Attanasio, and Porcu 2009; Retherford et al. 2010; Torrisi 2020) or time-varying (Erlandsson 2017; Fiori, Graham, and Feng 2014; Zang 2019). In addition to this, current enrolment can be accounted for through a time-varying indicator variable (Kravdal 2001, 2007; Piotrowski and Tong 2016; White et al. 2008). The corresponding parameter estimates are almost always negative – often significantly so.

Our highest educational qualification variable has four categories: Less than General Certificate of Secondary Education (GCSE) ('< GCSE'); GCSE; Advanced (A) Level ('A Level'); Degree.[8] Following Berrington, Stone, and Beaujouan (2015), we reclassify degree-educated women as 'A Level' if their age at leaving full-time education suggests a break in continuous education; we also impute missing responses. The youngest women are likely to exceed their highest qualification at interview and so have censored observations. We develop an imputation model to assign them more appropriate values, the key assumptions being that the older (pre-1983) cohorts have uncensored observations, and the probability of being in a given category depends on cohort and country of birth. Extrapolating the model for the younger cohorts, we adjust the number of women originally observed in each category to match the implied counts as closely as possible. Further details can be found in Ellison (2021).

### 4.3.2 Country of birth

Various hypotheses have been proposed addressing the degree to which immigrants adapt their fertility behaviour to that of their host nation (Kulu and González-Ferrer 2014). Substantial support for such hypotheses (Kulu 2005), combined with the large proportion of live births in England and Wales to non-UK-born women (Office for National Statistics (ONS) 2020a) justifies the inclusion of country of birth.

We group the responses using the Human Development Index (HDI), which quantifies the development level of a country by combining health, education, and standard of living indicators (United Nations Development Programme (UNDP) 2018). For examples of uses of the HDI in the fertility literature, see Myrskylä, Kohler, and Billari (2009) and Robards and Berrington (2016). We use the 2018 values (UNDP 2019) to assign each woman an HDI, assigning appropriate averages to ambiguous or missing responses. We then use the country groupings (low, medium, high, and very high human development) to categorise the HDI values, extracting the UK-born women into their own category.

[8] Note that GCSEs are taken at the end of compulsory schooling (age 16); A Levels are taken at age 18 and are typically a requirement for university admission.

# 5. Modelling

## 5.1 GAM specification

We model progression to first, second, third, and fourth or higher-order births and so consider current parities $p \in \{0,1,2,3+\}$. We take a period approach in the main text, and outline the findings from a cohort approach in Section 7. Here, for a parity $p$ we specify the most complex model that includes all of the covariates under consideration, noting that in practice we can include any subset of these:

- Clocks: $A$ = age, $Y$ = year (period), $T$ = TSLB top-coded at 11 years[9] (available for $p > 0$)
- Additional covariates: $Q$ = highest educational qualification, $H$ = HDI

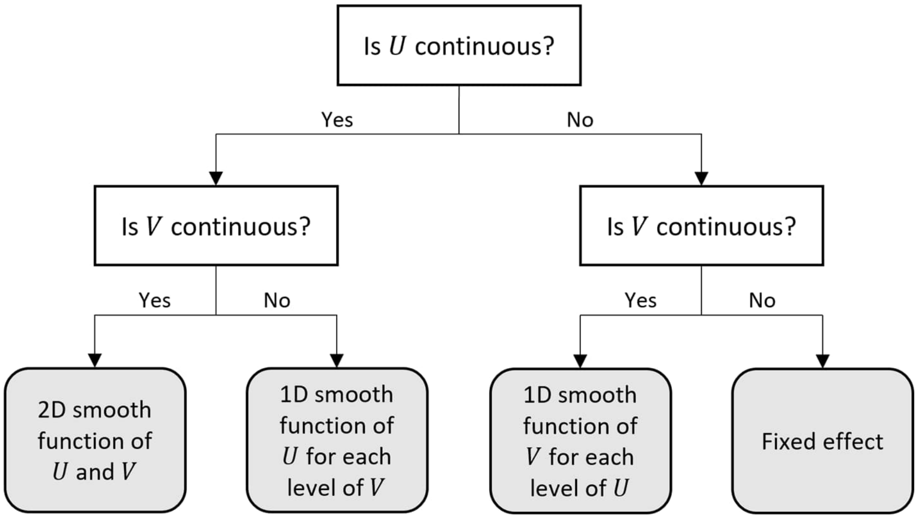
Let there be $N$ distinct covariate patterns (combinations of the covariate values) observed in the sample of person-years. For the group of records sharing the $i$th covariate pattern, let $n_i$ be the group size and $Z_i$ the sum of the $n_i$ binary responses, i.e., the number of birth events. We assume a binomial distribution for $Z_i$ with probability $r_i$, i.e., $Z_i \sim \text{Binomial}(n_i, r_i)$, noting that $r_i$ is the conditional probability of a birth event given the included covariates and $p$.

Our parity-specific binomial logistic GAM then sets the logit of $r_i$ equal to the sum of a series of terms:

- Fixed effects, denoted by $\mathbf{X}_i\boldsymbol{\beta}$ where $\mathbf{X}_i$ is the $i$th row of the fixed effects model matrix with corresponding parameters $\boldsymbol{\beta}$.
- One-dimensional (1D) smooth functions (i.e., curves), denoted by $f_U(u_i)$ for a generic continuous variable $U$ with $i$th observed value $u_i$.
- Smooth functions of the interaction between two variables, denoted by $f_{UV}(u_i, v_i)$ for generic variables $U$ and $V$. Figure 1 summarises the different forms that this can take, depending upon whether $U$ and $V$ are categorical or continuous. We note that if $U$ and $V$ are continuous, it is a two-dimensional (2D) smooth function (i.e., a surface); if $V$ is categorical (here, $Q$ or $H$), it is a 1D smooth function of $U$ for each level of $V$ (and vice versa); and if $U$ and $V$ are categorical, it is a fixed effect and therefore absorbed into $\mathbf{X}_i\boldsymbol{\beta}$.

[9] We top-code TSLB because in the preliminary modelling we found that its smooth functions exhibited noticeably more variability after around 11 years due to a lack of data; the choice is also consistent with Retherford et al. (2010) (see Section 2.1).

**Figure 1: Decision tree to illustrate the different forms that $\boldsymbol{f_{UV}(u_i, v_i)}$, the interaction between generic variables $\boldsymbol{U}$ and $\boldsymbol{V}$, can take**




Following on from this, we present the most complex form of our GAM:

$$\begin{aligned}\text{logit}(r_i) = \mathbf{X}_i\boldsymbol{\beta} &+ f_A(a_i) + f_Y(y_i) + f_T(t_i) + \\ &f_{AY}(a_i, y_i) + f_{AT}(a_i, t_i) + f_{YT}(y_i, t_i) + \\ &f_{AQ}(a_i, q_i) + f_{YQ}(y_i, q_i) + f_{TQ}(t_i, q_i) + \\ &f_{AH}(a_i, h_i) + f_{YH}(y_i, h_i) + f_{TH}(t_i, h_i). \end{aligned} \tag{1}$$

In Section 4.3 we introduced our covariates $Q$ and $H$ as categorical variables with four and five categories respectively. For model selection (see Section 5.2), we want to allow more parsimonious models to be chosen if they fit the data sufficiently well. Therefore, in addition to testing the inclusion of these original variables (which we henceforth denote by $Q_4$ and $H_5$), we also consider simpler versions formed by combining adjacent categories; we illustrate these variants in Figure 2. At its most complex, $\boldsymbol{\beta}$ includes the intercept term and parameters representing one of each of the $Q$ and $H$ variants as well as their interaction. The first level of each categorical variable is set as the baseline category. Following on from Section 4.3.1, for subsequent analyses we replace $Q$ with the 'mean imputation' from our imputation model.

**Figure 2: Illustration of the variants of the categorical variables**

| | $Q_4$ | $Q_{3a}$ | $Q_{3b}$ | $Q_{3c}$ | $Q_{2a}$ | $Q_{2b}$ | $Q_{2c}$ |
|---|---|---|---|---|---|---|---|
| < GCSE | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| GCSE | 2 | 2 | 2 | 1 | 2 | 1 | 1 |
| A Level | 3 | 3 | 2 | 2 | 2 | 2 | 1 |
| Degree | 4 | 3 | 3 | 3 | 2 | 2 | 2 |

| | $H_5$ | $H_{4a}$ | $H_{4b}$ | $H_{4c}$ | $H_{4d}$ | $H_{3a}$ | $H_{3b}$ | $H_{3c}$ | $H_{3d}$ | $H_{3e}$ | $H_{3f}$ | $H_{2a}$ | $H_{2b}$ | $H_{2c}$ | $H_{2d}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Low | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Medium | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 1 |
| High | 3 | 3 | 3 | 2 | 2 | 3 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 |
| Very high | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 1 |
| UK-born | 5 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 2 | 2 |

*Note:* $Q$ = highest educational qualification; $H$ = HDI. Numbers and colours indicate the levels of each variant; for example, '$Q_{3b}$' indicates that the $Q$ variant is the second one (b) with three categories.

## 5.2 GAM selection and fitting

Our selection process chooses one 'best' model for each parity $p$. We use forward selection, which involves fitting and comparing various forms of Equation (1). We perform the following steps:

1. Starting with the null model, separately include the main effect of each variable, i.e., $A$, $Y$, $T$ (for $p > 0$), $Q$ and $H$, considering all of the variants in Figure 2.
2. Add the variable that gives the greatest improvement in the Bayesian Information Criterion (BIC). This is a model selection criterion that balances goodness of fit with complexity.
3. Repeat this process using the remaining variables (excluding all other $Q/H$ variants if one has been added) until none provide any further improvement.
4. Test for the inclusion of interactions between all pairs of the chosen variables, again adding them in a stepwise fashion.

There is conjecture as to the most appropriate value for the 'number of observations' $n$ to use to calculate the BIC for discrete-time EHA. Raftery et al. (1996) and Lewis and Raftery (1999) choose the number of person-years, whereas Muggeo, Attanasio, and Porcu (2009) follow Raftery (1995) and choose the number of events. We opt for the latter and set $n$ as the number of events, as the dependence among person-years from the same woman makes the former unsatisfactory.

We fit the models in `R` (R Core Team 2020) using the `gam` function within the `mgcv` package (Wood 2017). For a comprehensive introduction to GAMs and their computation, see Wood (2017); for less technical descriptions in the context of EHA see Berger and Schmid (2018) and Bender, Groll, and Scheipl (2018). We use the restricted maximum likelihood (REML) estimation method as it automatically corrects the EDF (see Section 3) to account for uncertainty in the smoothing parameters (Wood 2017). Also, we weight the person-years using cross-sectional Wave 1 weights, standardised to the sample size.

### 5.3 Quantitative comparison with standard approaches

One aspect of our second aim is to quantify, for each parity, the impact of estimating smooth terms compared to using existing methods. We do this for the chosen models determined from Sections 5.1–5.2, with the following steps:

1. If the model has one continuous variable, refit it with a range of methods taken from the literature reviewed in Section 2 (see Appendix B).
2. If the model has two or more continuous variables not interacting with each other, repeat Step 1 for each variable in turn, varying its estimation method while keeping the other effects as smooth terms.
3. If the model has at least one interaction between any pair of these continuous variables, for each interacting pair refit the model with each combination of their representations from Step 2, while again keeping the other effects as smooth terms.
4. Compare the set of model variants using the BIC to assess which (combinations of) representations perform best for the different variables for each parity.

We fit all of the variants using the `gam` function.

## 6. Results

In this section we describe our findings from applying the methodology described in Section 5 and examine the results graphically. Implementing the model selection process detailed in Section 5.2, in Section 6.1 we present our chosen GAMs, providing a visual illustration of the smooth terms in Section 6.2. In Section 6.3 we discuss the results of the quantitative comparison with existing methods explained in Section 5.3, while in Section 6.4 we contrast the fitted probabilities graphically to support the conclusions.

## 6.1 Chosen GAMs

In Table 1 we present the chosen models – we will refer to the chosen model for parity $p$ as M$p$, for $p \in \{0,1,2,3+\}$. We represent each model as the sum of the main effects and interactions (written $U$ and $UV$ respectively for variables $U$ and $V$) that are included, using the same covariate initials as those in Section 5.1. Table 1 also displays the EDF corrected for smoothing parameter uncertainty (see Section 5.2), which we break down by model term. Appendix C summarises the steps for fitting the chosen GAMs in R.

**Table 1: Chosen models of progression to first birth (M0), second birth (M1), third birth (M2), and fourth or higher-order birth (M3+)**

| Model | Model description | Corrected effective degrees of freedom | |
|---|---|---|---|
| M0 | $A + Q_4 + Y + AQ_4 + AY$ | (1.00) + 7.67 + 3.00 + 5.90 + 12.85 + 32.67 | = 63.10 |
| M1 | $T + A + Y + Q_{2c} + AQ_{2c} + TQ_{2c}$ | (1.00) + 7.83 + 6.32 + 6.45 + 01.00 + 04.60 + 5.30 | = 32.49 |
| M2 | $A + T + Y + Q_{2c}$ | (1.00) + 5.62 + 7.51 + 6.19 + 01.00 | = 21.32 |
| M3+ | $A + Y + T + H_{2d} + Q_{2a}$ | (1.00) + 5.51 + 5.97 + 7.64 + 01.00 + 01.00 | = 22.12 |

*Note*: The corrected effective degrees of freedom are estimated by gam and disaggregated by model term (with the intercept in parentheses); values are given to 2 decimal places so sums may not be exact. $A$ = age, $Q$ = highest educational qualification, $Y$ = year, $T$ = time since last birth, $H$ = HDI; see Figure 2 for $Q$ and $H$ variant definitions. Single letters are main effects, pairs of letters are interaction effects. The terms are specified in the order that they were added in the forward selection.

We observe that all chosen models include the available clocks, i.e., age, period, and TSLB (for $p > 0$). These variables are added first in the forward selection process, apart from in M0 where qualification is added between age and period, providing strong support for their importance. There is also a simplification of the models across parity: M0 is the only model to include a 2D interaction (between age and period), while qualification is involved in interactions with age and TSLB in M0 and M1 but only appears as a main effect in M2 and M3+. Further testing suggests that this decreasing complexity can be partly explained by the declining number of birth events with parity. Therefore, care should be taken not to overinterpret the simpler models as indicators of simpler drivers of the decision-making processes for higher-order births.

### 6.2 Graphical representation of smooth terms from chosen GAMs

#### 6.2.1 Model of progression to first birth (M0)

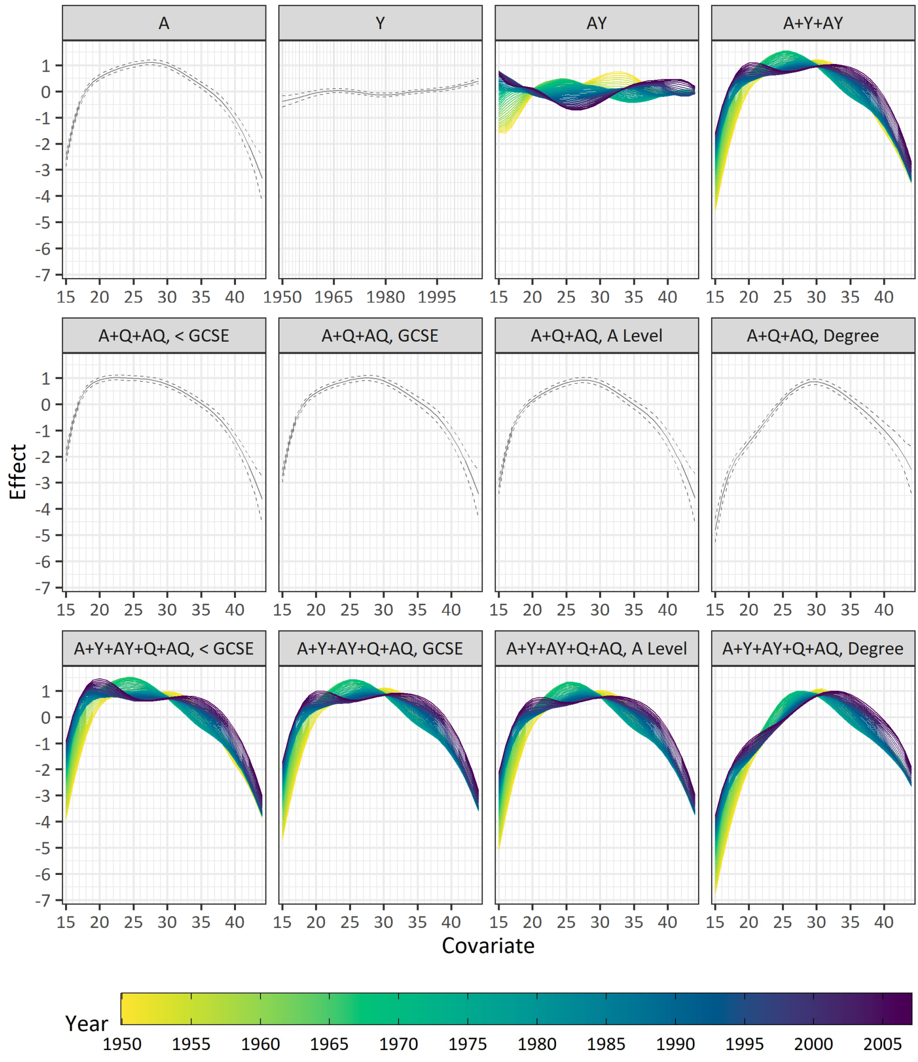
We consider M0 separately due to the presence of the 2D interaction, which complicates the interpretation of the other smooth terms. In Figure 3 we present the estimated main effects and interactions, both in isolation and combination (these are on the logit scale). We note that trends in the effects correspond to trends in the likelihood of a first birth under M0.

The first row in the figure shows the main effects of age ($A$) and year ($Y$), the $AY$ interaction, and their sum. The roughly symmetric rise and fall in the age effect is clear, giving rise to the standard bell-shaped curve on the probability scale (see Section 2.2). The mid-1960s peak in both the year effect and the $AY$ interaction for women in their 20s allows the 1960s baby boom (e.g., see Phillipson 2007 and ONS 2015) to be captured in the combined plot. This plot also highlights the declining fertility at these ages in the years since and the rises above age 30 (see ONS 2020b), and the recent emergence of the slight "hump" that had previously appeared for younger women (Chandola, Coleman, and Hiorns 1999).

The second row of Figure 3 shows the additive effect of age, qualification ($Q_4$), and the $AQ_4$ interaction for the qualification categories. As described in Section 4.3.1, we expect to see greater postponement for the more highly educated women. Indeed, the smooth effects change form from an asymmetric curve peaking at $A \approx 23$ for the '< GCSE' category, to roughly symmetric curves peaking in the late 20s for the 'GCSE' and 'A Level' categories, to a highly peaked asymmetric curve peaking at $A \approx 30$ for the 'Degree' category.

Adding $Y + AY$ to these effects gives the row 3 plots, which are equivalent to the logit of the fitted probabilities shifted by the negative of the M0 intercept. We note that as the $AY$ interaction is independent of qualification level, the recent bimodal age pattern (two peaks, one at younger ages and one at older ages) propagates through to the fitted probabilities. The shifting shapes of the age curves in row 2 mean that the earlier peak will dominate for the '< GCSE' category, while the later peak will dominate for the 'Degree' category. Appendix D provides plots of the fitted probabilities by calendar year and cohort year of birth, which demonstrate the weakening of this bimodality with qualification level.

**Figure 3: Estimates of smooth terms from the model of progression to first birth (M0)**




*Note:* Estimates are solid lines while dashed lines are approximate 95% confidence intervals. The scale represents an additive effect on the logit of the conditional probability of a birth event under M0 given the included covariates and being in parity 0. The effects are plotted against age ($A$, in years) or calendar year ($Y$, second subplot in row 1 only), and in some cases for a given highest educational qualification ($Q$) or value of $Y$ (coloured plots). Single letters are main effects, pairs of letters are interaction effects. Ranges of covariate values/covariate combinations match those observed in the data.

### 6.2.2 Models of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+)

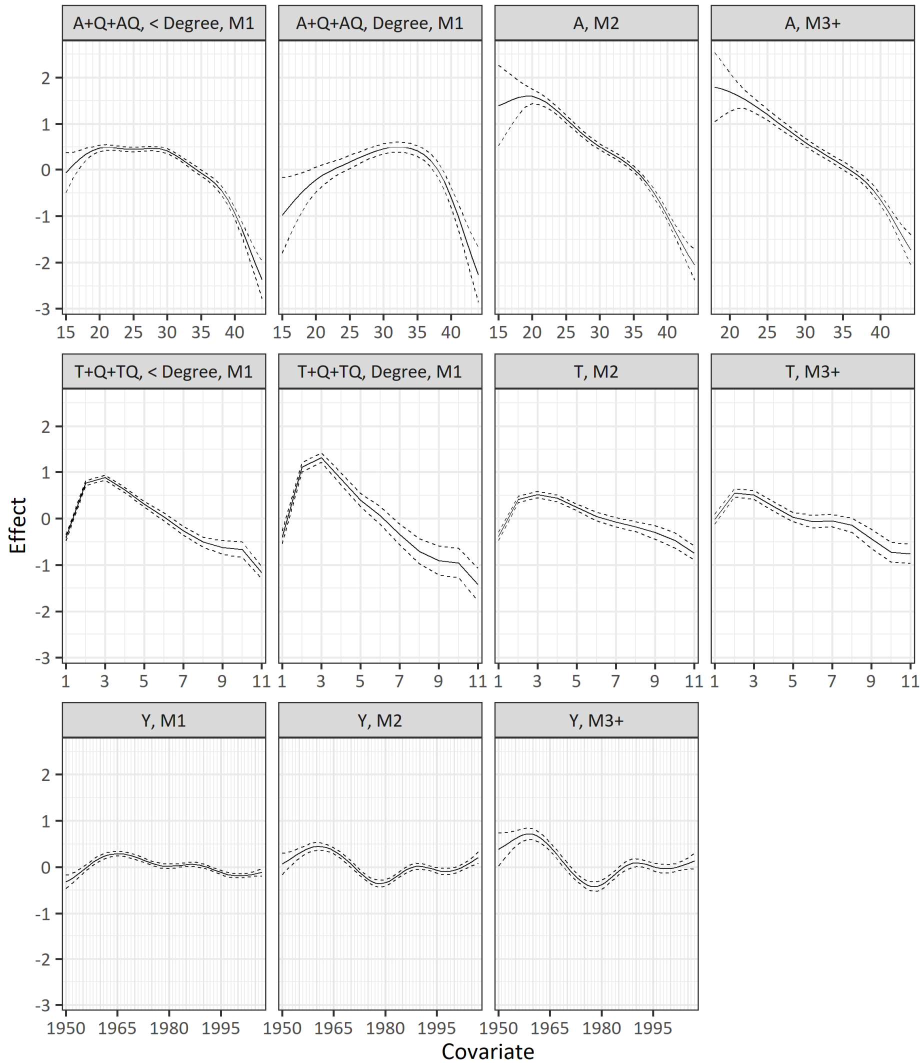
Next, in Figure 4 we present the estimates of the smooth terms in M1, M2, and M3+. This is straightforward due to the absence of 2D interactions. As with M0, trends in the effects correspond to trends in the likelihood of a second (M1), third (M2), and fourth or higher-order birth (M3+). First, we consider the age effects from each model (row 1), where for M1 we plot the additive effect $A + Q_{2c} + AQ_{2c}$ similarly to row 2 of Figure 3. The M1 '< Degree' curve is positive from the youngest ages and reasonably flat initially; this is very different to the steep increase for M0. However, the 'Degree' curve increases gradually to peak at $A \approx 32$ before proceeding to decline more steeply. This support for the delaying of second births by degree-educated women is consistent with their postponement of motherhood (see Section 6.2.1).

The effect for M2 and M3+ is an almost linear decline, with a steeper drop-off at older ages and large uncertainty at younger ages due to sparse data. This decreasing complexity of the age effect with parity is supported by its decreasing EDF values in Table 1. The general age patterns also correspond well with analyses such as Andersson (2004).

In the second row of Figure 4 we present the effect of time since last birth ($T$). The underlying effect is highly non-linear and complex, having the largest EDF in M1–M3+ (Table 1). Consistent with Section 2.1, for each model the likelihood of a subsequent birth peaks at $T = 2$ or $T = 3$ and then declines. The peaks weaken as parity increases, becoming more compact for M3+ where we note a resurgence in the decline for birth intervals of 6–9 years. This increased likelihood of a fourth or higher-order birth after a longer birth interval is likely due to repartnering (e.g., see Andersson 2021). The M1 curves indicate that degree-educated women are more likely than less-educated women to have a shorter first birth interval, and less likely to have a longer interval. This acceleration of childbearing is an illustration of the 'time-squeeze' phenomenon (Kreyenfeld 2002; Klesment et al. 2014) that has been detected for England and Wales (Rendall and Smallwood 2003).

We consider the period effects in the third row of Figure 4, which are consistently the second most complex after TSLB (Table 1). There is a clear distinction between M1 and M2/M3+. While the curve of the former increases steeply from a negative effect and peaks in the mid-1960s before slowly declining, the latter curves start from positive effects and peak earlier in around 1960 before declining steeply, with this drop more extreme for M3+. These differential trends are consistent with the work of Ní Bhrolcháin (1987, Figure 1). All of the effects then exhibit a second, smaller peak between the late 1970s and late 1990s, which again increases in intensity with parity before rising again in the early 2000s. These features are evident in the trend of the total fertility rate (ONS 2020b).

**Figure 4: Estimates of smooth terms from the models of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+)**



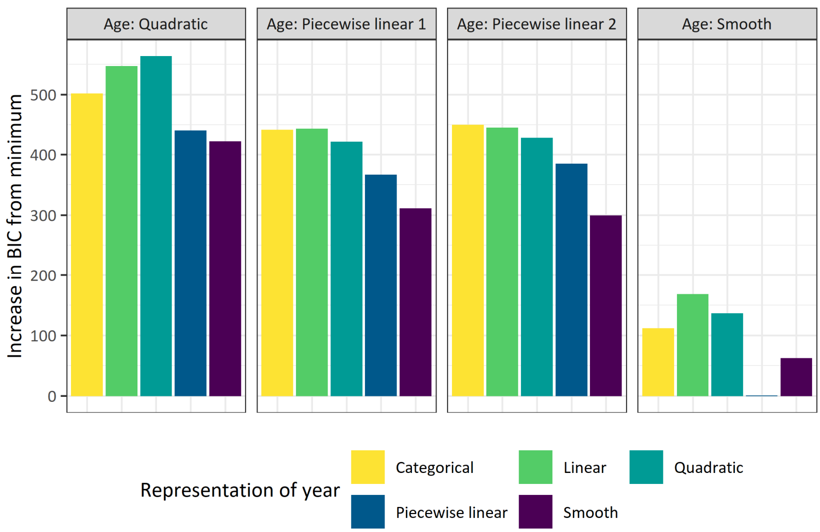

*Note:* Estimates are solid lines while dashed lines are approximate 95% confidence intervals. The scale represents an additive effect on the logit of the conditional probability of a birth event under the specified model given the included covariates and current parity. The subplots in rows 1–3 relate to $A$ = age (in years), $T$ = time since last birth (in years), and $Y$ = calendar year respectively; $Q$ = highest educational qualification; subplots within rows are ordered by model number, i.e., M1, M2, M3+. Single letters are main effects, pairs of letters are interaction effects. Ranges of covariate values/covariate combinations match those observed in the data.

### 6.3 Quantitative comparison with standard approaches

#### 6.3.1 Model of progression to first birth (M0)

As M0 includes two interacting continuous variables (age and year), in accordance with Section 5.3 we refit it with all possible combinations of their representations (see Appendix B). Figure 5 presents the increases in the BIC from the minimum value (i.e., the preferred combination), which is a smooth age effect and piecewise linear (PWL) year effect. Note that in this and subsequent figures we plot the alternative representations in order of increasing complexity. Note also that we omit the categorical age effect values as they are large ($\approx 2000$ here). To aid interpretation, we note that Raftery (1995) defines a "significant" improvement in BIC, where a decrease of 6–10 is strong and above 10 very strong.

**Figure 5: BIC increases for the model of progression to first birth (M0) comparative analysis**




*Note:* Each subplot relates to a different representation of age (in years); each colour relates to a different representation of calendar year. For ease of comparison, the vertical axis is the increase in the BIC from the minimum across all combinations of age and year representations, which has value 0 (age smooth, year piecewise linear).

There is clearly more variability in the BIC increase for a given year representation compared to a given age representation. Therefore, the choice of a smooth age effect, which is strongly preferable across the possible year representations, is highly consequential. The choice of year effect is less important; however, for a smooth age effect there are still large increases in BIC for anything other than a PWL year effect. If age is not smooth, a smooth effect for year is preferred. Repeating the analyses excluding the age–period interaction from M0, we find that smooth and PWL year effects are equally preferable across the age representations, followed by a categorical effect.

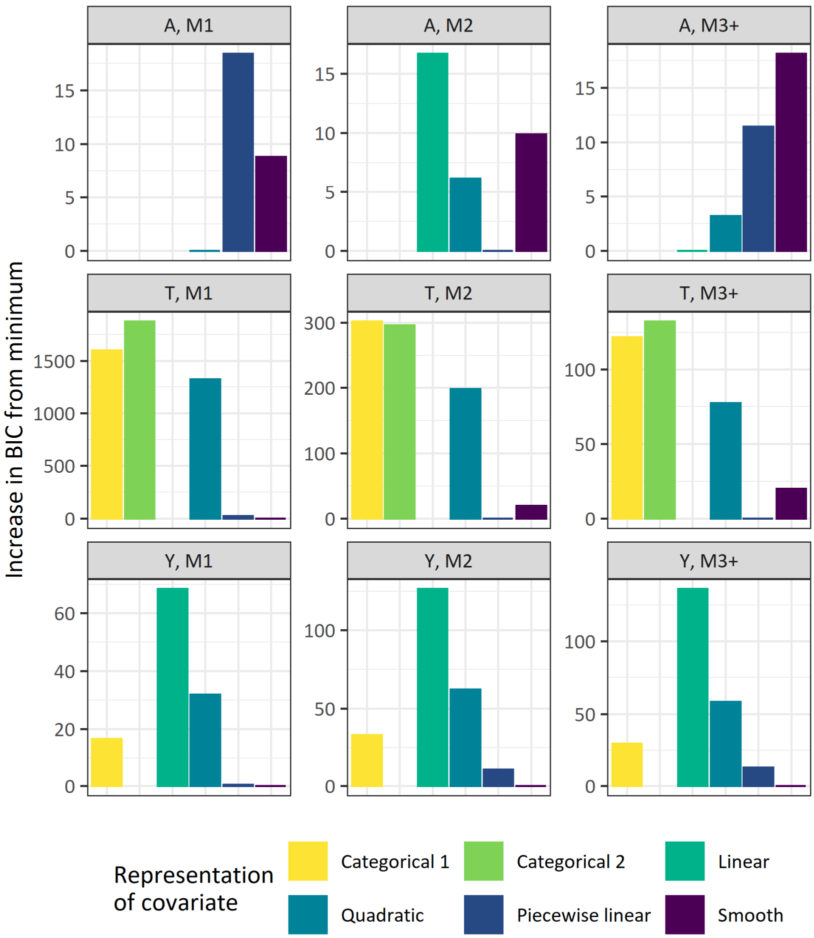
Following on from the discussion in Section 5.2, we also repeat this quantitative comparison using two alternative information criteria: the BIC with $n$ as the number of person-years rather than the number of events, and the Akaike Information Criterion (AIC) (results available upon request). These penalise model complexity more and less severely, respectively, compared to the original BIC. Using the alternative BIC does not change the findings; however, the AIC prefers a smooth year effect if age is smooth, and otherwise prefers smooth and PWL year effects roughly equally. Linear and quadratic effects also perform worse on the whole, due to their minimal parameter requirements.

#### 6.3.2 Models of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+)

As none of the chosen higher-order models interact pairs of continuous variables, we vary the representation of one effect while estimating the others as smooth terms (see Section 5.3). We present the BIC increases in Figure 6 for each variable–model combination. Considering age first (row 1), we again exclude the categorical values due to their magnitude. We note that the increases are substantially smaller than those in Figure 5, likely due to the lack of a 2D interaction. For M1 a quadratic age effect is preferred, followed closely by a smooth representation. PWL splines are least desirable here; however, for M2 they give the lowest BIC, with the quadratic form not far behind. For M3+ a linear effect is preferred, with the more complex methods performing increasingly worse. This is consistent with the simplification of the smooth age effect with parity (see Section 6.2).

In Section 6.2 we saw that the effects of TSLB and period were more consistent across parity, and this is reflected here. For TLSB (Figure 6, row 2) a smooth effect is preferred for M1, and PWL splines for M2 and M3+, with a strong improvement of around 20 in each case. The quadratic and categorical effects perform very poorly. A smooth representation is preferred for the year effect across the parities, with PWL splines second (albeit marginally for M1). Interestingly, the categorical year effect outperforms both polynomial forms.

**Figure 6: BIC increases for the model of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+) comparative analyses**




*Note:* For each variable–model combination, the effects of the other variables in the model are estimated as smooth terms. For ease of comparison, within each subplot the vertical axis is the increase in BIC from the minimum, which has value 0. The subplots in columns 1–3 relate to M1, M2, and M3+ respectively; the subplots in rows 1–3 relate to $A$ = age (in years), $T$ = time since last birth (in years), and $Y$ = calendar year, respectively. The bars corresponding to the categorical age representations are omitted due to their magnitude. Note the different scales.

The additional comparisons using the alternative BIC (see Section 6.3.1) do not change the preferred representations, apart from reversing the marginal case for period under M1; however, for M2 we find that quadratic and PWL age effects are now equally preferable, followed by the linear representation. Under the AIC the preferred age effects change quite drastically, with smooth representations chosen for all models and the less complex methods performing increasingly worse; smooth effects for TSLB also perform comparatively better for M2 and M3+.

## 6.4 Graphical comparison with standard approaches

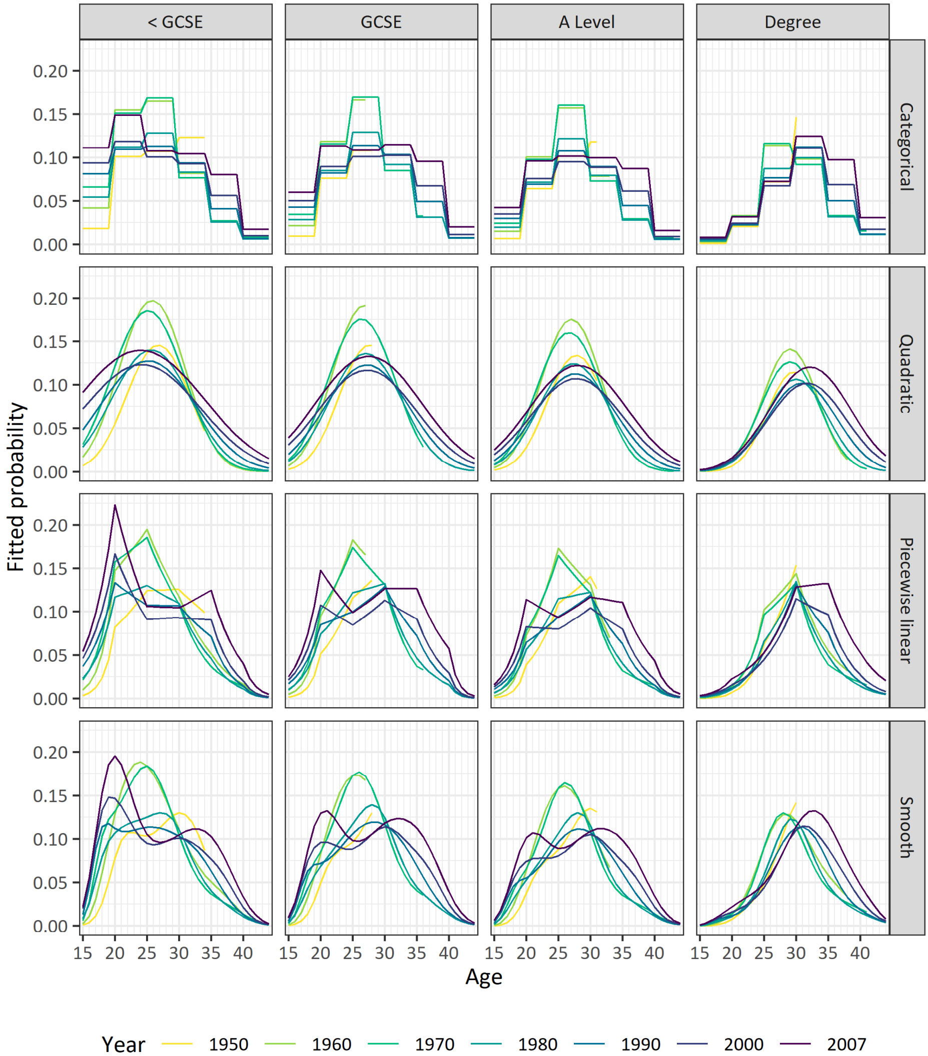
### 6.4.1 Model of progression to first birth (M0)

In this section we illustrate the findings from Section 6.3 by plotting fitted probabilities from the model variants. Starting with M0, in Figure 7 we present fitted probabilities by age, qualification, and selected years for the variants with a smooth year effect, allowing us to contrast the alternative age representations. We present the second PWL variant due to its lower BIC.

Figure 5 shows that the complex approaches perform best, and this is evident here. The categorical representation is clearly inappropriate, given the highly peaked nature of the underlying effect, thus having the worst fit despite its large EDF. With half the EDF and an improved fit the quadratic assumption is more reasonable; however, it is unable to capture the recent bimodality (see Section 6.2.1). PWL splines have the flexibility to do this, although the success strongly depends on the suitability of the knot locations, where the probability curves can change direction sharply and unnaturally. The smooth representation has the largest EDF, best fit, and lowest BIC, highlighting its ability to estimate non-linear patterns and complex interactions precisely and efficiently.

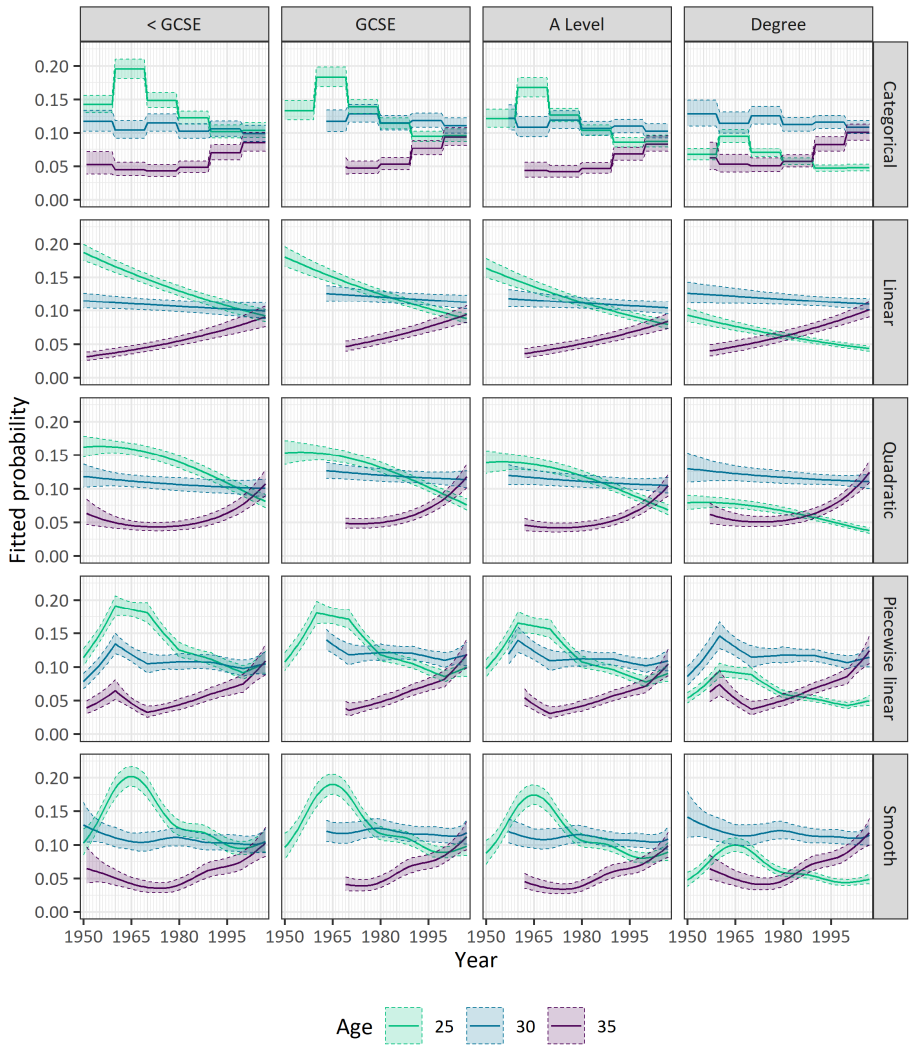
Next, we compare the alternative year representations, recalling from Section 6.3.1 that there was less to choose between them. We plot the fitted probabilities by year, qualification, and selected ages (25, 30, 35) for the variants with a smooth age effect in Figure 8, together with approximate 95% confidence intervals. The 1960s baby boom (see Section 6.2.1) is clearly experienced most strongly by those around age 25, with the age 30 pattern relatively flat across the entire period. The subsequent decline and rise for ages 25 and 35 respectively lead to all three intervals overlapping in the most recent years for the sub-Degree qualifications.

**Figure 7: Fitted probabilities for the model of progression to first birth (M0) comparative analysis, varying the representation of age**




*Note:* The effect of year is estimated as a smooth term throughout. Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against age (in years) for a given calendar year (colour), highest educational qualification (column) and representation of age (row). Ranges of covariate combinations match those observed without gaps in the data.

**Figure 8: Fitted probabilities and approximate 95% confidence intervals for the model of progression to first birth (M0) comparative analysis, varying the representation of year**




*Note:* The effect of age is estimated as a smooth term throughout. Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against calendar year for a given age (colour), highest educational qualification (column), and representation of year (row). Ranges of covariate combinations match those observed without gaps in the data. The shaded areas indicate the approximate 95% confidence interval associated with the corresponding fitted probability.

The PWL approach captures the main features of the smooth representation, with slight differences where there is sparse data, such as at the start of the year curves. The relative flatness of the year effect means that the aforementioned angularity of the PWL curves appears less stark here. These observations, together with the method having a considerably smaller EDF and only a slightly poorer fit, support its preference under the BIC. The increased EDF for the categorical representation over the polynomials allows it to detect peaks and troughs that they cannot, resulting in an improved fit and therefore BIC. Appendix E provides further analyses by age, year, and qualification level that estimate various simplified effects by taking appropriate weighted averages of the probabilities.

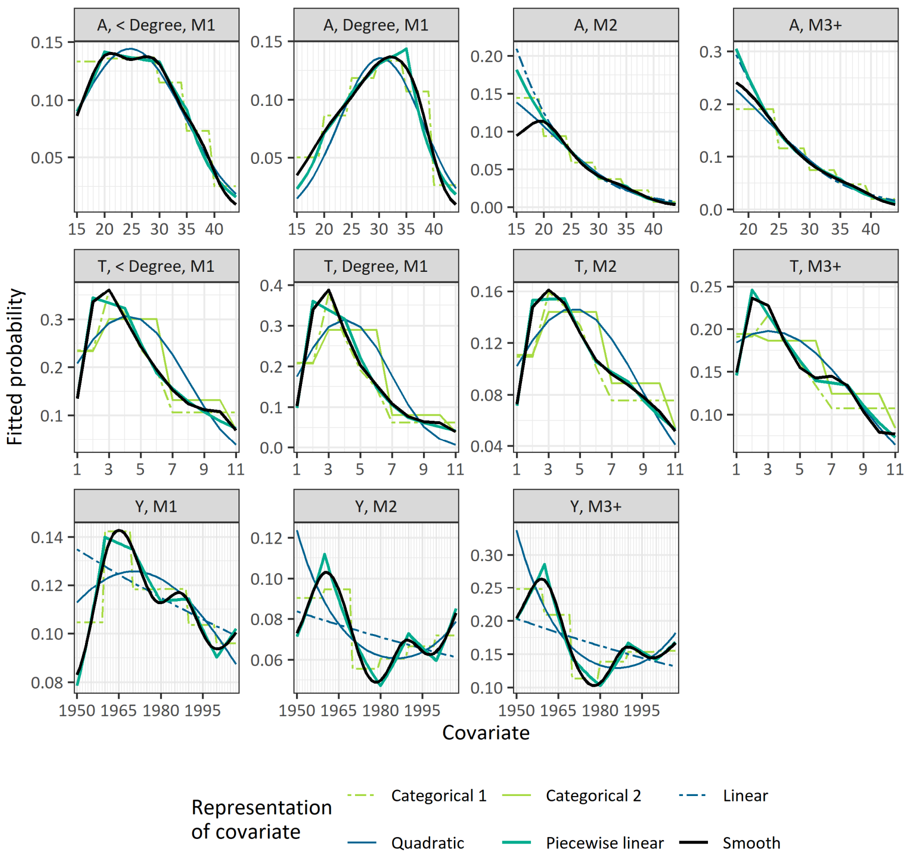
### 6.4.2 Models of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+)

In Figure 9 we graphically present the fitted probabilities from the higher-order models, overlaying the effects for each variable–model combination and fixing the other variables (see caption). The age curves (row 1) align very closely across the alternative methods, aside from the categorical representation, which performed considerably worse (see Section 6.3.2). The largest discrepancies occur where data is sparse, namely at young ages in the M1 'Degree' plot and the M2 and M3+ plots. This explains why the simple quadratic form is preferred for M1, as it can approximate the smooth '< Degree' curve very well, and the smooth 'Degree' curve well for older ages. Although the PWL representation has a similar EDF to the smooth effect, the considerably worse fit makes it less preferable; this is likely caused by its inability to precisely capture the 'Degree' peak. For M2, the smooth effect fits only slightly better than the PWL representation, despite an increase in EDF of 1.6 (required to estimate the initial peak), which is why the latter is preferred. For M3+ it is clear that little is gained from choosing anything more complex than a linear effect.

For TSLB (Figure 9, row 2), the smooth and PWL probabilities align closely, explaining their similar performance in Figure 6. The smooth effect is likely preferred for M1, as it captures the strong peak at $T = 3$ which the PWL splines, with even knot locations, cannot; the $TQ_{2c}$ interaction probably intensifies this advantage. PWL splines approximate the M2 and M3+ smooth terms better, aided by the less pronounced peak of the former and the change points of the latter aligning well with the knot locations. The categorical and quadratic curves are clearly poor approximations of the underlying effect, explaining the much larger BIC values.

For period (row 3), the PWL representation approximates the M1 smooth term very well, performing worse for the higher parities where the knot locations correspond less closely with the earlier peaks (see Section 6.2.2). The ability of the categorical representation to capture the undulations of the underlying effect better than the less flexible polynomial forms, but worse than the more complex representations, is also evident (see Appendix E).

**Figure 9: Fitted probabilities for the model of progression to second birth (M1), third birth (M2), and fourth or higher-order birth (M3+) comparative analyses**




*Note:* For each variable–model combination, the effects of the other variables in the model are estimated as smooth terms. The subplots in rows 1–3 relate to $A$ = age (in years), $T$ = time since last birth (in years), and $Y$ = calendar year, respectively; subplots within rows are ordered by model number, i.e., M1, M2, M3+. Fitted probabilities are conditional probabilities of a birth event under the specified model given the included covariates and current parity; they are plotted against the indicated covariate for a given model, covariate representation, and in some cases highest educational qualification. Note that we set $A = 25$, $T = 1$, and $Y = 1970$ when they are not the indicated covariate, and fix any categorical variables at their first levels unless specified otherwise. Also note the different scales.

## 7. Discussion

In this paper we present an application of GAMs to discrete-time EHA in the context of birth events, which to the best of our knowledge is the first of its kind in the literature. Our flexible approach allows every continuous variable to be incorporated as a smooth term, its shape and complexity being completely data-driven. Fitting a smooth function which is not restricted to a narrow class (e.g., polynomial) actually reflects what we believe about how continuously indexed quantities vary in the real world; i.e., smoothly but not necessarily coinciding with mathematical convenience.

This latter point is evident when we consider the substantive insights that we gain through our approach, such as accurately detecting the recent bimodality in the progression to first birth age curve caused by the earlier childbearing of less-educated women (see Appendices D and E); identifying the most likely first birth interval; and appropriately capturing the undulations in the time-pattern of fertility for higher parities. These advantages all lead to strong improvements in the BIC and therefore have quantitative as well as visual support. The inability of piecewise linear splines to capture these features with sufficient precision appears to be largely due to a lack of correspondence between the knot locations and the change points in the underlying effect. This could be improved with prior knowledge. However, in the absence of this, the relative insensitivity of GAM fits to knot location is an undoubted advantage that makes GAMs simple from a modeller perspective.

Our comparative analysis (Sections 6.3–6.4) comprehensively assesses the impact of using GAMs against the existing methods of discretisation, polynomials, and piecewise linear splines. There are four key findings:

1. Smooth terms offer the greatest gains when estimating highly complex and non-linear underlying effects and interactions (in our case, these involve age and period).
2. Relatively complex underlying effects, e.g., for time since last birth (TSLB), can sometimes be estimated appropriately using less sophisticated methods like piecewise linear splines, but the success depends on their specification.
3. Simpler underlying effects, e.g., higher-order age effects, can be very well approximated using more restrictive parametric forms like polynomials.
4. Categorical representations perform very poorly for age and TSLB, which is concerning given their ubiquitous adoption in this field.

These findings hold when we use an alternative BIC that penalises complexity more harshly, while the AIC prefers smooth higher-order age effects, opposing point 3. However, the highly consistent fitted probabilities across method (Figure 9, row 1) suggest that the smooth representations overfit to the data and are therefore unnecessary.

A key strength is the selection of comparator methods to coincide with examples from the literature, making our results highly relevant for informing future methodological development in discrete-time EHA both for births and more generally. However, we emphasise that these findings relate to a single dataset and a subset of all possible alternative methods and their specifications. As such, it will be vital to perform similar comparisons using other datasets and representations to test the applicability of our conclusions in a range of contexts. We anticipate that our findings regarding age and TSLB will persist, as their broad patterns are consistent across time and space, while for period they may vary, as the degree of non-linearity strongly depends on the chosen country and historical period.

A further advantage is the estimation of bivariate smooth terms, which are largely unexplored in the broader literature. Their consideration proves beneficial here, with the smooth 2D age–period interaction selected for inclusion in our model of first birth (M0). This elevates the sophistication of our inference, allowing us to get a visual handle on the complex age–period relationship. With the age–qualification interactions also preferably estimated using smooth methods, the ability of GAMs to efficiently estimate interactions is clearly a valuable property.

Despite this efficiency, we acknowledge that GAMs are more computationally expensive than existing methods, particularly when 2D interactions are present. As such, for the models compared in Section 6.3 the biggest time differential (of around 30 seconds) occurs for M0. This is not unreasonable, but would naturally scale with the number of 2D interactions. We argue that this is not problematic, first because this number is likely to be small: indeed, across our chosen models only one is included. Interactions between clocks are also likely to dominate as they are fundamental demographic variables, so alternative 2D interactions are unlikely to be selected. Second, GAMs are flexible in terms of the inclusion of smooth terms, with a combination of smooth main effects and linear interactions being perfectly acceptable and carrying much less computational cost.

This latter point highlights the fact that the models in our comparative analysis are all special cases of GAMs and so can be directly compared. As such, our proposal encapsulates the existing methods and expresses them in a more general framework that allows for smoothness in a formal, systematic way. Thus, GAMs retain the interpretability and transparency of the well-known generalized linear model framework, but with the advantage of greater flexibility. The resulting intuitiveness of the software (see Appendix C for example code) means that achieving competency requires a minimal time investment from the user, with simple decision trees such as Figure 1 providing guidance as to which interactions are suitable for different variable combinations. This is in contrast to less interpretable machine learning techniques such as random forests and neural networks, where the learning curve is much steeper and the main focus is on predictive performance rather than the explanatory power of the model.

We recognise that there are limitations to modelling progression to each birth order separately; for example, for the interpretation of educational effects (Kravdal 2001). Modelling the transitions jointly and accounting for unobserved differences between women via random effects would have been possible using GAMMs (see Section 3). However, the additional complexity would have increased computation time considerably and likely made it infeasible to understand the relative importance of the smooth terms for each parity. Consequently, given that the aim of this paper is to investigate the applicability and usefulness of GAMs for birth event analyses, it is preferable to model each transition separately.

Next, we discuss our use of covariates. Referring to the “five clocks” of Raftery et al. (1996), we incorporated all of them except cohort (using age, period, TSLB, and parity). We also considered taking a cohort approach (results available upon request). On performing these alternative analyses we find that similar models are chosen and a smooth cohort effect is preferred for most parities. The effect is simpler than that of period, with fewer undulations especially for the model of second birth where a quadratic term suffices. For higher parities, the smooth terms capture peaks for the mid-1920s and 1960s cohorts, and a trough for those cohorts born around 1950. We note that including both period and cohort as smooth terms is problematic due to the age–period–cohort identification problem (e.g., see Fienberg 2013).

In terms of education, we recognise the potential for our highest qualification variable to be endogenous; however, including education in this way is a standard approach in the demography literature, despite this known issue. Constructing time-varying enrolment and attainment variables is preferable (see Section 4.3.1), but in our case this would require imputation due to incomplete education histories, which has its own drawbacks (see Kravdal 2004). An instrumental variable strategy to correct for endogeneity is possible with GAMs (e.g., see Marra and Radice 2011), but this is beyond the scope of this paper.

In conclusion, our investigation into the application of GAMs to discrete-time EHA provides quantitative evidence that smooth methods are required to estimate complex underlying effects with sufficient precision and efficiency. We therefore strongly recommend the use of GAMs in this context – for initial exploratory analyses at the very least. This reiterates the guidance of Fahrmeir and Knorr-Held (1997), utilising recent computational developments to propose a transparent and systematic approach for selecting the most appropriate representation of each effect. The properties and benefits of GAMs could also prove useful in countries with sparse or less accurate data, or for estimating complex effects in other demographic processes (e.g., Hilton et al. (2019) applies GAMs to mortality modelling). This discussion strongly motivates our belief that

GAMs have the potential to become commonplace not only in the discrete-time EHA literature but also in the demographer's toolkit.

## 8. Acknowledgements

We thank an associate editor and two anonymous reviewers for their helpful comments on an earlier version of this paper. The doctoral programme of the first author was funded by the Engineering and Physical Sciences Research Council (award reference 1801045). This work was partly supported by the Economic and Social Research Council (ESRC) FertilityTrends project (grant ES/S009477/1) and the ESRC Centre for Population Change - phase II (grant ES/K007394/1). The UKHLS is an initiative funded by the ESRC and various government departments, with scientific leadership by ISER, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. We use code provided by Ann Berrington and Juliet Stone, which was used in the analyses for Berrington, Stone, and Beaujouan (2015), to clean the fertility histories.

## References

Andersson, G. (2004). Childbearing developments in Denmark, Norway, and Sweden from the 1970s to the 1990s: A comparison. *Demographic Research, Special Collection* 3(7): 155–176. doi:10.4054/DemRes.2004.S3.7.

Andersson, L. (2021). Lifetime parenthood in the context of single- and multiple-partner fertility. *Advances in Life Course Research* 47(100355). doi:10.1016/j.alcr.2020.100355.

Begall, K. and Mills, M. C. (2013). The influence of educational field, occupation, and occupational sex segregation on fertility in the Netherlands. *European Sociological Review* 29(4): 720–742. doi:10.1093/esr/jcs051.

Bender, A., Groll, A., and Scheipl, F. (2018). A generalized additive model approach to time-to-event analysis. *Statistical Modelling* 18(3–4): 299–321. doi:10.1177/1471082X17748083.

Berger, M. and Schmid, M. (2018). Semiparametric regression for discrete time-to-event data. *Statistical Modelling* 18(3–4): 322–345. doi:10.1177/1471082X17748084.

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. *Demographic Research* 33(26): 733–764. doi:10.4054/DemRes.2015.33.26.

Billari, F.C., Goisis, A., Liefbroer, A.C., Settersten, R.A., Aassve, A., Hagestad, G., and Spéder, Z. (2011). Social age deadlines for the childbearing of women and men. *Human Reproduction* 26(3): 616–622. doi:10.1093/humrep/deq360.

Blossfeld, H.-P. and Huinink, J. (1991). Human capital investments or norms of role transition? How women's schooling and career affect the process of family formation. *American Journal of Sociology* 97(1): 143–168. doi:10.1086/229743.

Chandola, T., Coleman, D.A., and Hiorns, R.W. (1999). Recent European fertility patterns: Fitting curves to 'distorted' distributions. *Population Studies* 53(3): 317–329. doi:10.1080/00324720308089.

Ellison, J. (2021). Stochastic modelling and projection of age-specific fertility rates [PhD thesis]. Southampton: University of Southampton, School of Mathematical Sciences. http://eprints.soton.ac.uk/450468/.

Erlandsson, A. (2017). Child home care allowance and the transition to second- and third-order births in Finland. *Population Research and Policy Review* 36(4): 607–630. doi:10.1007/s11113-017-9437-1.

Fahrmeir, L. (1994). Dynamic modelling and penalized likelihood estimation for discrete time survival data. *Biometrika* 81(2): 317–330. doi:10.1093/biomet/81.2.317.

Fahrmeir, L. and Knorr-Held, L. (1997). Dynamic discrete-time duration models: Estimation via Markov Chain Monte Carlo. *Sociological Methodology* 27(1): 417–452. doi:10.1111/1467-9531.271033.

Fahrmeir, L. and Wagenpfeil, S. (1996). Smoothing hazard functions and time-varying effects in discrete duration and competing risks models. *Journal of the American Statistical Association* 91(436): 1584–1594. doi:10.1080/01621459.1996.10476726.

Ferrara, G. and Vidoli, F. (2017). Semiparametric stochastic frontier models: A generalized additive model approach. *European Journal of Operational Research* 258(2): 761–777. doi:10.1016/j.ejor.2016.09.008.

Fienberg, S.E. (2013). Cohort analysis' unholy quest: A discussion. *Demography* 50: 1981–1984. doi:10.1007/s13524-013-0251-z.

Fiori, F., Graham, E., and Feng, Z. (2014). Geographical variations in fertility and transition to second and third birth in Britain. *Advances in Life Course Research* 21: 149–167. doi:10.1016/j.alcr.2013.11.004.

Galimberti, G. (2002). Time-varying coefficient models for event history analysis. *Statistica Applicata* 14(1): 81–96.

Hank, K. (2002). Regional social contexts and individual fertility decisions: A multilevel analysis of first and second births in western Germany. *European Journal of Population* 18: 281–299. doi:10.1023/A:1019765026537.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science* 1(3): 297–310. doi:10.1214/ss/1177013604.

Hastie, T. and Tibshirani, R. (1990). *Generalized additive models*. London/New York: Chapman and Hall.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient Models. *Journal of the Royal Statistical Society: Series B* 55(4): 757–796. doi:10.1111/j.2517-6161.1993.tb01939.x.

Hilton, J., Dodd, E., Forster, J.J., and Smith, P.W.F. (2019). Projecting UK mortality by using Bayesian generalized additive models. *Journal of the Royal Statistical Society: Series C* 68(1): 29–49. doi:10.1111/rssc.12299.

Hinde, A. (1998). *Demographic methods*. London: Arnold.

ISER, Institute for Social and Economic Research (2020). Questionnaires: Youth and adult [electronic resource]. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires.

Ivanova, K., Kalmijn, M., and Uunk, W. (2014). Fertility after repartnering in the Netherlands: Parenthood or commitment? *Advances in Life Course Research* 21: 101–112. doi:10.1016/j.alcr.2013.08.003.

Jefferies, J., Berrington, A., and Diamond, I. (2000). Childbearing following marital dissolution in Britain. *European Journal of Population* 16: 193–210. doi:10.1023/A:1026529300659.

Kim, K. (2014). Intergenerational transmission of age at first birth in the United States: Evidence from multiple surveys. *Population Research and Policy Review* 33: 649–671. doi:10.1007/s11113-014-9328-7.

Kim, K. (2017). Trends and determinants of premarital conception: Love in Korea. *International Journal of Japanese Sociology* 26(1): 23–38. doi:10.1111/ijjs.12062.

Kim, K. and Park, H. (2009). Landholding and fertility in Korea: 1914–1925. *Journal of Family History* 34(3): 275–291. doi:10.1177/0363199009337998.

Klesment, M., Puur, A., Rahnu, L., and Sakkeus, L. (2014). Varying association between education and second births in Europe: Comparative analysis based on the EU-SILC data. *Demographic Research* 31(27): 813–860. doi:10.4054/DemRes.2014.31.27.

Knies, G. (2018). Understanding Society: Waves 1–8, 2009–2017 and harmonised BHPS: Waves 1–18, 1991–2009, User Guide, November 2018. Colchester: University of Essex.

Kravdal, Ø. (2001). The high fertility of college educated women in Norway: An artefact of the separate modelling of each parity transition. *Demographic Research* 5(6): 187–216. doi:10.4054/DemRes.2001.5.6.

Kravdal, Ø. (2004). An illustration of the problems caused by incomplete education histories in fertility analyses. *Demographic Research, Special Collection* 3(6): 135–154. doi:10.4054/DemRes.2004.S3.6.

Kravdal, Ø. (2007). Effects of current education on second- and third-birth rates among Norwegian women and men born in 1964: Substantive interpretations and methodological issues. *Demographic Research* 17(9): 211–246. doi:10.4054/DemRes.2007.17.9.

Kravdal, Ø. and Rindfuss, R.R. (2008). Changing relationships between education and fertility: A study of women and men born 1940 to 1964. *American Sociological Review* 73(5): 854–873. doi:10.1177/000312240807300508.

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. doi:10.4054/DemRes.2002.7.2.

Kulu, H. (2005). Migration and fertility: Competing hypotheses re-examined. *European Journal of Population* 21: 51–87. doi:10.1007/s10680-005-3581-8.

Kulu, H. (2013). Why do fertility levels vary between urban and rural areas? *Regional Studies* 47(6): 895–912. doi:10.1080/00343404.2011.581276.

Kulu, H. and González-Ferrer, A. (2014). Family dynamics among immigrants and their descendants in Europe: Current research and opportunities. *European Journal of Population* 30: 411–435. doi:10.1007/s10680-014-9322-0.

Kulu, H. and Washbrook, E. (2014). Residential context, migration and fertility in a modern urban society. *Advances in Life Course Research* 21: 168–182. doi:10.1016/j.alcr.2014.01.001.

Lappegård, T. and Rønsen, M. (2005). The Multifaceted Impact of Education on Entry into Motherhood. *European Journal of Population* 21: 31–49. doi:10.1007/s10680-004-6756-9.

Lewis, S.M. and Raftery, A.E. (1999). Bayesian analysis of event history models with unobserved heterogeneity via Markov Chain Monte Carlo. *Sociological Methods and Research* 28(1): 35–60. doi:10.1177/0049124199028001003.

Liefbroer, A.C. and Corijn, M. (1999). Who, what, where, and when? Specifying the impact of educational attainment and labour force participation on family formation. *European Journal of Population* 15: 45–75. doi:10.1023/A:1006137104191.

Lynn, P. and Borkowska, M. (2018). Some indicators of sample representativeness and attrition bias for BHPS and Understanding Society. Colchester: University of Essex (Understanding Society Working Paper 2018-01). https://www.understandingsociety.ac.uk/research/publications/524851.

Marra, G. and Radice, R. (2011). A flexible instrumental variable approach. *Statistical Modelling* 11(6): 581–603. doi:10.1177/1471082X1001100607.

McDonald, J.W. and Rosina, A. (2001). Mixture modelling of recurrent event times with long-term survivors: Analysis of Hutterite birth intervals. *Statistical Methods and Applications* 10: 257–272. doi:10.1007/BF02511651.

Morosow, K. and Trappe, H. (2018). Intergenerational transmission of fertility timing in Germany. *Demographic Research* 38(46): 1389–1422. doi:10.4054/DemRes.2018.38.46.

Muggeo, V.M.R., Attanasio, M., and Porcu, M. (2009). A segmented regression model for event history data: An application to the fertility patterns in Italy. *Journal of Applied Statistics* 36(9): 973–988. doi:10.1080/02664760802552994.

Myrskylä, M., Kohler, H. P., and Billari, F.C. (2009). Advances in development reverse fertility declines. *Nature* 460: 741–743. doi:10.1038/nature08230.

Ní Bhrolcháin, M. (1987). Period parity progression ratios and birth intervals in England and Wales, 1941–1971: A synthetic life table analysis. *Population Studies* 41(1): 103–125. doi:10.1080/0032472031000142556.

Ní Bhrolcháin, M. and Beaujouan, É (2012). Fertility postponement is largely due to rising educational enrolment. *Population Studies* 66(3): 311–327. doi:10.1080/00324728.2012.697569.

ONS, Office for National Statistics (2015). Trends in births and deaths over the last century [electronic resource]. Cardiff: ONS. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/articles/trendsinbirthsanddeathsoverthelastcentury/2015-07-15.

ONS, Office for National Statistics (2020a). Births by parents' country of birth, England and Wales: 2019 [electronic resource]. Cardiff: ONS. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/parentscountryofbirthenglandandwales/2019.

ONS, Office for National Statistics (2020b). Births in England and Wales: 2019 [electronic resource]. Cardiff: ONS. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthsummarytablesenglandandwales/2019.

ONS, Office for National Statistics (2020c). Fertility rates by parity 1934 to 2018, England and Wales [electronic resource]. Cardiff: ONS. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/adhocs/11482fertilityratesbyparity1934to2018englandandwales.

ONS, Office for National Statistics (2020d). Childbearing for women born in different years, England and Wales: 2019 [electronic resource]. Cardiff: ONS. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/childbearingforwomenbornindifferentyearsenglandandwales/2019.

Peristera, P. and Kostaki, A. (2007). Modeling fertility in modern populations. *Demographic Research* 16(6): 141–194. doi:10.4054/DemRes.2007.16.6.

Phillipson, C. (2007). Understanding the Baby Boom generation: Comparative perspectives. *International Journal of Ageing and Later Life* 2(2): 7–11. doi:10.3384/ijal.1652-8670.07227.

Piotrowski, M. and Tong, Y. (2016). Education and fertility decline in China during transitional times: A cohort approach. *Social Science Research* 55: 94–110. doi:10.1016/j.ssresearch.2015.10.001.

R Core Team (2020). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.R-project.org/.

Raftery, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology* 25: 111–163. doi:10.2307/271063.

Raftery, A.E., Lewis, S.M., Aghajanian, A., and Kahn, M.J. (1996). Event history modeling of world fertility survey data. *Mathematical Population Studies* 6(2): 129–153. doi:10.1080/08898489609525426.

Rendall, M.S. and Smallwood, S. (2003). Higher qualifications, first-birth timing, and further childbearing in England and Wales. *Population Trends* 111(Spring): 18–26.

Retherford, R., Ogawa, N., Matsukura, R., and Eini-Zinab, H. (2010). Multivariate analysis of parity progression-based measures of the total fertility rate and its components. *Demography* 47(1): 97–124. doi:10.1353/dem.0.0087.

Reyes Santías, F., Cadarso-Suárez, C., and Rodríguez-Álvarez, M.X. (2011). Estimating hospital production functions through flexible regression models. *Mathematical and Computer Modelling* 54(7–8): 1760–1764. doi:10.1016/j.mcm.2010.11.087.

Riise, B.S., Dommermuth, L., and Lyngstad, T.H. (2016). Intergenerational transmission of age at first birth in Norway. *European Societies* 18(1): 47–69. doi:10.1080/14616696.2016.1141304.

Robards, J. and Berrington, A. (2016). The fertility of recent migrants to England and Wales. *Demographic Research* 34(36): 1037–1052. doi:10.4054/DemRes.2016.34.36.

Rondinelli, C., Aassve, A., and Billari, F.C. (2006). Socio-economic differences in postponement and recuperation of fertility in Italy: Results from a multi-spell random effect model. Colchester: University of Essex (ISER Working Paper 2006-46). https://www.iser.essex.ac.uk/research/publications/working-papers/iser/2006-46.

Rousselière, D. (2019). A flexible approach to age dependence in organizational mortality: Comparing the life duration for cooperative and non-cooperative enterprises using a Bayesian generalized additive discrete time survival model. *Journal of Quantitative Economics* 17: 829–855. doi:10.1007/s40953-019-00164-0.

Shao, S., Kauermann, G., and Smith, M.S. (2020). Whether, when and which: Modelling advanced seat reservations by airline passengers. *Transportation Research Part A: Policy and Practice* 132: 490–514. doi:10.1016/j.tra.2019.12.002.

Steele, F. (2005). Event history analysis. Southampton: National Centre for Research Methods (National Centre for Research Methods Review Paper NCRM/004). https://eprints.ncrm.ac.uk/id/eprint/4467.

Stone, J. and Berrington, A. (2017). Income, welfare, housing and the transition to higher order births in the UK. Southampton: Centre for Population Change (ESRC Centre for Population Change Working Paper Series 86). https://eprints.soton.ac.uk/414769/.

Timæus, I.M. and Moultrie, T.A. (2008). On postponement and birth intervals. *Population and Development Review* 34(3): 483–510. doi:10.1111/j.1728-4457.2008.00233.x.

Timæus, I.M. and Moultrie, T.A. (2020). Pathways to low fertility: 50 years of limitation, curtailment and postponement of childbearing. *Demography* 57(1): 267–296. doi:10.1007/s13524-019-00848-5.

Torrisi, O. (2020). Armed conflict and the timing of childbearing in Azerbaijan. *Population and Development Review* 46(3): 501–556. doi:10.1111/padr.12359.

UNDP, United Nations Development Programme (2018). Human Development Indices and Indicators: 2018 Statistical Update – Technical Notes. New York: UNDP. https://hdr.undp.org/sites/default/files/data/2020/hdr2018_technical_notes.pdf.

UNDP, United Nations Development Programme. (2019). Human Development Report 2019. New York: UNDP. https://hdr.undp.org/content/human-development-report-2019.

University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2017). *Understanding Society: Waves 1-7, 2009-2016: Special Licence Access* [data collection]. 7th Edition. UK Data Service. SN: 6931. doi:10.5255/UKDA-SN-6931-7.

Van Bavel, J. and Nitsche, N. (2013). 'The proper age for parenthood' and second birth rates in Europe. *European Sociological Review* 29(6): 1149–1161. doi:10.1093/esr/jct003.

Van Hook, J. and Altman, C.E. (2013). Using discrete-time event history fertility models to simulate total fertility rates and other fertility measures. *Population Research and Policy Review* 32(4): 585–610. doi:10.1007/s11113-013-9276-7.

White, M.J., Muhidin, S., Andrzejewski, C., Tagoe, E., Knight, R., and Reed, H. (2008). Urbanization and fertility: An event-history analysis of Coastal Ghana. *Demography* 45(4): 803–816. doi:10.1353/dem.0.0035.

Wood, S.N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC Press. doi:10.1201/9781315370279.

Wood, S.N. (2020). Inference and computation with generalized additive models and their extensions. *TEST* 29: 307–339. doi:10.1007/s11749-020-00711-5.

Wu, L.L. (2003). Event history models for life course analysis. In: Mortimer, J.T. and Shanahan, M.J. (eds.). *Handbook of the life course*. New York: Kluwer Academic/Plenum Publishers: 477–502. doi:10.1007/978-0-306-48247-2_22.

Zang, E. (2019). Women's educational attainment and fertility among Generation X in the United States. *Population Studies* 73(3): 335–351. doi:10.1080/00324728.2019.1658799.

## Appendix A: Sample selection process

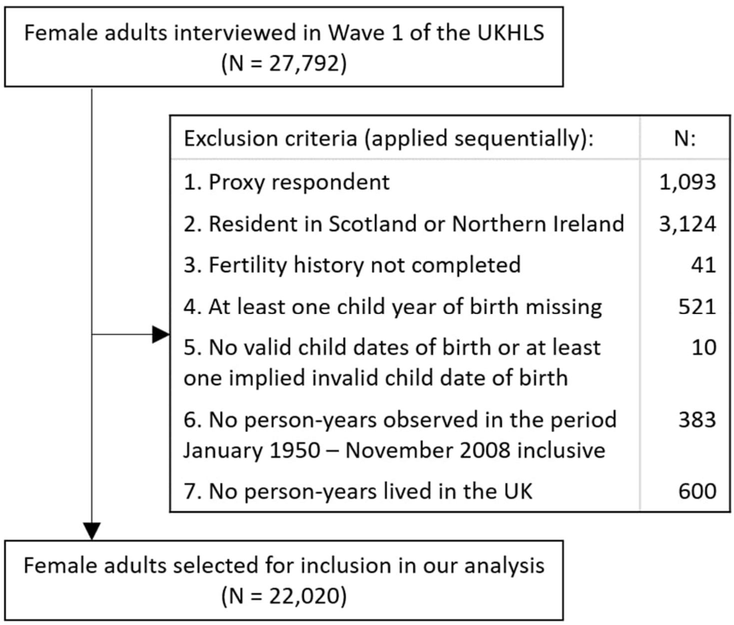
The process of obtaining our sample of 22,020 women from the initial 27,792 interviewed in Wave 1 of the UKHLS is outlined in Figure A-1. We justify the numbered exclusion criteria below.

1. We needed the respondent to report their details and hence only included full respondents.
2. We are keen to extend the modelling presented in this paper to incorporate the parity-specific fertility rates provided by ONS (2020c), which are for England and Wales only. Therefore, in order for our UKHLS sample to be as representative as possible of this population, we exclude women living in Scotland and Northern Ireland.
3. Self-explanatory.
4. As our interest is in the timing of parity progression, we need to be confident that each fertility history is accurate. If a woman reaches a parity above 1 and has a missing child year of birth (CYOB) there is no way of knowing the child's position in the birth order; hence the accuracy of the entire fertility history is severely compromised. Even if the woman only reaches parity 1, in which case the birth order is trivially known, with the CYOB missing there is still great uncertainty about when the single transition actually occurred. Therefore, we believe that the safest course of action is to exclude women with missing CYOBs.
5. We classify a child date of birth (CDOB) as invalid if the age of the mother at the CDOB is less than 12. We treat the invalid CDOBs as missing, instead assuming that the births occurred some time after the mother's 12th birthday but before the first birth with a valid CDOB (if this exists); we then only include the woman in our analysis from the parity reached following this birth. However, if the mother is aged 12 at the first valid CDOB then our assumption is invalidated (the CDOB is implied invalid) and the woman is excluded.
6. The years after 1955 have person-years observed for the whole reproductive age range (15–44); those immediately prior to this cover all but the oldest ages, where fertility rates are low. Therefore, January 1950 is a suitable choice for the start of our observation period. It makes sense for our observation period to end at a fixed point prior to the start of the Wave 1 interviews in 2009 – choosing December 2008 would mean that for the cohorts still being observed, their last fully observed age could only be reached by the January-born women and therefore the number of person-years corresponding to the 2008 period would be small. Moving the endpoint back by a month to November 2008 means that this age cannot be fully observed

for any cohort members, and so the last fully observed age (corresponding to the 2007 period) can be reached by all members as desired.

7. Foreign-born respondents are asked for the year they came to Britain – where this is provided, we remove all person-years up to and including the implied age at moving, to make our sample of person-years more representative of the United Kingdom.

**Figure A-1: Exclusion flowchart to illustrate the selection of the women in our sample**




We note that the 4,107 women satisfying criteria 2, 6, or 7 are outside the scope of our study, while the 1,665 women satisfying criteria 1, 3, 4, or 5 would ideally have been included (as long as they did not also satisfy any of criteria 2, 6, or 7). Whereas the 1,093 proxy respondents in this latter group can reasonably be considered missing at random, the remaining 572 women satisfying criteria 3–5 cannot, as the reasons for their removal are fertility-related.

Also, we note that for the women with missing years and/or months of birth (103 of the 22,020), we computed approximate values using the woman's age at the time of interview and the date of interview. Where only the month of birth was missing but these two pieces of information were not consistent with the year of birth provided, we imputed the month of birth at the midyear; we did the same for any missing child months of birth. We also decided to ignore the presence of birth intervals of length 1–8 months, occurring for 139 of the women in our sample, as we felt that the benefits of their inclusion outweighed the potentially negative impact that slight inaccuracies in their fertility histories could have on our inferences. For any women with repeated CDOBs, we treated these as multiple births and adjusted the parity accordingly.

## Appendix B: Selected alternatives to smooth terms

**Table B-1: Selected categorical representations for the variables age ($A$), calendar year ($Y$), and time since last birth ($T$)**

| Term | Range | Current parity | Categories | References |
|---|---|---|---|---|
| $A$ | 15–44 | 0, 1, 2 | 15–19, 20–24, 25–29, 30–34, 35–39, 40–44 | Jefferies, Berrington, and Diamond (2000), Van Hook and Altman (2013), Stone and Berrington (2017) |
| | | 3+ | 18–24, 25–29, 30–34, 35–39, 40–44 | |
| $Y$ | 1950–2007 | 0, 1, 2, 3+ | 1950–1959, 1960–1969, 1970–1979, 1980–1989, 1990–1999, 2000–2007 | Riise, Dommermuth, and Lyngstad (2016) |
| $T$ | 1–11 | 1, 2, 3+ | 1: 1–2, 3, 4, 5, 6, 7+ | Stone and Berrington (2017) |
| | | | 2: 1–2, 3–6, 7–10, 11+ | Torrisi (2020) |

**Table B-2: Selected polynomial representations for the variables age ($A$), calendar year ($Y$), and time since last birth ($T$)**

| Term | Range | Current parity | Polynomial | References |
|---|---|---|---|---|
| $A$ | 15–44 | 0, 1, 2, 3+ | Quadratic | Erlandsson (2017), Hank (2002), Kim (2014, 2017), Kim and Park (2009), Piotrowski and Tong (2016), White et al. (2008), Zang (2019) |
| | | 2, 3+ | Linear | – |
| $Y$ | 1950–2007 | 0, 1, 2, 3+ | Linear | Kulu (2013), Lewis and Raftery (1999), Muggeo, Attanasio, and Porcu (2009) |
| | | | Quadratic | – |
| $T$ | 1–11 | 1, 2, 3+ | Quadratic | Piotrowski and Tong (2016), Zang (2019) |

**Table B-3: Selected piecewise linear spline representations for the variables age ($A$), calendar year ($Y$), and time since last birth ($T$)**

| Term | Range | Current parity | Knots | References |
|---|---|---|---|---|
| $A$ | 15–44 | 0 | 1: 20, 23, 27, 32, 37 | Kravdal (2007), Kravdal and Rindfuss (2008), Begall and Mills (2013) |
| | | | 2: 20, 25, 30, 35, 40 | Kravdal (2001), Kulu (2013), Kulu and Washbrook (2014) |
| | | 1 | 20, 25, 30, 35 | Kravdal (2007), Kravdal and Rindfuss (2008) |
| | | 2, 3+ | 25, 30, 35 | Kravdal (2007), Kravdal and Rindfuss (2008), Begall and Mills (2013) |
| $Y$ | 1950–2007 | 0, 1, 2, 3+ | 1960, 1970, 1980, 1990, 2000 | - |
| $T$ | 1–11 | 1, 2, 3+ | 2, 4, 6, 8 | Kravdal (2001, 2007), Kravdal and Rindfuss (2008), Begall and Mills (2013) |

## Appendix C: Fitting the chosen GAMs using the `gam` function in `R`

Following on from Section 6.1, here we provide the data structure and code for fitting the chosen GAMs, M0-M3+, in `R`.

M0 ($A + Q_4 + Y + AQ_4 + AY$):

The data frame `dat0` has the following structure:

| age | year | q4 | n | z | r |
|---|---|---|---|---|---|
| 15 | 1950 | 1 | | | |
| 16 | 1950 | 1 | | | |
| | | | | | |
| 44 | 2007 | 4 | | | |

The `age` and `year` columns are self-explanatory; the `q4` column is the qualification variable, taking values 1–4 according to Figure 2 (note that it needs to be encoded as a categorical variable or ‘factor’ before fitting the model). Following on from Section 5.1, `n` gives the weighted number of person-year records sharing the corresponding covariate pattern, while `z` gives the weighted number of these records which have a binary response of 1 and `r` is the quotient of `z` and `n`; i.e., `z/n`. The code to fit M0 is given below:

```
M0 <- gam(r ~ s(age, bs = "ps", k = 9, m = c(2, 1)) + q4 +
       s(year, bs = "ps", k = 9, m = c(2, 1)) +
       s(age, by = q4, bs = "ps", k = 9, m = c(2, 1)) +
       ti(age, year, bs = c("ps", "ps"), k = c(9, 9),
       d = c(1, 1), m = list(c(2, 1), c(2, 1))),
       family = binomial, weights = n, data = dat0,
       method = "REML")
```

We note the following:

- The `s()` terms denote 1D smooth functions, with the ‘`by = q4`’ argument in the third line of the above code telling the model to estimate a smooth function of `age` for each level of `q4`.
- The `ti()` term denotes a 2D smooth function.
- The `bs`, `k`, and `m` arguments within the `s()` and `ti()` terms allow the user to specify the construction of the functions and how their smoothness is determined.
- The `family` argument indicates the distribution being assumed (here, binomial – see Section 5.1).
- The `method` argument indicates the method by which the smoothing parameters are estimated (here, REML – see Section 5.2).

Once the model is fitted, useful functions include:

- `summary(M0)`: gives a summary of the fitted model, including fixed effect estimates and EDF values for the smooth functions.
- `plot(M0)`: plots the smooth terms fitted in the model.
- `predict(M0)`: produces predictions from the fitted model with associated uncertainty, on the probability scale (`type = "response"`), logit probability scale (`type = "link"`), and even the scale of the smooth terms (`type = "terms"`).

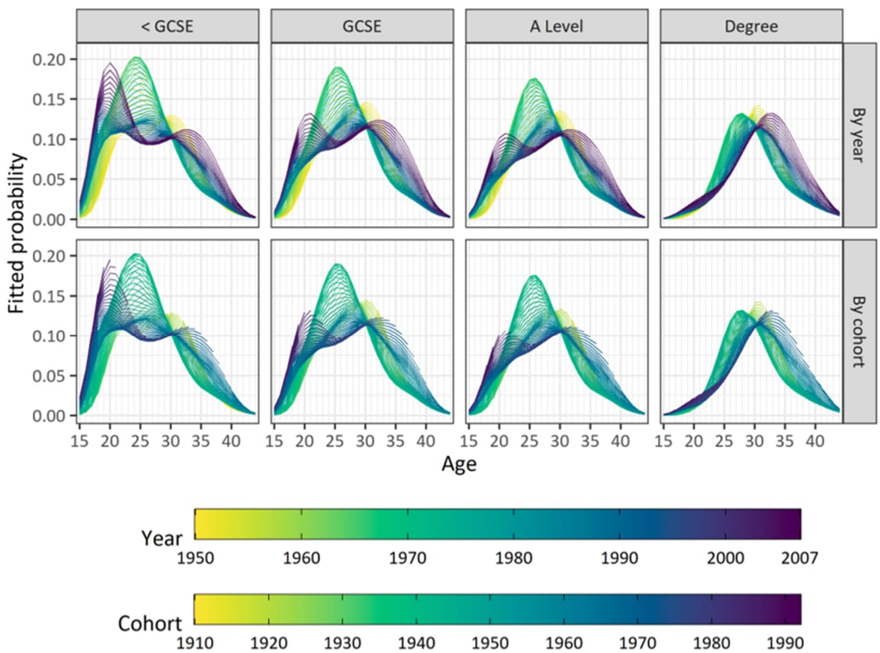
M1-M3+:

The simpler models for the higher-order parity progressions can be fitted similarly, by creating data frames like `dat0` which include columns relating to each of the variables in the chosen model, and then adjusting the `M0` code appropriately.

## Appendix D: Fitted probabilities from the model of progression to first birth (M0)

To facilitate the substantive interpretation of M0, in Figure D-1 we plot the fitted probabilities estimated under M0 by calendar year (row 1) and cohort year of birth (row 2). As described in Section 6.2.1, we note the intensification of the earlier peak in childbearing and subsequent strong bimodality in the recent age curves for the lower qualification categories, and the intensification of the later peak and subsequent unimodal pattern for the 'Degree' category. We also observe that despite M0 not directly accounting for cohort, it clearly captures key features of cohort fertility such as the peak for women born in the mid to late 1930s and the delaying of childbearing to older ages for the more recent cohorts (see ONS 2020d).

**Figure D-1: Fitted probabilities from the model of progression to first birth (M0) by year and cohort**

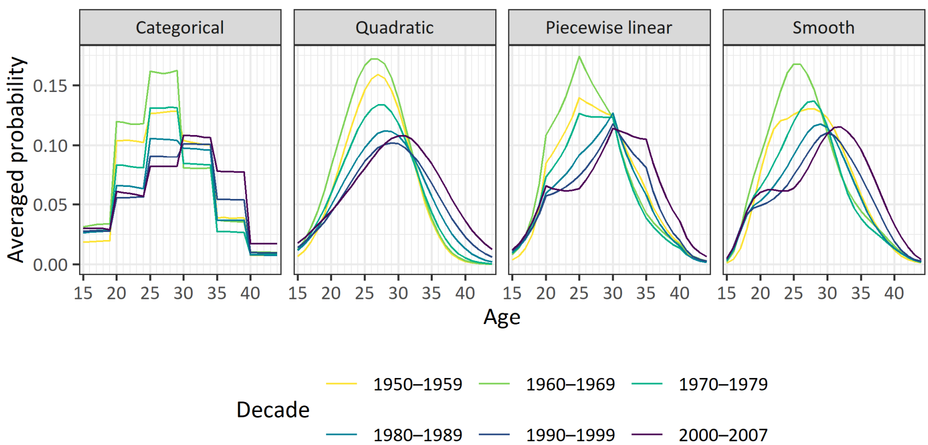



*Note:* Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against age ($A$, in years) for a given calendar year ($Y$, row 1) or cohort year of birth (row 2), and highest educational qualification ($Q$, columns), and are identical in the two rows. Ranges of $AYQ$ combinations match those observed in the data.

## Appendix E: Further simplified graphical comparisons for the model of progression to first birth (M0) by age, year, and qualification level

To simplify fertility changes over age and time, for each age representation and decade we compute the weighted average of the fitted probabilities across qualification, plotting the age curves in Figure E-1. This allows us to compare the success with which the comparator methods capture the recent hump in early fertility (see Section 6.2.1) which has led to the emergence of bimodality in the age curve. It is hard to identify the hump among the step functions in the categorical plot and it is non-existent in the quadratic representation. The piecewise linear and smooth methods perform best, consistent with their ability to estimate the peaks in early and late fertility in Figure 7. However, for the piecewise linear approach the location of the hump is restricted by the change points, resulting in an angular appearance.

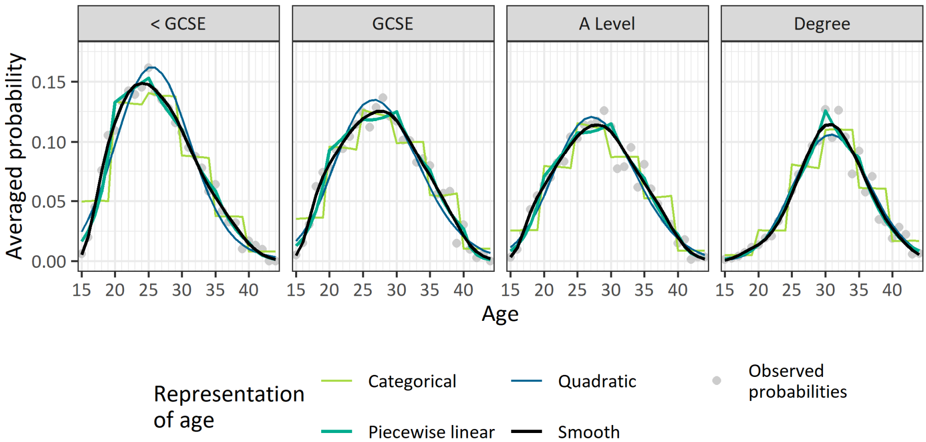
**Figure E-1: Fitted probabilities averaged over highest educational qualification for the model for progression to first birth (M0) comparative analysis**




*Note:* The effect of year is estimated as a smooth term throughout. Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against age (in years) for a given range of calendar years (colour) and representation of age (subplot). Ranges of covariate combinations match those observed in the data.

To simplify fertility changes across age and qualification, we also compute the weighted average across calendar year, overlaying the age curves corresponding to the various methods for each qualification category in Figure E-2, as well as the observed probabilities for comparison. This provides further support for our conclusions, as the smooth curves approximate the observed patterns most closely. Inappropriate knot locations clearly force the piecewise linear effect to deviate from the smooth curve near these points, while the quadratic and categorical effects can only achieve increasingly poorer approximations.

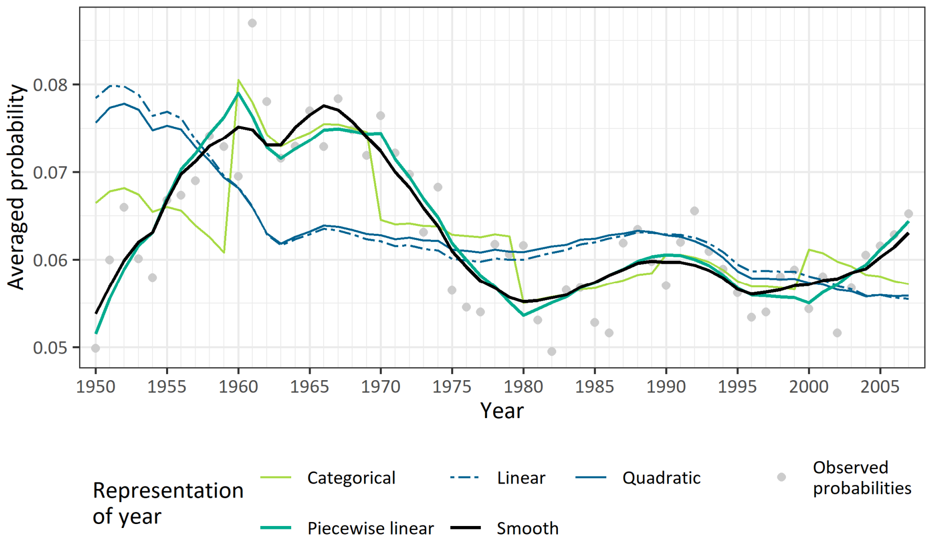
**Figure E-2: Fitted probabilities averaged over calendar year for the model for progression to first birth (M0) comparative analysis**




*Note:* The effect of year is estimated as a smooth term throughout. Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against age (in years) for a given highest educational qualification (subplot) and representation of age (colour). Observed probabilities are overlaid. Ranges of covariate combinations match those observed in the data.

Lastly, to simplify fertility changes over time, in Figure E-3 we plot the year curves computed from taking a weighted average of the fitted probabilities across age and qualification for each year representation. The period features identified in Section 6.2, namely the 1960s baby boom, the second, smaller peak in the 1980s and 1990s, and the more recent rise in the 2000s are all captured by the smooth and piecewise linear approaches. The categorical method performs well at times, but fails when there are sustained fertility increases or decreases within a category such as the 1950s, 1970s, and 2000s. Finally, we note the striking deficiencies of using simple polynomials to estimate the year effect.

**Figure E-3: Fitted probabilities averaged over age and highest educational qualification for the model for progression to first birth (M0) comparative analysis**




*Note:* The effect of age is estimated as a smooth term throughout. Fitted probabilities are conditional probabilities of a birth event under M0 given the included covariates and being in parity 0; they are plotted against calendar year for a given representation of year. Observed probabilities are overlaid.