# Can incorporating parity information improve the reliability of fertility projections? Insights from a Bayesian generalized additive model approach

Joanne Ellison[1], Jakub Bijak[2], Erengul Dodd[3]

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

**Abstract**

Fertility projections inform population projections and are used to plan for the future provision of vital services such as maternity care and schooling. Existing fertility forecasting models tend to use aggregate births data indexed by age and time alone, thereby neglecting to include information about parity, i.e. the number of previous live-born children. This omission risks ignoring a crucial mechanism of fertility dynamics and consequently producing biased predictions. We propose a Bayesian parity-specific fertility projection model within a generalized additive model (GAM) framework. The use of GAMs enables a smooth age-cohort rate surface to be estimated for each parity simultaneously. We constrain our model using aggregate data and additionally introduce random walk priors on completed family size and parity progression ratios, which are summary fertility measures known to change relatively slowly over time. Using Hamiltonian Monte Carlo methods, we fit our model to Canada, Czechia, England & Wales and the Netherlands. We compare our forecasts with the best-performing existing models to quantify the impact of including the parity dimension on predictive accuracy. Our findings indicate that a parity-specific approach could lead to more plausible and reliable fertility projections, aiding government planners in their decision-making and enabling more tailored policy solutions.

*Keywords*: Bayesian methods; Cohort fertility; Forecasting; Generalized additive models; Parity progression

# 1 Introduction

Fertility projections are used in population projections and to plan for future service provision, such as maternity care and schooling. Methodological literature on fertility forecasting is extensive, with comprehensive reviews in Booth (2006) and Bohk-Ewald et al. (2018). Existing models tend to use aggregate data on the total number of births indexed only by age, and period (calendar year) or birth cohort. These variables constitute three of the "five clocks" of fertility (Raftery et al. 1996:1), the other two being time since last birth and number of previous live-born children (parity). While little information is typically collected about time since last birth at the population level, disaggregation of births by birth order is increasingly common (Jasilioniene et al. 2015). As childbearing is a sequential process, the current number of children is likely to be an important determinant of the decision to have an additional child. Therefore, omitting birth order information risks ignoring a crucial mechanism of fertility dynamics and producing biased predictions. A parity-specific approach can thus translate into more reliable projections and improved decision-making in a range of policy and planning domains.

While the consideration of parity is underexplored in fertility projections (Petropoulos et al. 2022:797–798), it is commonplace in the fertility analysis literature. A key reason for this is the ability to gain a more detailed understanding of the changes in childbearing behaviour underlying past trends in fertility levels. The approaches taken include: computing period parity progression ratios (Feeney and Yu 1987; McDonald et al. 2015; Murphy and Berrington 1993; Ní Bhrolcháin 1987); proposing parity-adjusted summary fertility measures such as the total fertility rate (Bongaarts and Sobotka 2012); decomposing changes in parity progression ratios and total fertility rates by parity (Andreev et al. 2002; Hellstrand et al. 2020, 2021) and changes in the cohort total fertility rate by parity progression ratios (Zeman et al. 2018); and calculating standardized rates by birth order (Andersson 1999; Rallu and Toulemon 1994). Event history analyses of births also typically account for parity, by fitting separate parity-specific models (e.g. Ellison et al. 2022) or joint models that control for parity and unobserved heterogeneity between women (e.g. Begall and Mills 2013; Kravdal 2001; Van Hook and Altman 2013).

Analysing childbearing patterns by parity can also better inform future fertility assumptions (Smallwood 2002), which are needed when producing projections. Indeed, Kohler and Ortega (2002) argue that appropriate cohort fertility projections require plausible assumptions for the quantum (level) and tempo (timing) of fertility, which in turn necessitate parity information to account for those exposed to each birth order more precisely; concurrently, this captures the aforementioned mechanistic nature of childbearing, whereby decisions are made and realized sequentially. This additional differentiation by parity increases the homogeneity of the resulting groups while also incorporating women's previous childbearing experience (Rallu and Toulemon 1994). Hobcraft (1996) provides a historical example, stating that parity-specific analyses would likely have helped to anticipate the baby boom and bust of the 1960s and 1970s in England & Wales, which were not predicted by the corresponding forecasts. Regarding current use, a recent European survey of 32 national statistical agencies found that while eight use parity-specific data to inform their fertility projections, only two actually produce parity-specific projections (Gleditsch et al. 2021).

In the fertility projections literature, the production of parity-specific forecasts has been suggested, for example by Schmertmann et al. (2014) and Shang and Booth (2020),

with data limitations preventing the latter from implementing the approach. Ellison et al. (2023) proposes a Bayesian parity-specific projection model combining survey and vital registration data. Aggregate forecasts are generated but each parity is modelled independently – this means that measures depending on multiple parities, such as age-specific fertility rates (ASFRs) or summary measures, cannot be constrained. The authors recommend the subsequent development of a generalizable approach that models the parities jointly and incorporates interpretable constraints on summary measures. A Bayesian setup enables coherent integration of multiple data sources and prior knowledge, and appropriate quantification of predictive uncertainty; in the field of demographic forecasting, Bayesian methods are becoming increasingly prevalent (Mazzuco and Keilman 2020).

This discussion motivates the aims of this paper. Firstly, we aim to develop a fertility forecasting method that incorporates parity information, advancing the existing literature in line with the suggestions of Ellison et al. (2023). Secondly, we aim to assess whether the proposed method has increased forecast reliability. Within a Bayesian generalized additive model (GAM) framework and informed by parity-specific rate estimates, we construct smooth parity-specific age-cohort rate surfaces. We additionally incorporate constraints on ASFRs and summary fertility measures at the parity-specific and aggregate levels, using vital registration data and priors informed by pooling data from the Human Fertility Database (Human Fertility Database 2023). We obtain forecasts for Canada, Czechia, England & Wales and the Netherlands, comparing their performance with some of the best-performing existing models and an aggregate version of our proposed model.

The paper proceeds as follows. After describing our data sources in Section 2, we specify our modelling approach in Section 3. In Section 4 we present the model outputs, namely the parity-specific rates and summary measures, and summarize the comparative findings. We then provide a discussion in Section 5.

## 2 Data

### 2.1 Setup

In this paper we address the fertility projection problem of completing cohort fertility: forecasting the average number of children that women from a particular birth cohort will have had by the end of their reproductive lives (cohort total fertility rate, CFR). Contrastingly, period fertility considers the average number of children a pseudo-cohort of women would have if they were subject to the fertility rates of a given period at every age (total fertility rate, TFR). The pros and cons of these approaches are well known (Ní Bhrolcháin 1992); however, we prefer the cohort approach for several reasons. Firstly, where period data on parity distributions are absent or lacking in quality and/or frequency, a cohort perspective allows their estimation by cumulating across each reproductive age (Section 2.2 provides further details). Secondly, the reduced fluctuation of cohort fertility is beneficial for forecasting, as it justifies more strongly the assumption of the future smooth progression of summary measures.

Next we introduce the modelling setup. Taking the reproductive age range to be from $a_{min} = 15$ to $a_{max} = 44$ years, we let $JOY$ be a particular 'jump-off' year whereby its associated data are the most recent data used in model fitting. We consider four $JOY$

values – 2000, 2005, 2010 and 2015 – to test the forecast performance of our model over different periods and forecast horizons. In a similar setup to Schmertmann et al. (2014), for a given $JOY$ we consider $n_c = a_{max} - a_{min} + 11 = 40$ cohort years of birth, ranging from $c_{min}^{JOY} = (JOY - a_{max} - 10)$ to $c_{max}^{JOY} = (JOY - a_{min})$. For example, $c_{min}^{2000} = 1946$ and $c_{max}^{2000} = 1985$, while $c_{min}^{2015} = 1961$ and $c_{max}^{2015} = 2000$. The first 11 cohorts (the 1961–1971 cohorts for $JOY = 2015$) have reached at least age $a_{max}$ by year $JOY$ and therefore have already completed their reproductive lives; the subsequent $(a_{max} - a_{min})$ cohorts are only partially observed up to age $(a_{max} - 1), \ldots, a_{min}$. We illustrate this setup graphically in panel a of Figure 1, for the England & Wales ASFRs when $JOY = 2015$. The rates are displayed in the form of a Lexis surface (left plot) and cohort-specific rate curves (right plot).

The data requirements of the model are population-level birth counts (by age, cohort and birth order), and exposures (by age and cohort). Births and exposure data are typically reported annually on a period basis (note that as is standard, we approximate the exposures using mid-year population estimates). Sophisticated cohort conversion methods exist and are used for example by the Human Fertility Database (HFD) (Jasilioniene et al. 2015). However, for convenience and ease of model implementation, we simply assign data collected about women aged $a$ last birthday in year $JOY$ to cohort $(JOY - a)$. This means that the cohort assignment is only approximate; however, the resulting ASFR and summary measure estimates correspond closely with the HFD values (Human Fertility Database 2023). Bohk-Ewald et al. (2018) also take this simplified approach.

We also consider four countries: England & Wales (data from Office for National Statistics 2022a); and Canada, Czechia and the Netherlands (births data from Human Fertility Database 2023; exposure data from Human Mortality Database 2023)[1]. The country-specific data need to cover the period from 1961 (when the $c_{min}^{2000} = 1946$ cohort is aged 15) to 2015 (most recent $JOY$ value). This excludes many HFD countries, mainly due to the order-specific births data which often start to be collected much more recently than the aggregate data. Of the nine HFD countries available for inclusion, we remove several due to data quality issues. Our chosen countries exhibit reasonable geographical spread, and differing population sizes and social, cultural and historical contexts. This leads to a diverse set of childbearing patterns upon which we can suitably test the forecast performance of our model.

## 2.2 Age-parity- and age-specific fertility rates

The primary data source that informs our fertility projection model is conditional parity-specific fertility rate estimates by age and cohort. Their conditionality means that they are of occurrence/exposure form and so appropriately account for the population at risk. For women of a given age $a$ and cohort $c$, we define the conditional fertility rate for birth order $i$, $l_{ac}^i$, in the following way:

$$l_{ac}^i = \frac{\text{births of order } i \text{ to women aged } a \text{ in cohort } c}{\text{person-years lived at parity } i-1 \text{ by women aged } a \text{ in cohort } c} = \frac{b_{ac}^i}{E_{ac}^{i-1}}. \qquad (1)$$

[1] We also require exposure estimates corresponding to future age-cohort combinations: for England & Wales we use the 2020-based National Population Projections (Office for National Statistics 2022b), while for the HFD countries we use the 2022 Revision of *World Population Prospects* (United Nations, Department of Economic and Social Affairs, Population Division 2022).

We assume that parities $J$ and above, and birth orders $J+1$ and above, are combined. Therefore parity takes values $0, 1, \ldots, J+$ and birth order takes values $1, 2, \ldots, (J+1)+$. We take $J = 3$ and therefore consider parities 0–3+, due to small cell counts and similar age patterns for the conditional rates of the highest parities.

Data for the numerator comes from vital registration. Information on true (i.e. biological) birth order is now collected by most countries, with many having only requested parity information for marital births until relatively recently (Jasilioniene et al. 2015). In England & Wales this change happened in 2012, with population-level estimates of births by true birth order up to 2012 computed using the General Household Survey (Office for National Statistics 2023). Post-2012, birth registration data alone is used, however misreporting concerns have been identified (Office for National Statistics 2018); we adjust the post-2012 birth counts to correct for this (Appendix A provides further details). For the remaining countries we source the order-specific births data from the HFD, since the time series are available from 1950 or earlier for each country (pre-2012 estimates for England & Wales are not provided).

We estimate the denominators by cumulating cohort fertility, following the method of Smallwood (2002); the HFD produces its fertility tables using a similar approach (Jasilioniene et al. 2015). The method first assumes that all women in a given cohort are childless at age $a_{min}$. Then, at each subsequent age, the relevant order-specific birth counts are used to update the number of women exposed to each parity. These are then scaled to sum to the corresponding mid-year population estimate, giving the final exposure estimates. We specify the method formally below.

For a given cohort $c$, our goal is to estimate parity-specific exposures $E^i_{ac}$ for $i \in \{0, 1, \ldots, J+\}$ and $a \in \{a_{min}, \ldots, a_{max}\}$. We define temporary exposures $\tilde{E}^i_{ac}$, mid-year estimates $E_{ac}$, and omit the cohort subscripts for clarity. We perform the following steps as in Smallwood (2002):

1. Assume all women at age $a_{min}$ are childless:

$$E^0_{a_{min}} = E_{a_{min}}, E^i_{a_{min}} = 0 \text{ for } i \in \{1, 2, \ldots, J+\}.$$

2. For $a = a_{min} + 1$, obtain temporary exposures $\tilde{E}^i_a$ by updating each $E^i_{a-1}$ according to the following equations:

$$\begin{aligned}
\tilde{E}^0_a &= E^0_{a-1} - b^1_{a-1}; \\
\tilde{E}^i_a &= E^i_{a-1} - b^{i+1}_{a-1} + b^i_{a-1} \text{ for } i \in \{1, 2, \ldots, (J-1)\}; \\
\tilde{E}^{J+}_a &= E^{J+}_{a-1} + b^J_{a-1}.
\end{aligned}$$

   Then, scale each $\tilde{E}^i_a$ so that $\sum_i \tilde{E}^i_a = E_a$, and define this as $E^i_a$:

$$E^i_a = E_a \times \frac{\tilde{E}^i_a}{\sum_i \tilde{E}^i_a} \text{ for } i \in \{0, 1, \ldots, J+\}.$$

3. Repeat step 2 for $a \in \{(a_{min} + 2), \ldots, a_{max}\}$.

A drawback to this approach is that it ignores any potential sensitivity of parity to mortality and migration, which would ideally be accounted for if data were available (Smallwood

2002). We also note that although the HFD provides census- or register-based estimates of the female age-parity distribution for Canada, Czechia and the Netherlands, we do not make use of this for several reasons. Firstly, they are only available for selected years (Canada: 1991; Czechia: 1950–2011, roughly decennial; Netherlands: 1995–2009, annual), the infrequency making it difficult to incorporate the data into the cohort cumulation process. The HFD achieves this by using a single census or register estimate for the starting exposures of its period fertility tables, when it is possible and appropriate (Jasilioniene et al. 2015); however, this limits the range of years for inclusion. Secondly, the absence of such estimates for England & Wales, and the differing data availability for the other countries, means that any proposed approach would be inconsistently applied across countries. Thirdly, it will be necessary to replicate the chosen estimation method within our projection model, further elevating the importance of a simple and consistent methodology.

Next, for women of a given age $a$ and cohort $c$, we define the unconditional fertility rate for birth order $i$, $m^i_{ac}$, as:

$$m^i_{ac} = \frac{\text{births of order } i \text{ to women aged } a \text{ in cohort } c}{\text{person-years lived by women aged } a \text{ in cohort } c} = \frac{b^i_{ac}}{E_{ac}}. \tag{2}$$

The only difference from the conditional rate in equation (1) is the change in denominator from women at risk of the particular birth order to those at risk of any birth.

Lastly, for women of a given age $a$ and cohort $c$, we define the ASFR or age-specific fertility rate, $m_{ac}$, as:

$$m_{ac} = \frac{\text{births to women aged } a \text{ in cohort } c}{\text{person-years lived by women aged } a \text{ in cohort } c} = \frac{\sum_i b^i_{ac}}{E_{ac}} = \frac{b_{ac}}{E_{ac}}.$$

We note that $m_{ac} = \sum_i m^i_{ac}$, i.e. for fixed $a$ and $c$, the sum of the unconditional rates gives the ASFR.

## 2.3 Summary fertility measures

We describe the computation of the summary fertility measures required for our proposed model, namely the CFR (see Section 2.1) and parity progression ratios (PPRs). These are defined for women in cohort $c$ after age $a_{max}$. We define the CFR, $CFR_c$, representing the average completed family size, as $CFR_c = \sum_{a_{min}}^{a_{max}} m_{ac}$. We define the CFR for birth order $i$, $CFR^i_c$, representing the average number of births of order $i$ per woman, as $CFR^i_c = \sum_{a_{min}}^{a_{max}} m^i_{ac}$. We note that $CFR_c = \sum_i CFR^i_c$. Then we define the PPR from parity $i$ to $i+1$, $PPR^{i\to i+1}_c$, representing the proportion of women progressing from parity $i$ to $i+1$, as:

$$\begin{aligned}
PPR^{0\to 1}_c &= CFR^1_c; \\
PPR^{i\to i+1}_c &= \frac{CFR^{i+1}_c}{CFR^i_c} \text{ for } i \in \{1,2,\ldots,(J-1)\}; \\
PPR^{J+\to(J+1)+}_c &= \frac{CFR^{(J+1)+}_c}{CFR^J_c + CFR^{(J+1)+}_c}.
\end{aligned}$$

For a given jump-off year $JOY$, using estimates of the $m^i_{ac}$ values from the HFD, we calculate these summary measures for all country-cohort combinations from the 1945 cohort onwards who have completed their reproductive lives by year $JOY$. We use these pooled datasets to inform constraints that we incorporate into our proposed model (Section 3.3 provides further details). Having described the model setup, the required data sources and fertility measures, in Section 3 we specify our parity-specific projection model.

# 3 Methods

## 3.1 Model structure

We illustrate the model structure in Figure 2, in the same spirit as Figure 3 of Zhang and Bryant (2019). We distinguish between three levels – L1, L2 and L3 – in the columns. Each level has a different demographic array of interest (white rectangles):

- L1: true conditional birth rates by age, cohort and birth order, i.e. the true $l^i_{ac}$ values (equation (1) in Section 2.2);
- L2: true unconditional birth rates by age, cohort and birth order, i.e. the true $m^i_{ac}$ values (equation (2) in Section 2.2);
- L3: true summary fertility measures, namely the true CFR and PPRs (Section 2.3).

The observed quantities (gray rectangles) refer to the data requirements of the model (Sections 2.1–2.3):

- birth counts by age, cohort and birth order (from vital registration; L1), which then sum to give birth counts by age and cohort (L2);
- exposures by age and cohort (from mid-year population estimates and projections; L2), and exposures by age, cohort and parity (from high-quality direct estimates if available, and otherwise approximated using the approach of Smallwood 2002 described in Section 2.2; L1);
- pooled cross-country dataset of CFR and PPR estimates from the HFD (L3).

The two prior models (rounded rectangles) allow us to incorporate assumptions regarding the progression of the demographic arrays of interest across age and/or time. The first (L1) assumes underlying smoothness of the true conditional order-specific birth rates across age and cohort for each birth order. The second (L3) assumes slow changes of the CFR and PPRs over time informed by the HFD data. The two data models (hexagons) describe the processes through which the birth count data are generated from the true rates. These govern the relationships between the births by age, cohort and birth order and the true conditional birth rates (L1), and the births by age and cohort and the true unconditional rates (L2). We compute the true unconditional rates from the true conditional rates by again applying the Smallwood (2002) method to generate order-specific birth counts according to the true conditional rates, which we then divide by the age-cohort exposures which are assumed known. The summary measures then follow straightforwardly from the formulas in Section 2.3. The remainder of this section motivates and specifies the prior and data models corresponding to Levels 1–2 and 3 in Sections 3.2 and 3.3 respectively.

## 3.2 Modelling rates by age, cohort and parity (Levels 1 and 2)

### 3.2.1 Prior model

First we specify the L1 prior model for the true conditional birth rates by age, cohort and birth order. We will interchangeably refer to these as parity-specific rates for current parities 0, 1, 2 and 3+, and order-specific rates for progression to first, second, third, or fourth and higher-order birth. Returning to the age-cohort ASFR surface (panel a of Figure 1), we now consider four such surfaces of the conditional parity-specific rates, illustrated in panel b. Analogous to panel a, we present the Lexis surfaces and the cohort-specific rate curves on the top and bottom rows respectively. While the first birth (parity 0) rates progress smoothly across age and cohort and display similar patterns to the ASFRs, the higher-order rates are more erratic, particularly at the youngest ages where few women are exposed. Our approach smooths these rates across age and cohort for each parity on the logarithmic scale, within a generalized additive model (GAM) framework. GAMs enable the individual and joint effects of covariates to be represented as smooth functions; see Wood (2017) for an in-depth introduction.

In the demographic forecasting literature, GAMs have been applied in the context of mortality. Hilton et al. (2019, 2021) develop Bayesian approaches estimating smooth functions of age, age-specific improvement factors, and cohort. Aburto et al. (2021) fit a GAM including smooth effects of age, week and their interaction to forecast weekly deaths, finding it to outperform existing methods. Currie et al. (2004) perform two-dimensional (2D) smoothing of mortality rates across age and period under a P-spline approach (Eilers and Marx 1996). Camarda (2019) extends such an approach to incorporate shape constraints and thus improve forecast plausibility. Concerning fertility, the Bayesian forecasting model of Ellison et al. (2023) (see Section 1) is also based on a GAM framework. The model estimates smooth effects of age, cohort and their interaction, as well as time since last birth.

We now specify our prior model. Letting $\lambda^i_{ac}$ be the true conditional rate for age $a$, cohort $c$ and birth order $i$, our prior model sets:

$$\log\left(\lambda^i_{ac}\right) = f_i(a,c),$$

where $f_i$ is a smooth function of age and cohort estimated using the observed data. Similarly to Currie et al. (2004) and Camarda (2019), we take a P-spline approach. In the simple one-dimensional (1D) case this considers a cubic B-spline basis, consisting of a series of smooth curves, or basis functions, that cover the covariate range and are non-zero for different parts of it. Following the advice of Currie et al. (2004) we space the basis functions regularly at five-year intervals, requiring eight and ten in the age and cohort dimensions respectively. The smooth curve is then constructed as a linear combination of the basis functions, where the coefficients are constrained to change smoothly across the covariate. Appendix B provides a visualization of this 1D case.

Extending this to smooth the log-rates across age *and* cohort, the basis functions are now smooth 2D surfaces constructed by multiplying the curves from the age and cohort dimensions together; Figure 2 of Camarda (2019) provides a depiction of this. Letting there be $P = 8$ and $Q = 10$ 1D basis functions covering the age and cohort ranges respectively, for a given age $a$ and cohort $c$ we represent the values of these functions by $b^1_A(a), \ldots, b^P_A(a)$ and $b^1_C(c), \ldots, b^Q_C(c)$. Then, for a given $p \in \{1, \ldots, P\}$ and $q \in \{1, \ldots, Q\}$ we represent the value of the corresponding 2D basis function as $b^{pq}_{AC}(a,c)$, where $b^{pq}_{AC}(a,c) = b^p_A(a) \times b^q_C(c)$. We then

express $f_i(a,c)$ as the weighted sum of these basis functions with parity-specific weights $\beta_i^{pq}$:

$$f_i(a,c) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \beta_i^{pq} b_{AC}^{pq}(a,c).$$

In our case we therefore have $P \times Q = 8 \times 10 = 80$ basis functions covering the full age-cohort surface. To constrain the $\beta_i^{pq}$ values to change smoothly across age and cohort, we penalize their first-order differences. We illustrate this in Figure 3, an $8 \times 10$ grid of the basis function coefficients where the cell in row $p$ and column $q$ corresponds to $\beta^{pq}$ (omitting the $i$ subscript for clarity). First-order differences can be penalized between adjacent coefficients horizontally or vertically in the cohort or age directions respectively (represented respectively by vertical or horizontal lines between cells).

We color code Figure 3 to indicate areas with different types of penalization. We constrain the coefficients in the first and last rows (rows 1 and $P$) to equal row-specific values $\beta^1$ and $\beta^P$ respectively (gray areas):

$$\beta^{11} = \cdots = \beta^{1Q} = \beta^1; \tag{3}$$

$$\beta^{P1} = \cdots = \beta^{PQ} = \beta^P. \tag{4}$$

This allows us to achieve forecast plausibility in two ways: firstly, by borrowing strength across the youngest ages where there is large rate uncertainty, particularly for higher parities (panel b of Figure 1); and secondly, by ensuring negligible rates at the oldest ages.

For the remaining rows we penalize the first-order differences in the cohort direction (white and hatched areas) and the age direction (white area only). We implement these penalties through normal priors for the differences, specified in equations (5) and (6) respectively:

$$\beta^{pq} - \beta^{p(q-1)} \sim N\left(0, \sigma_C^2\right), p \in \{2,\ldots,P-1\}, q \in \{2,\ldots,Q\}; \tag{5}$$

$$\beta^{pq} - \beta^{(p-1)q} \sim N\left(0, \sigma_A^2\right), p \in \{3,\ldots,P-2\}, q \in \{1,\ldots,Q\}. \tag{6}$$

Standard deviation parameters $\sigma_C$ and $\sigma_A$ govern the degree of smoothing in the cohort and age directions respectively. We give $\sigma_C/0.01$ and $\sigma_A/0.01$ Student's t priors with 7 degrees of freedom; these informative priors express our belief that neighbouring coefficients should have similar values. When fitting the model, the priors in (5) and (6) combine to achieve smoothness across age and cohort simultaneously.

To respect the biological constraints of fertility we do not penalize the differences where the white and hatched areas meet, because experiments found that doing so allowed fertility to peak at implausibly old ages. Furthermore, we do not penalize the differences where the gray and white/hatched areas meet, because the coefficients in rows 1 and $P$ take much smaller values than those in the interior rows.

### 3.2.2 Data models

Next we specify the data models for Levels 1 and 2 (Section 3.1). For birth order $i$, the L1 data model considers the births by age and cohort, $b_{ac}^i$, to follow a negative binomial distribution with mean equal to the relevant parity-specific exposure, $E_{ac}^{i-1}$, multiplied by the true conditional birth rate, $\lambda_{ac}^i$. The overdispersed Poisson interpretation of the

negative binomial distribution enables extra variability to be accounted for, here through the dispersion parameter $\phi^i$:

$$b^i_{ac} \sim \text{NegBinomial}\left(E^{i-1}_{ac}\lambda^i_{ac}, \phi^i\right).$$

We specify a half-normal $N^+(0,5^2)$ prior for each $\phi^i$ to indicate our uncertainty about the parity-specific rate estimates resulting from the denominator approximation (Section 2.2).

Analogously, the L2 data model considers the total births by age and cohort, $b_{ac}$, to follow a negative binomial distribution with mean equal to the exposure by age and cohort, $E_{ac}$, multiplied by the true ASFR, defined as $\mu_{ac}$:

$$b_{ac} \sim \text{NegBinomial}\left(E_{ac}\mu_{ac}, \phi^0\right).$$

We specify a uniform $U(0,\infty)$ prior for the dispersion parameter $\phi^0$ because unlike the parity-specific rate estimates, we can be very confident about the ASFR estimates as they only require aggregate data.

### 3.3 Modelling summary fertility measures (Level 3)

In this section we specify the L3 prior model for the summary measures, namely the true CFR and PPRs. Continuing from Section 2.3, we first consider our pooled cross-country dataset of observable CFR and PPR estimates from Human Fertility Database (2023) for a given $JOY$ value. This means that we have estimates from the 1945 cohort to the $(JOY - a_{max})$ cohort, who complete their childbearing in year $JOY$. Appendix C provides plots of these estimates for $JOY = 2015$. Although the country-specific curves fluctuate substantially across the cohort range, generally consecutive changes are small. We therefore base our prior model on the assumption that the true CFR and PPRs change relatively slowly over time. To inform this assumption, for each summary measure $SM$, i.e. the CFR and different PPRs, we calculate the standard deviation of the consecutive differences from all of the country-specific curves, denoted by $\sigma_{SM}$. For $JOY = 2015$ these take the following values (to 3 significant figures):

$$\begin{aligned} \sigma_{CFR} &= 0.0335; \\ \sigma_{PPR0->1} &= 0.0151;\ \sigma_{PPR1->2} = 0.00939; \\ \sigma_{PPR2->3} &= 0.00722;\ \sigma_{PPR3+->4+} = 0.00683. \end{aligned}$$

To compare the standard deviations we scale them by the average value of their corresponding summary measure. We then find that actually $PPR^{2->3}$ and $PPR^{3+->4+}$ exhibit the greatest variability (with scaled standard deviations of 0.0216 and 0.0215 respectively), followed by the CFR (0.0174), $PPR^{0->1}$ (0.0169) and then $PPR^{1->2}$ (0.0123).

We then express the prior model for true summary measure $SM$, denoted by $\theta^{SM}_c$ for cohort $c$, as a normal prior for a first-order random walk:

$$\theta^{SM}_c \sim N\left(\theta^{SM}_{c-1}, \sigma^2_{SM}\right), c = \left(c^{JOY}_{min}+1\right), \ldots, c^{JOY}_{max}.$$

In this way the prior model indicates that values of the true summary measure for consecutive cohorts should be close to each other, with the degree of closeness informed by the typical variation observed in the historical HFD data. For the HFD countries, the

construction of this prior from observed data that already enters the model has links to Empirical Bayes methods. However, the prior will act mainly to achieve plausibility in the projected summary measures for partially observed cohorts, with any effects for fully observed cohorts likely to be negligible.

## 3.4 Model fitting

For a given country and $JOY$ value, we fit our Bayesian parity-specific fertility projection model using the package `rstan` (Stan Development Team 2023) within `R` (R Core Team 2023). This provides an interface to the `Stan` software, which enables efficient posterior sampling from complex models. We obtain 750 samples and discard the first 250 as burn-in, leaving 500 samples. Assessment of standard convergence diagnostics indicate that the samples mix well. We also fit an aggregate version that simply smooths the age-cohort ASFR surface (panel a of Figure 1) and only constrains the CFR. Both the parity-specific and aggregate models fit quickly, taking on average 3 minutes and less than 30 seconds respectively on a 2.7GHz Intel Core i7 Windows machine. For each posterior sample, we then generate order-specific and aggregate birth counts according to the data models in Section 3.2.2, from which we obtain posterior distributions of quantities of interest that we present in Section 4.

# 4 Results

## 4.1 Parity-specific and aggregate models

In this subsection we present results from the proposed aggregate and parity-specific models fitted to the data from our chosen countries. We focus on $JOY = 2015$ for simplicity, with the other $JOY$ values considered in Section 4.2 when we compare the predictive performance of our proposed models with that of existing methods from the literature.

### 4.1.1 Fitted rates

Figure 4 presents the fitted rates for England & Wales (rows 1 and 2) and Czechia (rows 3 and 4). Rows 1 and 3 display the posterior median ASFRs from the two models (columns 1 and 2) followed by the parity-specific rates, which are only estimable by the parity-specific model. Rows 2 and 4 plot the corresponding observed rates for comparison. The dashed lines indicate rate forecasts (row 1/3), and rate estimates observed after 2015 (row 2/4).

For England & Wales, the first two columns show reasonably similar ASFR forecasts for the aggregate and parity-specific models, with continued declines and recuperation at younger and older ages respectively (albeit with these trends slightly more extreme for the aggregate model). However, the parity-specific model allows us to learn about the parity dynamics underlying these forecasts. In the subsequent columns we see that increasing smoothing is required due to the observed rates being more erratic. The ASFR declines at younger ages appear to be driven predominantly by decreases in the rate of progression to motherhood (parity 0 rates), while all parities contribute to the higher rates at older ages. We also note that the constraint applied for plausibility at the oldest ages (equation (4) in Section 3.2.1) ensures that the level of these increases diminishes at the upper end of the age range.

Czechia presents a starkly different forecasting challenge, with the ASFR collapse acoss ages

15–25 that occurred for the older half of the cohorts following the end of the communist regime (Billingsley 2010). This was counteracted by a steep rise in the ASFRs at later ages which both models predict to continue for the younger cohorts. Interestingly, both models also anticipate a decrease in the peak childbearing age for the youngest cohorts from around age 30 to age 28. The breakdown by parity determines the main drivers of these changes to be at parities 0 and 1, with the projections for the higher parities remaining reasonably close to the recent observed values.

Appendix D provides the corresponding figures for Canada and the Netherlands. The forecasts for Canada are generally similar to those for England & Wales, their main feature being increases at older ages; however, these increases are considerably steeper for parity 0 compared to England & Wales. The Netherlands differs notably from the other countries in that the variability of the observed ASFRs at each age is very low, while the variability of the consecutive ASFR differences is relatively high for ages 25–35; furthermore, the cohort schedules are consistently highly-peaked and so have fast-changing ASFRs across age. The parity-specific forecasts are remarkably stable, with the greatest change observed for parity 0 where we again see a postponement of childbearing but with the level of the peak decreasing slightly – this contrasts with England & Wales and Canada.

We also note that the Netherlands is the only country with a large difference between the ASFR forecasts under the aggregate and parity-specific models, the former predicting a more extreme shift of the cohort schedules to the right. The greater stability in the latter is due to the PPR constraints, which adjust the parity-specific forecasts to be more realistic while maintaining the CFR projection; this also occurs for the other countries, but less adjustment is required as these forecasts are already reasonable. For the Netherlands, the fluctuations and recent declines within the central reproductive ages, together with their dominance, lead to severe extrapolations that the CFR constraint alone cannot mitigate with sufficient plausibility.

#### 4.1.2 Summary measures

Figure 5 presents the forecasts of the five summary measures, namely the CFR and PPRs, with their associated uncertainty. In row 1 we plot the posterior median CFR and corresponding 95% prediction interval under the aggregate and parity-specific models for each country. We first observe that the posterior medians for the two models generally align closely. For Canada, England & Wales and the Netherlands, both models predict a similar stable or increasing trend, followed by a gradual decline. Only for Czechia do the projections diverge considerably, the initial decrease giving way to a strong recovery under the parity-specific model and a plateau under the aggregate model. The difference is not surprising given the steeply decreasing trend in the observed CFR values for Czechia, contrasting with the much greater stability exhibited by the other countries.

Regarding the level of uncertainty, while the intervals for the two models overlap extensively, those for the parity-specific model are consistently narrower, meaning that the projections have greater precision. Lastly, although the additional data points from 2016 onwards can be used to assess forecast performance, these correspond to cohorts who were already nearing the end of their reproductive lives and thus their CFR values can be predicted with high accuracy. In Section 4.2 we assess the forecast performance for the earlier $JOY$ values, where longer time series of out-of-sample data are available.

We now consider the posterior distributions of the four PPRs, which are presented analogously in row 2 of Figure 5 (for the parity-specific model only). Czechia again differs from the other countries, for two main reasons: (1) the progression to motherhood ($PPR^{0->1}$) is much higher and shows a decreasing trend, and (2) the progressions to parities 3 and 4+ are much lower and closely overlapping, whereas for the other countries their distributions are more distinct and with $PPR^{2->3}$ at the higher level. This demonstrates how a variety of parity dynamics can lead to a similar CFR (e.g., see Zeman et al. 2018).

Concerning the projections, those for $PPR^{0->1}$ exhibit a generally similar trend to the corresponding CFR projections due to the closeness of the underlying rates (Section 3.2.1). The progression from parity 1 to 2 is predicted to remain stable for Canada, gently fluctuate for England & Wales, gradually decrease and then stabilize for Czechia, and the reverse of this for the Netherlands. Lastly, the higher-order PPRs are projected to diverge slightly for Czechia but converge for the other countries. The gradual nature of these predicted PPR and CFR changes reflects the addition of their respective constraints to incorporate prior knowledge regarding their relatively slow progressions in the past (Section 3.3).

## 4.2 Comparison with existing methods

In this subsection we assess the predictive performance of our proposed parity-specific projection model. We make two comparisons: (1) against the aggregate version of the model to directly evaluate the effect of incorporating parity information, and (2) against a selection of the best-performing (aggregate) methods in the existing literature to determine whether our parity-specific model has competitive forecast accuracy and precision. We consider three comparator methods:

- freeze rates approach: this simply fixes the ASFRs at their most recent observed values;
- model of Myrskylä et al. (2013): this extrapolates the recent age-specific ASFR trend for five years and then freezes the rates;
- model of Schmertmann et al. (2014): this is a Bayesian approach which incorporates prior information from historical HFD data regarding typical ASFR patterns across age and cohort to obtain projections that are plausible in both dimensions.

The models of Myrskylä et al. (2013) and Schmertmann et al. (2014) were found to forecast completed cohort fertility with the highest accuracy in the recent comparative paper of Bohk-Ewald et al. (2018). The freeze rates approach performed well despite its simplicity, leading the authors to recommend its usage as a benchmark when proposing any new fertility forecasting method.

### 4.2.1 Summary statistics

We first compare the predictive performance of these methods across all countries and *JOY* values simultaneously using five different summary statistics relating to the ASFR and CFR forecasts, which we present in Table 1. Note that of the three comparator methods from the existing literature, only the model of Schmertmann et al. (2014) is taken to have probabilistic forecasts[2]. The first three statistics are standard measures, namely the mean absolute error

[2]Although Myrskylä et al. (2013) provide a method for computing forecast intervals, their coverage has been shown to be poor (Bohk-Ewald et al. 2018; Ellison et al. 2020). We therefore only consider the point forecasts of this approach.

(MAE), mean absolute percentage error (MAPE) and the root mean square error (RMSE), where we take the point estimate for the probabilistic forecasts to be the posterior median of the forecast distribution[3].

For both the ASFR and CFR forecasts, these measures indicate that our proposed parity-specific model is preferred marginally over the aggregate version, and considerably over the three comparator methods. The next two rows contain the coverage of the 90% and 50% prediction intervals for the three probabilistic methods. The values for the ASFR forecasts are considerably lower than the nominal levels, with the aggregate and Schmertmann et al. (2014) models having the highest coverage for the 90% and 50% prediction intervals respectively. Contrastingly, the values for the CFR forecasts are generally much closer to the nominal levels, with the proposed parity-specific and aggregate models seemingly very well calibrated.

The final row presents the average value of a particular scoring rule. Scoring rules have been applied in the context of fertility forecasting by Ellison et al. (2020). They are designed to assess probabilistic rather than point forecasts and so can measure the performance of the entire forecast distribution (Gneiting and Raftery 2007). We compute the continuous ranked probability score (CRPS) (Matheson and Winkler 1976), which evaluates the quality of a forecast distribution based on how close its cumulative distribution function (CDF) is to that of the true value; it can also be calculated where only point forecasts are available (and is equivalent to the MAE), meaning that we can compare all of our methods simultaneously (Ellison et al. 2020). We note that smaller CRPS values indicate greater predictive accuracy. Table 1 indicates that the ASFR forecasts from the models of Myrskylä et al. (2013) and Schmertmann et al. (2014) are marginally preferable under the CRPS, with our parity-specific model being very close. However, for the CFR forecasts the parity-specific model is quite strongly preferable.

The summary statistics imply that overall, our proposed parity-specific model tends to outperform the corresponding aggregate model and is also competitive with some of the best-performing existing models. However this is at a high level of aggregation, and so next we summarize the differences we observe when repeating this exercise for a given country or $JOY$ value (results not shown). The conclusion regarding the comparison with the aggregate model generally holds, with only a few instances where the aggregate model is preferable based on at least one measure (ASFR: Czechia and 2000; CFR: Netherlands and 2015). The findings regarding the comparison with the existing methods are less consistent across countries and $JOY$ values. Across both the ASFR and CFR forecasts, the model of Myrskylä et al. (2013) and the freeze rates approach are preferable for the Netherlands, as well as for the later $JOY$ values under some of the measures (ASFR: 2005 and later; CFR: 2010 and later). This suggests that the generally stronger performance by our proposed parity-specific model for the earlier $JOY$ values, which contribute a larger proportion of the forecast errors, dominates the overall results.

Next, in Figure 6 we inspect the results broken down for each country-$JOY$ combination. For the ASFR and CFR forecasts (top and bottom rows respectively) and for each model (column), we plot the relative difference in the average CRPS compared to the parity-specific

[3]Note that to ensure a fair comparison between the deterministic models (which provide a perfect fit to the observed ASFRs) and the probabilistic models, the CFR summary statistics for the latter in Table 1 (excluding the coverages) and Figure 6 ignore the uncertainty about the ASFRs observed before the jump-off year, instead treating them as known.

model for each $JOY$ value ($x$-axis) and country (color). For our first comparison we look at the bars in the leftmost column. For both the ASFR and CFR forecasts, nearly all of the differences are positive, meaning that the average CRPS for the aggregate model is nearly always higher than that for the parity-specific model; this supports our findings from Table 1. For our second comparison we make a similar assessment of the next three columns, i.e. of the comparator methods from the existing literature. Considering the comparison with the 'benchmark' freeze rates approach first (rightmost column), we see that our parity-specific model outperforms this method in around half of the country-$JOY$ combinations for the ASFR forecasts, but nearly all of the combinations for the CFR forecasts. For the models of Myrskylä et al. (2013) and Schmertmann et al. (2014) the results are more mixed, with the parity-specific model preferable in the majority of combinations only for the CFR forecasts of the Schmertmann et al. (2014) model.

#### 4.2.2 Visual assessment

Next we visually inspect the projections, starting with plotting the CFR forecast distributions for the four countries for $JOY$ values 2000 and 2005 in rows 1 and 2 of Figure 7. We choose these earliest jump-off years as they provide the longest time series of out-of-sample data points. In most cases the prediction intervals or point forecasts for the five methods overlap to some extent. However, the 2000-based forecasts for Czechia provide an interesting example whereby the predictions diverge sharply from around the 1970 cohort; the parity-specific and aggregate approaches anticipate a recovery to replacement level, while the other methods predict a steep decline to unprecedented levels. The reality lies in the middle, with a gradual CFR decline being observed that lies much closer to the point forecasts from our proposed models; this explains the poor performance of the comparator methods in Figure 6. In contrast, the 2000-based forecasts for the Netherlands demonstrate an instance where the naive freeze rates method performs best, with the alternative methods predicting a recovery in the CFR that does not materialize. However, in the 2005-based forecasts for all countries except the Netherlands, the freeze rates approach is outperformed by the more optimistic trajectories produced by the methods that account for the recent trend in some way.

Lastly, in rows 3–5 of Figure 7 we examine the ASFR forecasts for the 1975 cohort at $JOY$ values 2000, 2005 and 2010, when women in this cohort were aged 25, 30 and 35 respectively (we will refer to the age of the cohort in the year $JOY$ as the truncation age). This allows us to appreciate the challenge of completing the cohort schedule at young truncation ages. For example, the cause of the divergent 2000-based CFR forecasts for Czechia in row 1 of Figure 7 is clear looking at the underlying ASFRs for the 1975 cohort – the age pattern has peaked and is stable, but when it will begin to decline (or whether it will peak further) is difficult to determine, hence the very different forecasts. However, the projections for the same cohort under the later $JOY$ values tell a very different story, exhibiting much greater consistency due to a larger proportion of the cohort schedule having been observed. A similar pattern is visible in the Netherlands panels, with the methods unable to agree on the location and height of the peak in ASFRs in the 2000-based forecast, which becomes much less of a concern in the later forecasts. The same is true for Canada and England & Wales but to a lesser extent.

## 5 Discussion and conclusions

In this paper we propose, to the best of our knowledge, the first integrated methodological framework for producing fertility projections by parity. The model generates smooth age-cohort surfaces of rates for each parity which are combined to give an aggregate rate surface, allowing summary fertility measures such as completed family size (via the cohort total fertility rate, or CFR) and parity progression ratios to be computed within the model. Taking a Bayesian approach makes it straightforward to apply constraints both to these measures and to the smooth surfaces themselves, thus incorporating prior knowledge and improving forecast plausibility; it also allows the model to be informed by data sources at the parity-specific and aggregate levels in a coherent way.

The key methodological challenge to overcome is how to model *conditional* parity-specific fertility rates, i.e. rates that depend on the number of women at risk of each particular birth order, when in most cases data of sufficient quality is only available for the numerator, i.e. the number of order-specific births. To this end, we exploit the sequential progression of parity through a person's life and approximate the denominators by cumulating across cohort fertility (Jasilioniene et al. 2015; Smallwood 2002). This is similar to the approach taken by Ellison et al. (2023), however an important difference is that our cumulation occurs during model fitting rather than only being applied post-hoc to separate model outputs for each parity. The implementation of such an approximation would be necessary even if the denominators for the fitting period were known, as estimates consistent with the rate projections would be required for the forecast period. Despite this, the fact remains that fitting the proposed model relies on having at least some parity-specific data available at the population level, which may pose a difficulty in less-developed countries.

A second challenge is how to obtain forecasts that are plausible, especially given that we are extrapolating multiple two-dimensional smooth surfaces within the same model. We do this by applying various constraints that express our prior knowledge about the likely progression of demographic rates (see Camarda 2019 for an example of such an approach in the context of mortality forecasting). The first instance is at the parity-specific level (Section 3.2.1), constraining the parameters to borrow strength across cohorts where the rates are erratic at the youngest ages, and to enforce the known biological limitations of fertility at the oldest ages. The second application is at the aggregate level (Section 3.3), where we constrain summary fertility measures to change slowly from cohort to cohort, informed by the historical variability of these measures in a pooled dataset from the Human Fertility Database (Human Fertility Database 2023). These constraints, when taken together, should lead to increased forecast plausibility in terms of the shapes of the cohort schedules of rates at the parity-specific level, and the progression of summary measures at the aggregate level.

To assess the forecast performance of our proposed model we fit it to four countries, each across four cohort ranges with a different jump-off year ($JOY$), i.e. the last year of data used for model fitting. We also fit an analogous 'aggregate' model, which only smooths the overall age-specific fertility rates (ASFRs) and constrains the CFR, as well as several competitive aggregate models from the literature, informed by the results of Bohk-Ewald et al. (2018). We perform a comprehensive quantitative and visual comparison of the resulting ASFR and CFR forecasts to ascertain whether the incorporation of parity information improves forecast accuracy and precision. We can address this question directly by comparing the parity-specific and aggregate models, which only differ in this respect; we find that the former is strongly preferable under nearly all of the measures. In contrast, the comparison

with existing approaches has mixed findings, with the comparator methods tending to perform better for particular countries and *JOY* values. However, this comparison only addresses the question indirectly – there are many differences between our proposed model and the existing methods which do not relate to parity and cannot be eliminated. We consider the discrepancies in more detail below.

Under the 'benchmark' freeze rates approach, it is interesting that noticeably poorer ASFR and CFR forecasts under our proposed model are evidenced only for the Netherlands, for *JOY* values 2000 and 2010 (Figure 6); we also note that these are the only occasions where freeze rates is the preferred model overall. In both cases, our projected parity 0 cohort schedules appear quite extreme, with steep increases across cohorts at the central reproductive ages that are not realized. In contrast with the 2015-based forecasts (Section 4.1.1), it seems likely that this is driven by the strongly *rising* trends in the observed rates at the end of the fitting period. Clearly this is also a challenge for the comparator methods, however it indicates that our proposed model could benefit from further methodological development. This could include applying additional constraints to improve forecast plausibility, modelling countries jointly (Schmertmann et al. 2014 provides an example of this), and accounting for inter-parity correlations.

Other areas for future investigation include extending the comparative analysis, for example to cover a wider range of countries and time periods, or to use as case studies countries that produce annual age-parity population estimates, such as Nordic countries through population registers, to assess the accuracy of our approximation method. Conversely, we could consider countries with limited parity-specific data availability to explore model adaptations to borrow strength across countries. Our proposed model could be refined further, for example by testing the sensitivity to different implementations of the constraints or prior specifications. We could also extend the model to incorporate the last of the "five clocks" (Raftery et al. 1996:1), i.e. time since last birth, to capture the important dependence of birth events on this variable (Ellison et al. 2022). Furthermore, scenario testing could be of great interest to policymakers, whereby the effect of various parity-specific interventions on overall childbearing levels could be explored. Lastly, our proposed model could be incorporated into a kinship projection model, where advancements have already been made through an efficient matrix formulation and the inclusion of parity as a characteristic (Caswell 2020).

In conclusion, our proposed model provides a sophisticated Bayesian methodology for obtaining parity-specific fertility projections. The ability of such projections to express aggregate rates in terms of the underlying parity dynamics that interact to produce them undoubtedly adds a great deal of richness and incorporates an awareness of the sequential nature of childbearing, features which are currently absent from existing methods. The model is also efficient to run and appears to have reasonable forecast performance when compared to some of the best-performing models from the literature. This demonstrates that enhancing explanatory capacity, in this case through explicitly modelling the childbearing mechanism by parity, yields a model that is at least comparable in terms of its predictive performance and errors. This provides a strong argument for the use and further development of fertility projections that more accurately capture the process of childbearing, thus having the potential to enable more tailored policy solutions.

## Acknowledgements

This work was partly supported by the Economic and Social Research Council (ESRC) FertilityTrends project (grant ES/S009477/1) and the Centre for Population Change Connecting Generations Centre (grant ES/W002116/1). Earlier versions of this work have been presented at the Nordic Demographic Symposium (June 2022), the European Population Conference (July 2022), the British Society for Population Studies conference and the Royal Statistical Society Annual Conference (both September 2022). The authors thank Ann Berrington, Sarah Christison, Bernice Kuang, Hill Kulu, Jon Forster, Jason Hilton and Peter W. F. Smith for their feedback during the research phase of the project.

## Data availability

The data that support the findings of this study are openly available from various sources, with specific references provided in the paper. Code to replicate all results in this paper can be accessed at https://doi.org/10.5281/zenodo.10535250. Where code has been sourced from a third party, full details of where to obtain the code and any modifications made by the authors have been provided.

## References

Aburto, J. M., Kashyap, R., Schöley, J., Angus, C., Ermisch, J., Mills, M. C., and Dowd, J. B. (2021). Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *Journal of Epidemiology and Community Health*, 75(8):735–740.

Andersson, G. (1999). Childbearing Trends in Sweden 1961–1997. *European Journal of Population*, 15:1–24.

Andreev, E. M., Shkolnikov, V. M., and Begun, A. Z. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, 7:499–522.

Begall, K. and Mills, M. C. (2013). The Influence of Educational Field, Occupation, and Occupational Sex Segregation on Fertility in Netherlands. *European Sociological Review*, 29:720–742.

Billingsley, S. (2010). The Post-communist Fertility Puzzle. *Population Research and Policy Review*, 29(2):193–231.

Bohk-Ewald, C., Li, P., and Myrskylä, M. (2018). Forecast accuracy hardly improves with method complexity when completing cohort fertility. *Proceedings of the National Academy of Sciences*, 115(37):9187–9192.

Bongaarts, J. and Sobotka, T. (2012). A Demographic Explanation for the Recent Rise in European Fertility. *Population and Development Review*, 38(1):83–120.

Booth, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting*, 22:547–581.

Camarda, C. G. (2019). Smooth constrained mortality forecasting. *Demographic Research*, 41(38):1091–1130.

Caswell, H. (2020). The formal demography of kinship II: Multistate models, parity, and sibship. *Demographic Research*, 42(38):1097–1146.

Currie, I. D., Durban, M., and Eilers, P. H. C. (2004). Smoothing and forecasting mortality rates. *Statistical Modelling*, 4:279–298.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science*, 11(2):89–121.

Ellison, J., Berrington, A., Dodd, E., and Forster, J. J. (2022). Investigating the application of generalized additive models to discrete-time event history analysis for birth events. *Demographic Research*, 47:647–694.

Ellison, J., Berrington, A., Dodd, E., and Forster, J. J. (2023). Combining individual- and population-level data to develop a Bayesian parity-specific fertility projection model. *Journal of the Royal Statistical Society: Series C*, qlad095.

Ellison, J., Dodd, E., and Forster, J. J. (2020). Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society: Series A*, 183(3):829–856.

Feeney, G. and Yu, J. (1987). Period Parity Progression Measures of Fertility in China. *Population Studies*, 41(1):77–102.

Gleditsch, R. F., Syse, A., and Thomas, M. J. (2021). Fertility Projections in a European Context: A Survey of Current Practices among Statistical Agencies. *Journal of Official Statistics*, 37(3):547–568.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Hellstrand, J., Nisén, J., Miranda, V., Fallesen, P., Dommermuth, L., and Myrskylä, M. (2021). Not Just Later, but Fewer: Novel Trends in Cohort Fertility in the Nordic Countries. *Demography*, 58(4):1373–1399.

Hellstrand, J., Nisén, J., and Myrskylä, M. (2020). All-time low period fertility in Finland: Demographic drivers, tempo effects, and cohort implications. *Population Studies*, 74(3):315–329.

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*, 68:29–49.

Hilton, J., Dodd, E., Forster, J. J., and Smith, P. W. F. (2021). Modelling Frontier Mortality Using Bayesian Generalised Additive Models. *Journal of Official Statistics*, 37(3):569–589.

Hobcraft, J. (1996). Fertility in England and Wales: A Fifty-Year Perspective. *Population Studies*, 50(3):485–524.

Human Fertility Database (2023). Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). http://www.humanfertility.org. Data downloaded 5 July 2023.

Human Mortality Database (2023). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). http://www.mortality.org. Data downloaded 6 July 2023.

Jasilioniene, A., Jdanov, D. A., Sobotka, T., Andreev, E. M., Zeman, K., Shkolnikov, V. M., and with contributions from J Goldstein, E J Nash, D Philipov, and G Rodriguez (2015). *Methods Protocol for the Human Fertility Database*. http://www.humanfertility.org/Docs/methods.pdf. Accessed 27 April 2017.

Kohler, H.-P. and Ortega, J. A. (2002). Tempo-adjusted period parity progression measures, fertility postponement and completed cohort fertility. *Demographic Research*, 6:91–144.

Kravdal, Ø. (2001). The High Fertility of College Educated Women in Norway: An Artefact of the Separate Modelling of Each Parity Transition. *Demographic Research*, 5:187–216.

Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. *Management Science*, 22(10):1087–1096.

Mazzuco, S. and Keilman, N. (2020). *Developments in Demographic Forecasting*. Springer Nature.

McDonald, P., Hosseini-Chavoshi, M., Abbasi-Shavazi, M. J., and Rashidian, A. (2015). An assessment of recent Iranian fertility trends using parity progression ratios. *Demographic Research*, 32:1581–1602.

Murphy, M. and Berrington, A. (1993). Constructing period parity progression ratios from household survey data. In Ní Bhrolcháin, M., editor, *New Perspectives on Fertility in Britain*, pages 17–33. London: HMSO.

Myrskylä, M., Goldstein, J. R., and Cheng, Y. A. (2013). New cohort fertility forecasts for the developed world: Rises, falls, and reversals. *Population and Development Review*, 39(1):31–56.

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.

Ní Bhrolcháin, M. (1992). Period paramount? A critique of the cohort approach to fertility. *Population and Development Review*, 18(4):599–629.

Office for National Statistics (2018). *Quality assurance of data from the number of previous children question at birth registrations, England and Wales: 2016*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/methodologies/qualityassuranceofdatafromthenumberofpreviouschildrenquestionatbirthregistrationsenglandandwales2016. Accessed 5 October 2022.

Office for National Statistics (2022a). *Fertility rates by parity 1934 to 2020, England and Wales*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/adhocs/14363fertilityratesbyparity1934to2020englandandwales. Accessed 22 September 2022.

Office for National Statistics (2022b). *Zipped population projections data files, Great Britain and England and Wales*. https://www.ons.gov.uk/peoplepopulationandcommunit

y/populationandmigration/populationprojections/datasets/z2zippedpopulationprojectionsdatafilesgbandenglandandwales. Accessed 22 September 2022.

Office for National Statistics (2023). *User guide to birth statistics*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/methodologies/userguidetobirthstatistics. Accessed 6 November 2023.

Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M. Z., Barrow, D. K., Taieb, S. B., Bergmeir, C., Bessa, R. J., Bijak, J., Boylan, J. E., et al. (2022). Forecasting: theory and practice. *International Journal of Forecasting*, 38(3):705–871.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

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.

Rallu, J.-L. and Toulemon, L. (1994). Period Fertility Measures: The Construction of Different Indices and their Application to France, 1946-89. *Population: An English Selection*, 6:59–93.

Schmertmann, C., Zagheni, E., Goldstein, J. R., and Myrskylä, M. (2014). Bayesian Forecasting of Cohort Fertility. *Journal of the American Statistical Association*, 109(506):500–513.

Shang, H. L. and Booth, H. (2020). Synergy in fertility forecasting: improving forecast accuracy through model averaging. *Genus*, 76(27).

Smallwood, S. (2002). New estimates of trends in births by birth order in England and Wales. *Population Trends*, 108:32–48.

Stan Development Team (2023). *RStan: The R interface to Stan*. R package version 2.26.13. http://mc-stan.org/.

United Nations, Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022, Online Edition. *File POP/01-3: Female population by single age, region, subregion and country, annually for 1950-2100 (thousands)*. https://population.un.org/wpp/Download/SpecialAggregates/EconomicTrading/. Accessed 13 October 2022.

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*, 32:585–610.

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R, Second Edition*. Chapman & Hall/CRC Press.

Zeman, K., Beaujouan, E., Brzozowska, Z., and Sobotka, T. (2018). Cohort fertility decline in low fertility countries: Decomposition using parity progression ratios. *Demographic Research*, 38:651–690.

Zhang, J. L. and Bryant, J. (2019). Combining multiple imperfect data sources for small area estimation: a Bayesian model of provincial fertility rates in Cambodia. *Statistical Theory and Related Fields*, 3(2):178–185.

Table 1: Summary statistics for the ASFR and CFR forecasts under various models

| Forecast | Measure | P-S | A | S | M | FR |
|---|---|---|---|---|---|---|
| ASFR | MAE | **0.0063** | 0.0070 | 0.0097 | 0.0073 | 0.0083 |
| | MAPE (%, 1 decimal place) | **11.5** | 12.9 | 21.9 | 14.3 | 18.3 |
| | RMSE | **0.011** | **0.011** | 0.015 | 0.012 | 0.013 |
| | Coverage of 90% PI (%) | 66 | **73** | 69 | - | - |
| | Coverage of 50% PI (%) | 30 | 37 | **38** | - | - |
| | Average CRPS | 0.0074 | 0.0088 | **0.0073** | **0.0073** | 0.0083 |
| CFR | MAE | **0.024** | 0.028 | 0.041 | 0.033 | 0.048 |
| | MAPE (%, 1 decimal place) | **1.3** | 1.5 | 2.3 | 1.8 | 2.6 |
| | RMSE | **0.051** | 0.057 | 0.084 | 0.074 | 0.096 |
| | Coverage of 90% PI (%) | 82 | **83** | 69 | - | - |
| | Coverage of 50% PI (%) | **52** | 58 | **48** | - | - |
| | Average CRPS | **0.024** | 0.030 | 0.032 | 0.033 | 0.048 |

*Notes*: Summary statistics are calculated across all countries and *JOY* values for the ASFR and CFR forecasts under the proposed parity-specific (P-S) and aggregate (A) models, the Schmertmann et al. (S) and Myrskylä et al. (M) models, and the freeze rates approach (FR). Values are given to 2 significant figures unless stated otherwise. MAE = mean absolute error; MAPE = mean absolute percentage error; RMSE = root mean square error; PI = prediction interval; CRPS = continuous ranked probability score. The most preferable value/values is/are in **bold**.

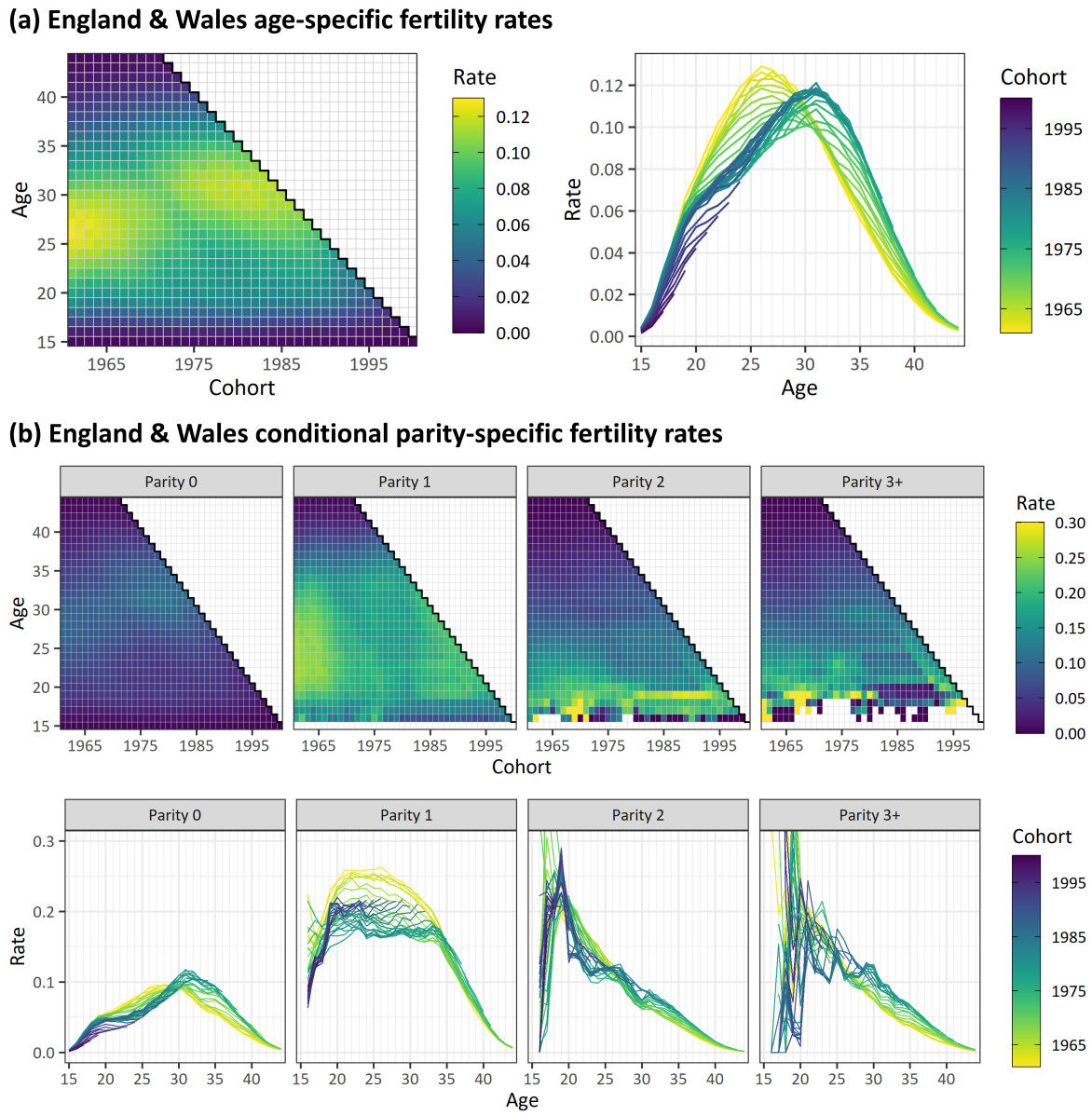


Figure 1: (a) Left: Lexis surface of England & Wales age-specific fertility rate estimates for the 1961–2000 cohorts, observed up to $JOY = 2015$, from Office for National Statistics (2022a). Right: same data presented in the form of cohort-specific rate curves. (b) Top row: Lexis surfaces of England & Wales conditional parity-specific fertility rate estimates for the 1961–2000 cohorts and parities 0, 1, 2 and 3+, observed up to $JOY = 2015$, from Office for National Statistics (2022a); note that the rate estimates have been top-coded at 0.3. Bottom row: same data presented in the form of cohort-specific rate curves. Note the non-standard age-cohort axes in the Lexis surfaces.

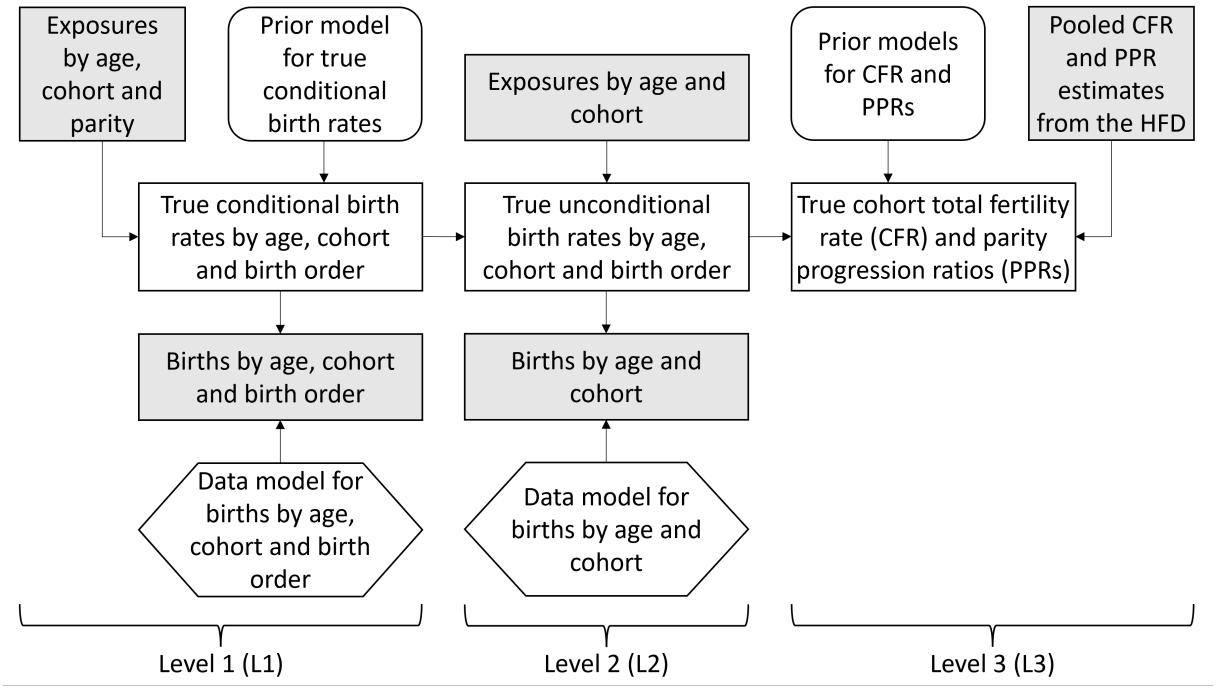


Figure 2: Overview of the proposed modelling framework. Straight-edged rectangles are demographic arrays, rounded rectangles are prior models and hexagons are data models. Only the gray rectangles are observed.

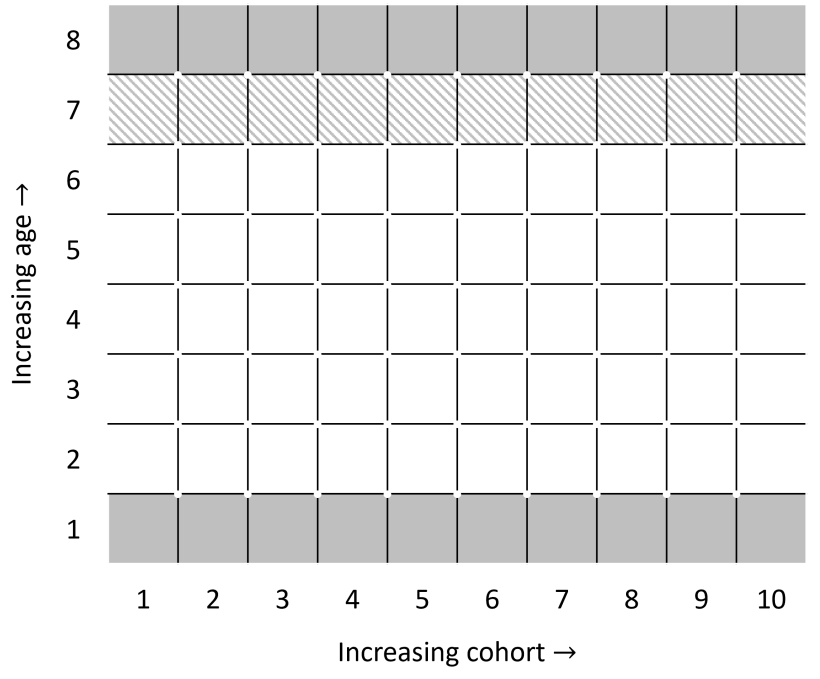


Figure 3: Illustration of penalties applied to basis function coefficients for a 2D age-cohort surface with $P = 8$ basis functions across age and $Q = 10$ across cohort. Gray areas: coefficients fixed across cohort; White and hatched areas: first-order differences penalized in cohort direction; White area: first-order differences penalized in age direction.

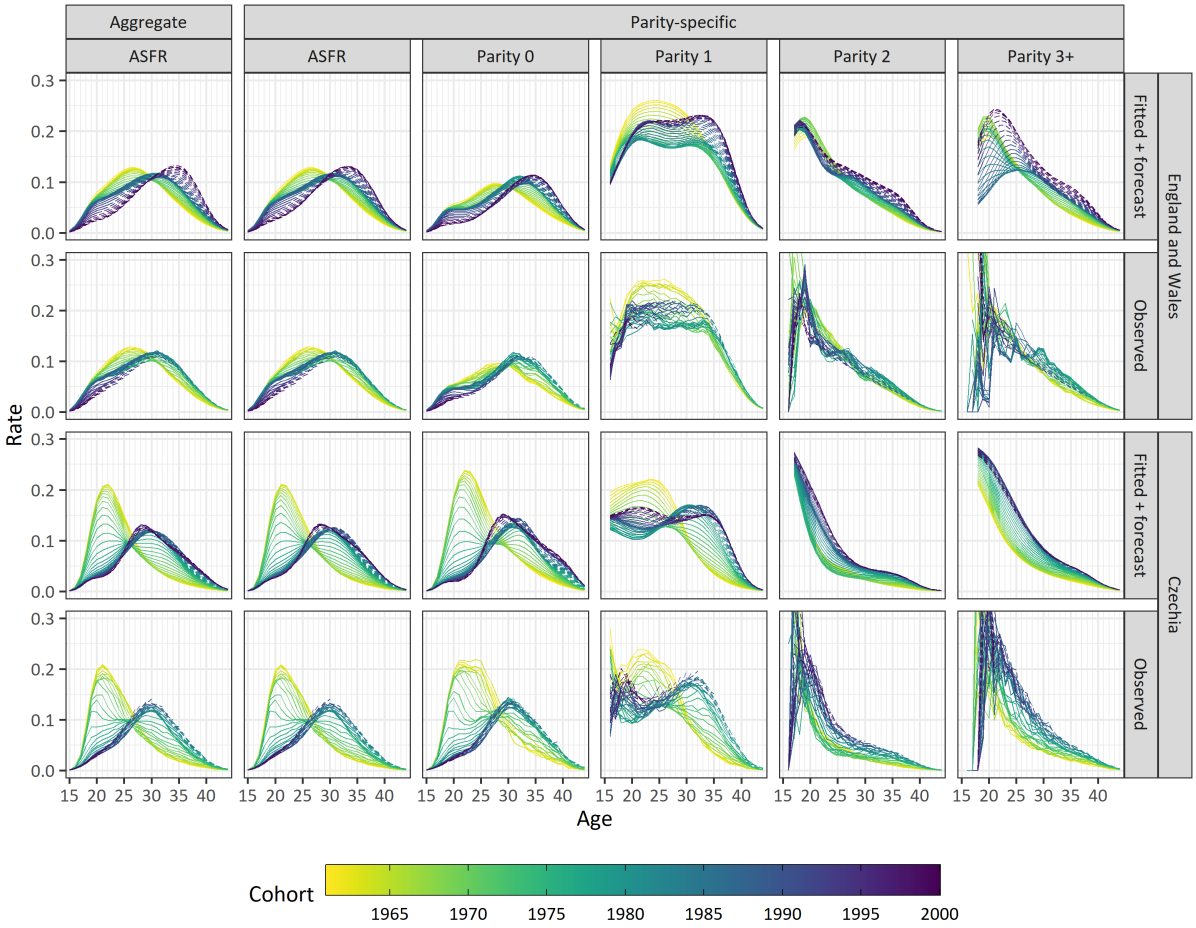


Figure 4: Row 1: England & Wales posterior median ASFRs and parity-specific rates from the aggregate and parity-specific fertility projection models for $JOY = 2015$; Row 2: corresponding observed rates from Office for National Statistics (2022a). Rows 3 and 4: Equivalent plots for Czechia, with observed rates from Human Fertility Database (2023). Dashed lines indicate values after the year $JOY$.

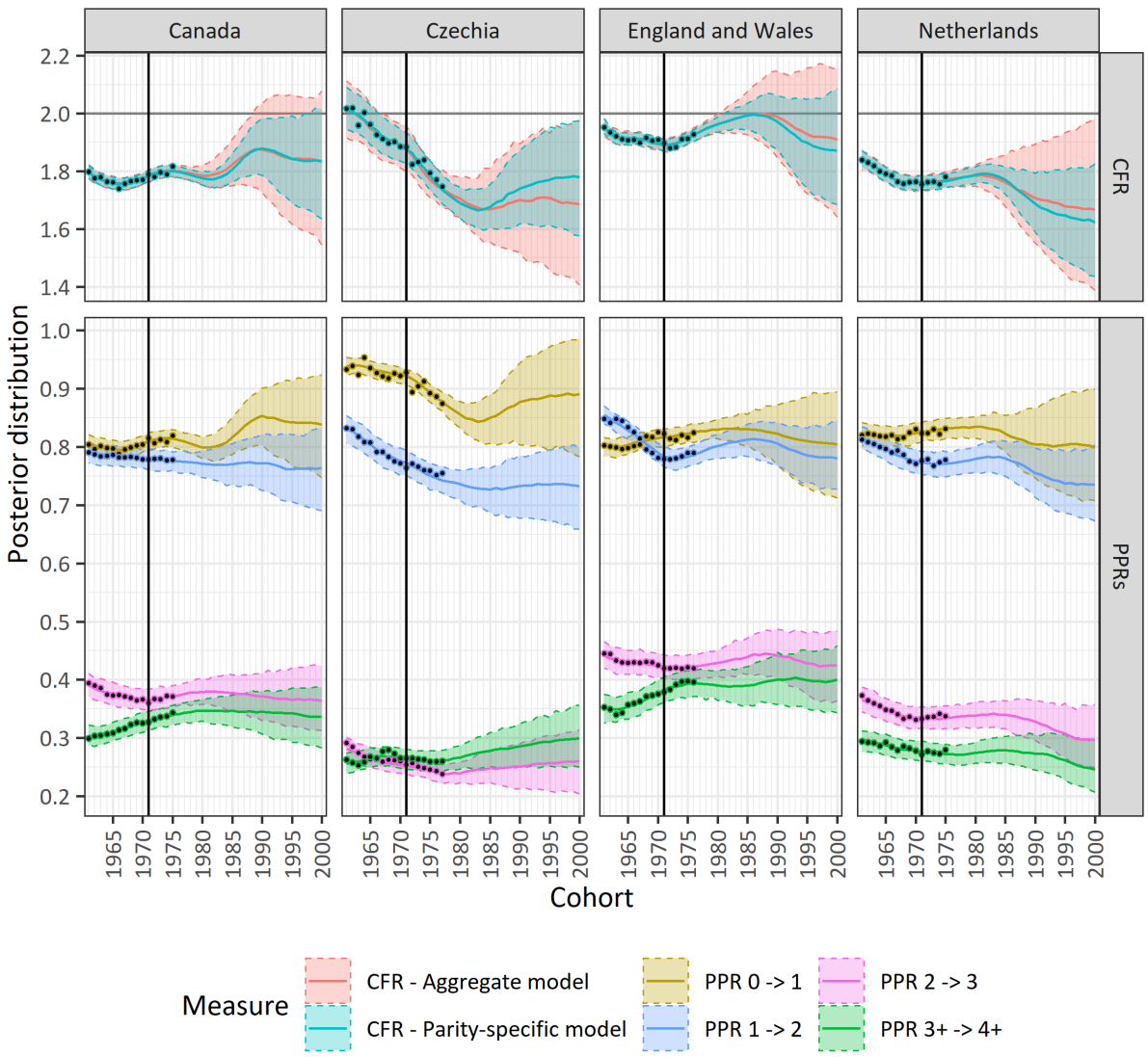


Figure 5: CFR (top row) and PPR (bottom row) posterior medians and 95% prediction intervals for the aggregate and parity-specific fertility projection models for $JOY = 2015$ and the selected countries; note that the PPRs are only estimable by the parity-specific model. Points are observed values from Office for National Statistics (2022a) and Human Fertility Database (2023); vertical lines indicate the last fully-observed cohort (born in the year 1971 = $JOY - a_{max}$).

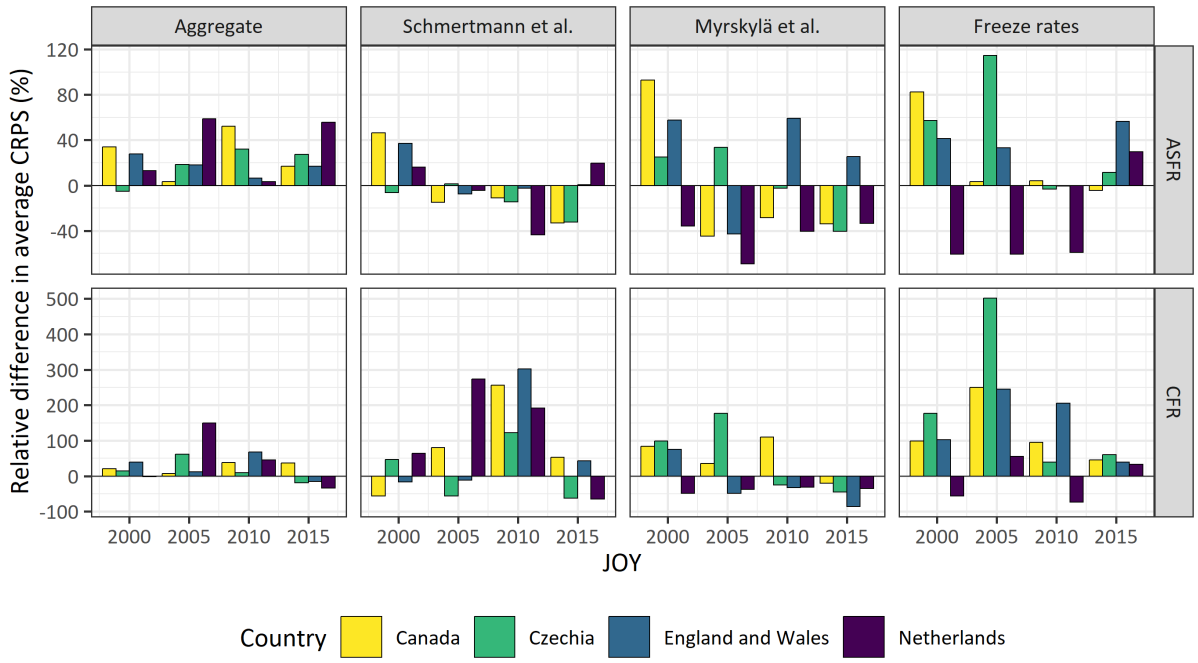


Figure 6: Relative differences in the average CRPS compared to the proposed parity-specific model for the aggregate, Schmertmann et al. and Myrskylä et al. models, and the freeze rates approach, for each country-*JOY* combination and the ASFR (top row) and CFR (bottom row) forecasts.

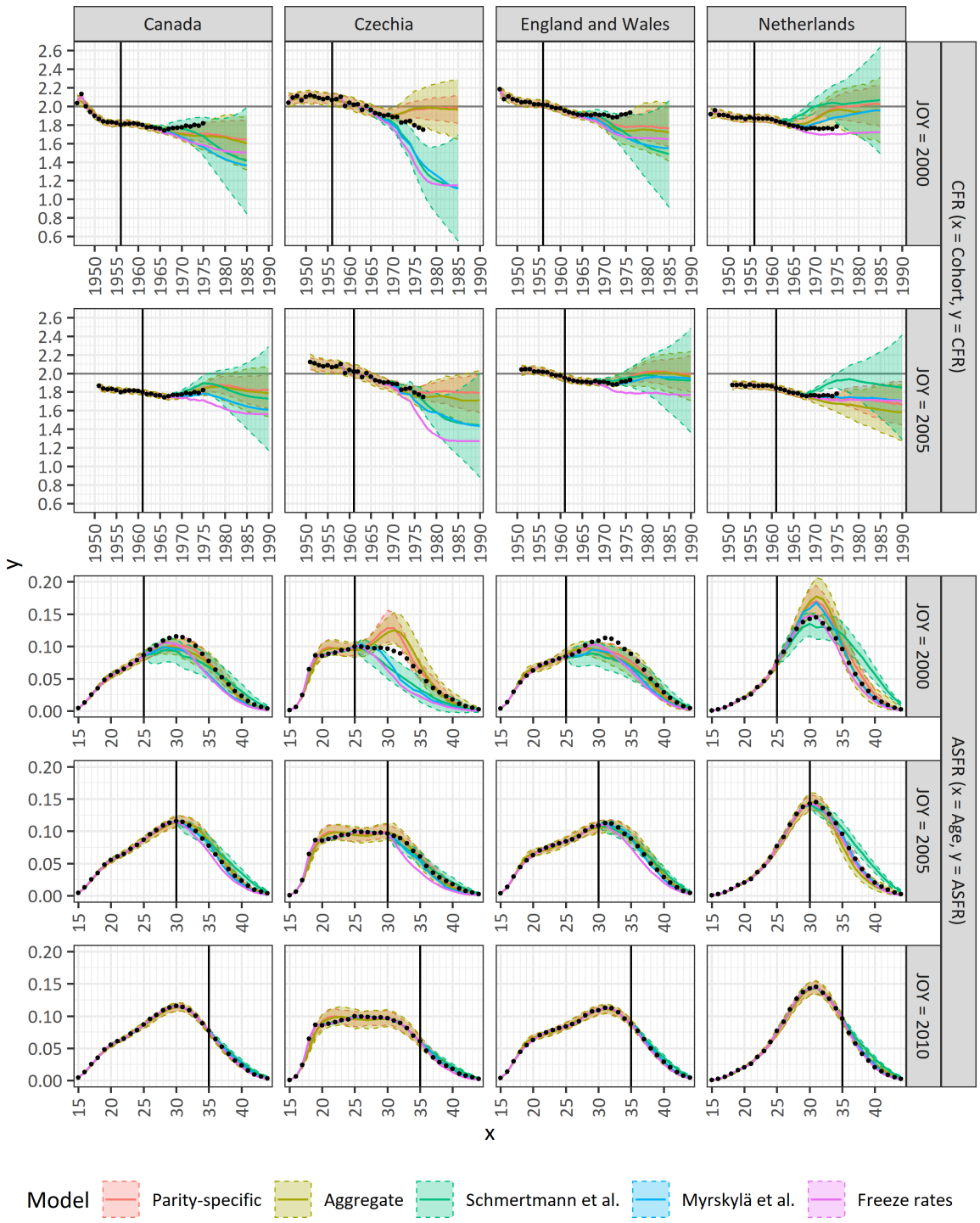


Figure 7: Rows 1 and 2: CFR posterior medians and 95% prediction intervals for the proposed parity-specific and aggregate models and the Schmertmann et al. model, and point forecasts for the Myrskylä et al. model and the freeze rates approach, for $JOY$ values 2000 and 2005, and the selected countries; vertical lines indicate the last fully-observed cohort (born in the year $JOY - a_{max}$). Rows 3–5: For the same models and the selected countries, ASFR posterior medians and 95% prediction intervals for the 1975 cohort at $JOY$ values 2000, 2005 and 2010; vertical lines indicate the truncation age (equal to $JOY - 1975$). Points are observed values from Office for National Statistics (2022a) and Human Fertility Database (2023).

# Appendices

## A Adjusting the post-2012 England & Wales birth counts by birth order

In May 2012, a change in legislation came into effect stating that information on true birth order was to be requested at all birth registrations in England & Wales (Office for National Statistics 2023). Analysis of the births by birth order data collected following this change found that there was a large decrease in the proportion of women reporting that they had no previous live births compared to 2011 estimates (Office for National Statistics 2018). This was thought to be due to a lack of clarity in the wording of the question or the way in which it was being asked, leading to misreporting whereby the current child, i.e. the child being registered, was included in the count of previous children. The greatest decrease was observed for teenagers, where the proportion with no previous children should be the highest. In subsequent years the age-specific proportions of women reporting no previous children increased across the reproductive age range, and in 2016 were very close to the 2011 estimates. This is thought to be in part due to an amendment made to the question on the online registration system in 2016 to improve its clarity.

To improve the accuracy of the England & Wales data that we use to fit our parity-specific fertility projection model, we make an adjustment to the birth counts. We consider ages 14–19 and the years 2012–2016, where the phenomenon was observed most strongly. For each age, we perform the following steps:

1. For each year, calculate the proportion of total births attributable to each birth order.
2. Adjust these proportions so that they change linearly from 2011–2017 for each order.
3. Adjust the order-specific birth counts to be in accordance with the new proportions, ensuring that the total number of births remains fixed.
4. Recalculate the approximate parity-specific exposures using the method of Smallwood (2002) (see Section 2.2).

In Figure A.1 we plot the underlying order-specific birth proportions, with the original and adjusted proportions represented by the solid and dashed lines respectively and the 2011–2017 period indicated by the shaded area. The largest changes are in the proportions of first and second births, with the adjusted proportions exhibiting substantial increases and decreases respectively compared to the original proportions. The adjusted proportions for third and higher-order births, which have been combined for simplicity, are also lower than the original proportions, but only marginally. In this way, the main adjustment being made is the reclassification of second births as first births, thereby reducing the proportion of women reporting a previous birth as desired.

Figure A.2 plots the original and adjusted parity-specific fertility rate estimates in the top and bottom rows respectively. Following the discussion of Figure A.1, we should expect to see increased parity 0 rates and decreased parity 1 rates at ages under 20 over the 2012–2016 period. Indeed we see that the parity 1 rates are noticeably lower at the youngest ages, much more consistent with earlier and later years. The parity 2 rates are also generally lower at these ages, but to a lesser extent, and we see very little difference for parities 3 and 4+. Interestingly, the adjusted parity 0 rates are very similar to the original rates despite the increase in first births – this is likely due to the large exposures meaning that the relatively large increase in the numerator makes a negligible difference on the scale of the rates.

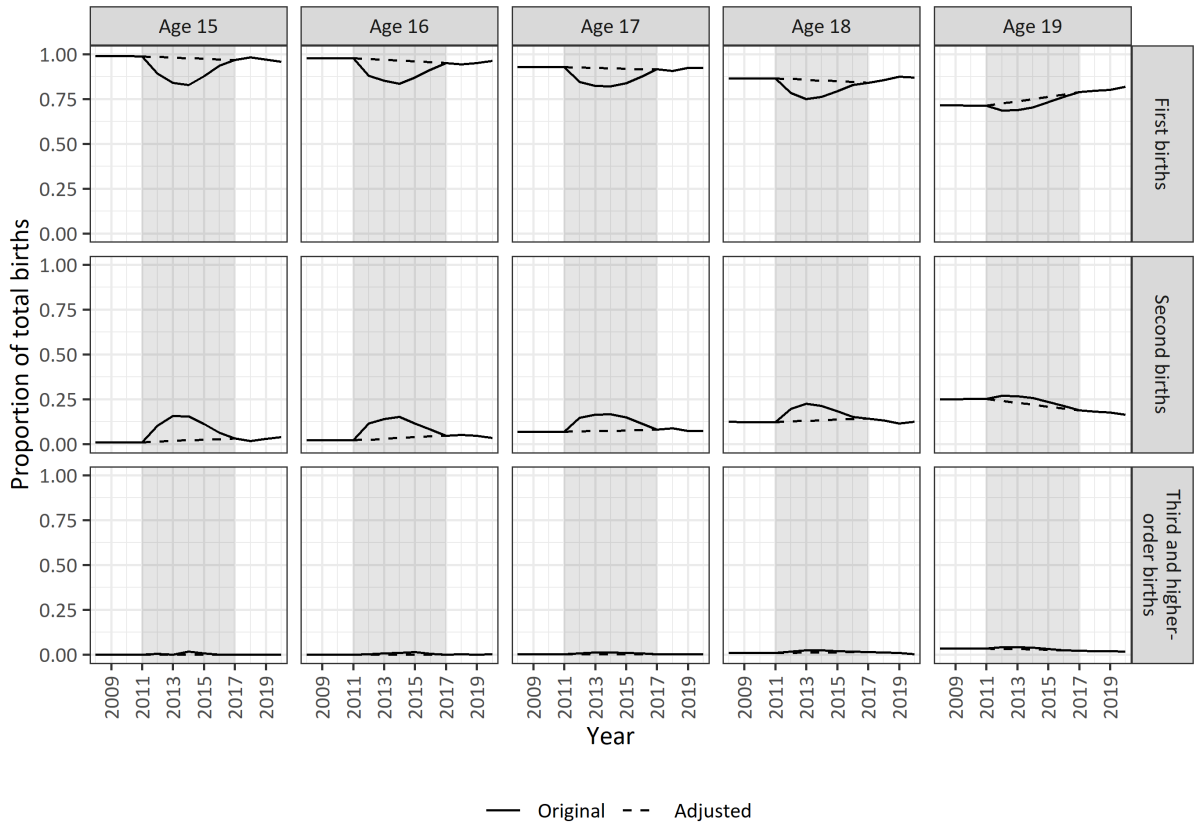


Figure A.1: England & Wales original (solid lines) and adjusted (dashed lines) order-specific birth proportions by year for the period 2008–2020, birth orders 1, 2 and 3+, and ages 15–19 (Office for National Statistics 2022a); the shaded area indicates the period over which the adjustment was made (2011–2017).

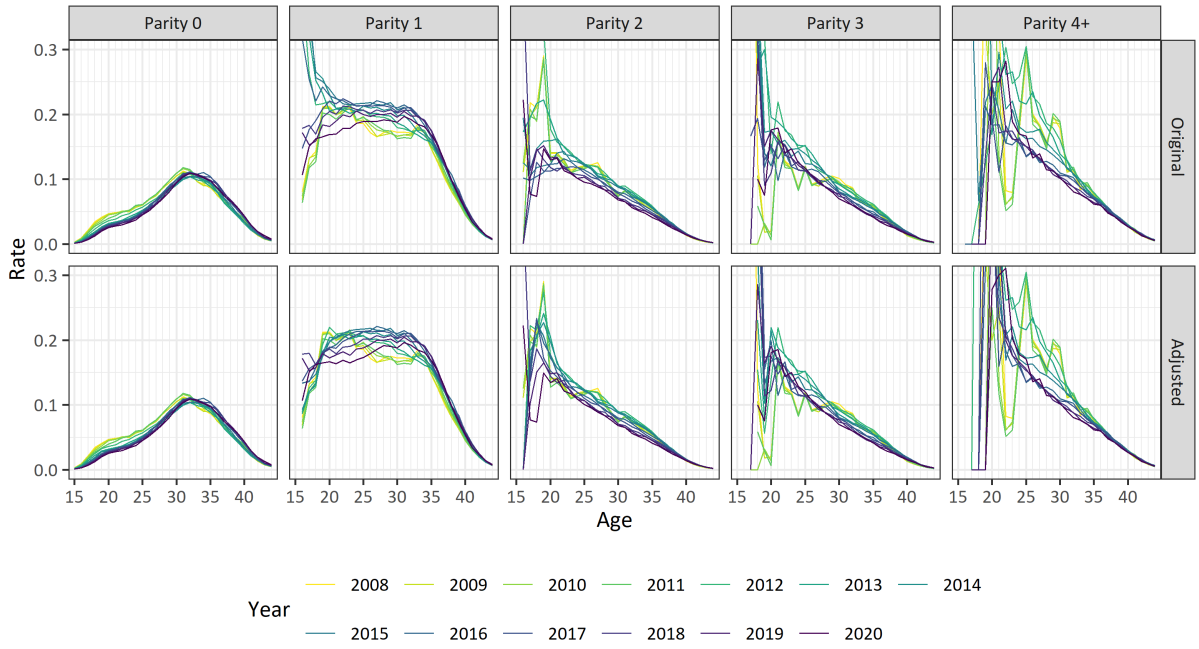


Figure A.2: England & Wales original (top row) and adjusted (bottom row) conditional fertility rate estimates for 2008–2020 and parities 0–4+ (Office for National Statistics 2022a).

## B Illustration of P-spline approach for the univariate case

We illustrate the P-spline approach in Figure B.1 for the simple one-dimensional case where we are smoothing across a single variable, in this case age. Below we describe the three steps depicted in the panels of the figure:

1. Step 1 displays the cubic B-spline basis with eight basis functions.
2. Step 2 scales each of the basis functions by different weights which are constrained to change smoothly across age. Summing the values of the weighted curves for each age then gives the black smooth curve of fitted log-rates.
3. Step 3 exponentiates the fitted log-rates to give the smooth fitted rate curve.

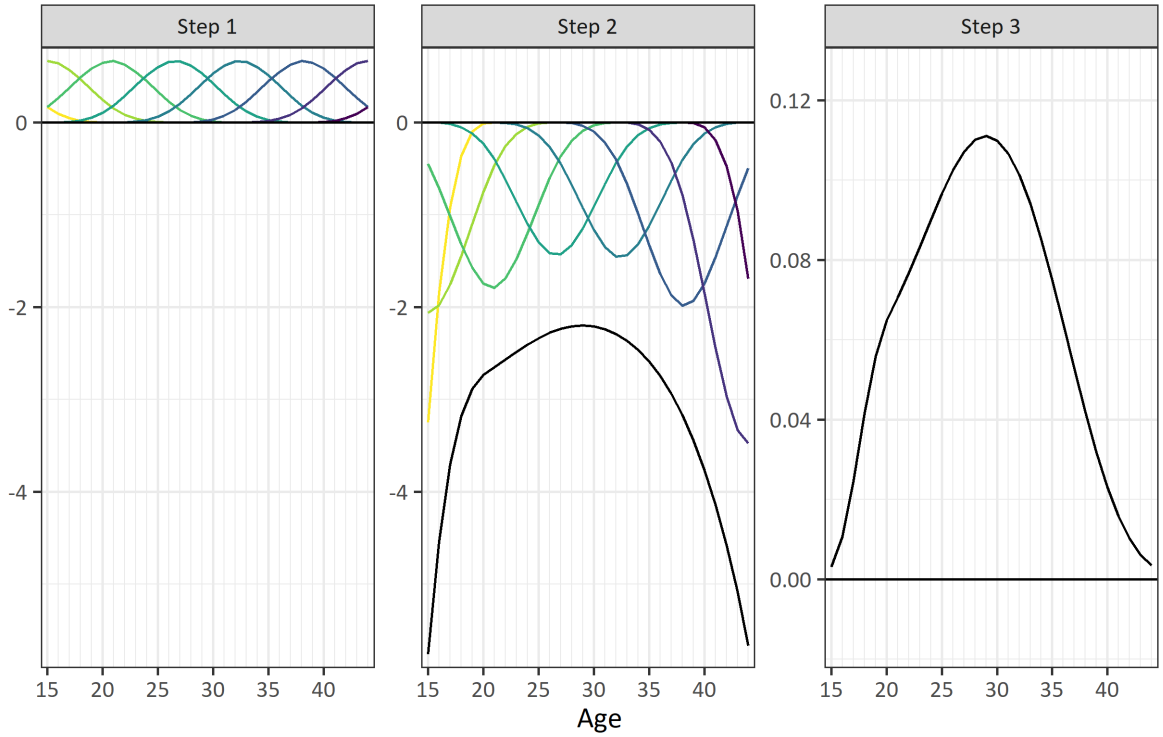


Figure B.1: Illustration of smoothing of age pattern of log-rates under a P-spline approach.

## C Summary fertility measure estimates for $JOY = 2015$

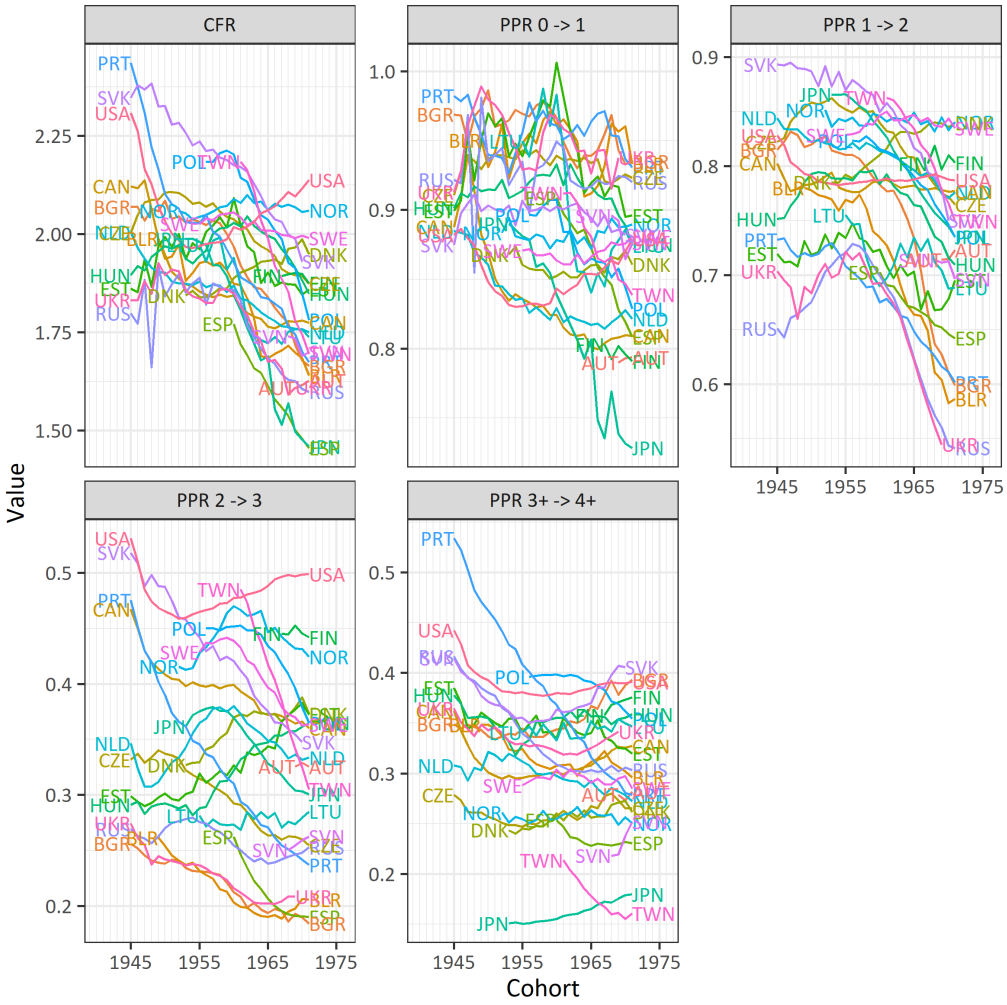


Figure C.1: CFR and PPR estimates for the 1945–1971 cohorts, computed from Human Fertility Database (2023).

# D Additional rate forecasts for $JOY = 2015$

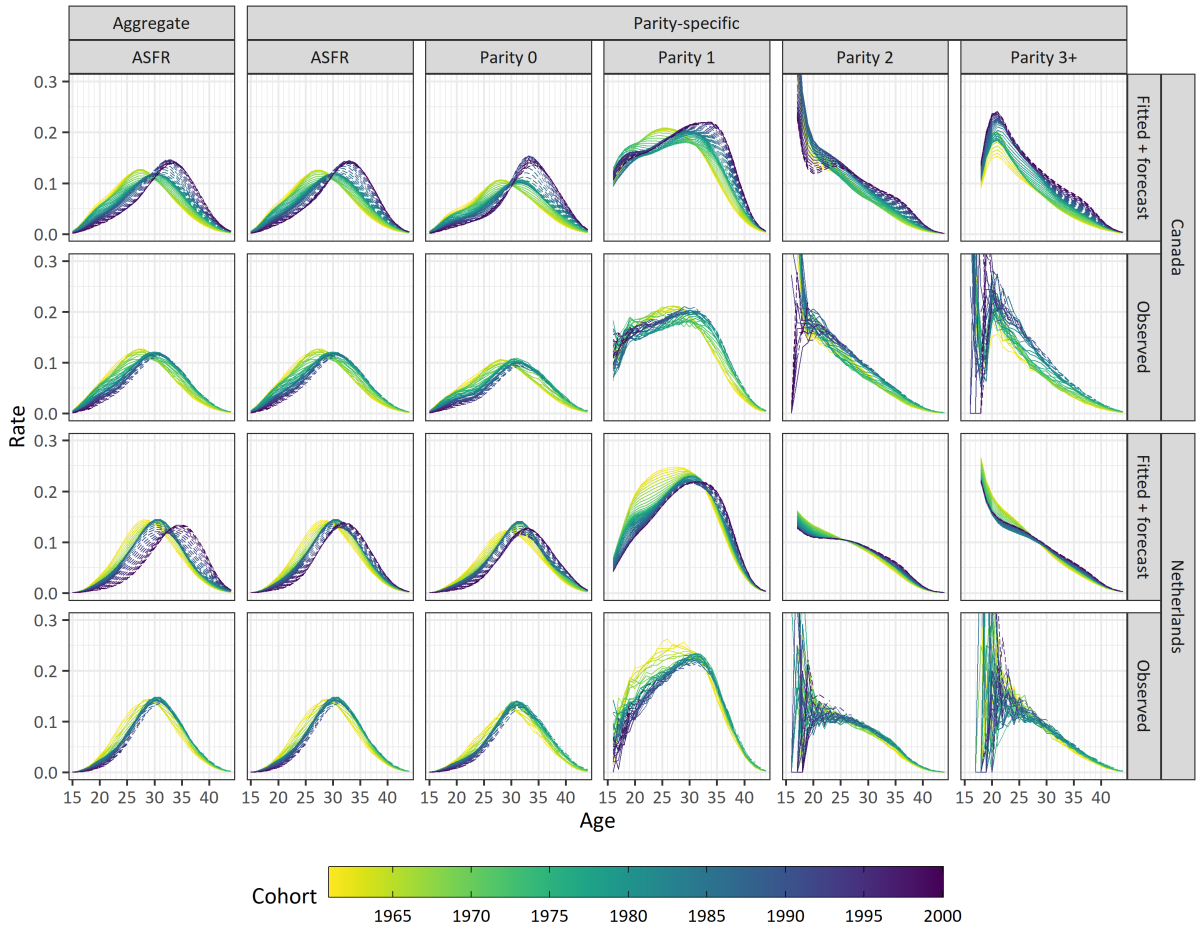


Figure D.1: Row 1: Canada posterior median ASFRs and parity-specific rates from the aggregate and parity-specific fertility projection models for $JOY = 2015$; Row 2: corresponding observed rates from Human Fertility Database (2023). Rows 3 and 4: Equivalent plots for the Netherlands. Dashed lines indicate values after the year $JOY$.