Chapter 5
Projecting Proportionate Age–Specific Fertility Rates via Bayesian Skewed Processes

Emanuele Aliverti, Daniele Durante, and Bruno Scarpa

5.1   Introduction

There is an extensive interest on models for fertility rates in statistics and demography (Hoem et al. 1981; Scarpa 2014). Several approaches have demonstrated a satisfactory fit for age–specific fertility rates via standard routine formulations such as the Hadwiger model (Hadwiger 1940), the Gompertz model (Murphy and Nagnur 1972) and the Gamma model (Hoem et al. 1981). These analyses have led to important insights on relevant population patterns and on how education, fertility control and marriage practices have played a key role in determining the shapes of fertility curves (Rindfuss et al. 1996; Billari and Kohler 2004). However, recent studies on developed countries have observed that age–specific fertility rates require more flexible models which are able to capture both symmetric and asymmetric patterns (Mazzuco and Scarpa 2015; Peristera and Kostaki 2007; Chandola et al. 1999).

E. Aliverti (✉)
Department of Statistical Sciences, University of Padova, Padova, Italy
e-mail: aliverti@stat.unipd.it

D. Durante
Department of Decision Sciences and Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy
e-mail: daniele.durante@unibocconi.it

B. Scarpa
Department of Statistical Sciences and Department of Mathematics "Tullio Levi-Civita", University of Padova, Padova, Italy
e-mail: scarpa@stat.unipd.it

The above findings have stimulated new research questions and the development of more flexible statistical models which are able to adequately describe these non–standard shapes and characterize their dynamic evolution. Recent approaches include models relying on mixtures of symmetric distributions (Peristera and Kostaki 2007; Bermúdez et al. 2012), smoothing splines (Schmertmann 2003) and skewed distributions (Mazzuco and Scarpa 2015), with some parametric assumptions sometimes relaxed via nonparametric alternatives (Kostaki et al. 2009; Canale and Scarpa 2015). Clearly, the improved fit of these models comes at a price in terms of interpretability. For example, smoothing splines generally provide an excellent fit, but interpretation of the parameters is difficult (Hoem et al. 1981; Peristera and Kostaki 2007). Besides this, few attention has been devoted to forecasts. In fact, until 2011, most demographic projections were based on deterministic predictions of fertility rates produced by the World Population Prospect report of the United Nations (Lutz and Samir 2010). In these forecasts, potential variability is only included via low and high fertility scenarios obtained by manipulating the Total Fertility Rates' (TFR) projections (Alkema et al. 2011; Raftery et al. 2013). However, such an approach does not properly quantify predictive uncertainty, and the extent to which these low or high level scenarios are realistic is still an open question (Alkema et al. 2011).

More recently, United Nations and other agencies have started moving to probabilistic approaches for population forecasting. However, in most of the cases, only summary indicators such as TFR and life expectancy at birth (e_0) are stochastically projected. This means that, in a cohort–component perspective, these indicators have to be converted into age–specific—fertility or mortality—rates, in order to project the population counts. A naive solution would be to assume a standard age schedule that is applied for every year, but this strategy has two major drawbacks. First, it has been shown that mean, variance and even skewness of the age schedule of fertility are not fixed, but time–varying (Mazzuco and Scarpa 2015; Keilman and Pham 2000). Second, in this way a component of uncertainty is missing, whereas we would like to incorporate in our forecasts the uncertainty due to varying age schedules (Ediev 2013).

Motivated by the above considerations, recent approaches for probabilistic forecasting have focused on Bayesian hierarchical models (Alkema et al. 2011; Raftery et al. 2013, 2014; Ševčíková et al. 2016). These methods aim at projecting TFR and life expectancies at birth, while deriving related quantities—such as the age–specific fertility rates—via Markov chain Monte Carlo (MCMC) (Alkema et al. 2011; Ševčíková et al. 2016). Indeed, Bayesian models facilitate probabilistic forecasts via posterior predictive distributions, and incorporate uncertainty in estimation and prediction. For high and medium fertility countries, the proposal to project age schedules of fertility consists in a linear interpolation among a starting fertility age pattern and a target model chosen among different possible age schedules of fertility (Ševčíková et al. 2016). For low fertility countries, it is assumed that a target model will be reached by 2025–2030. Such assumptions are coherent with the United Nations population forecasts, in which both fertility and mortality levels of all countries are assumed to eventually converge to a global value. However, in

single population forecasting settings, it is preferable to use a more data–driven approach, without considering target schedules.

In this contribution we propose a Bayesian dynamic model for *proportionate age–specific fertility rates* (PASFRS)—i.e. the age–specific fertility rates divided by the TFR to obtain values summing up to one. Our goal is to provide a parsimonious, yet flexible, representation of PASFRS based on densities of skew–normal variables with moments evolving in time via flexible Gaussian process priors. Such a specification allows to model proportionate age–specific fertility rates across different years via a skewed process, and to characterize their temporal evolution flexibly, while quantifying the uncertainty in estimation and prediction. We refer to our Bayesian skewed processes as BSP. Unlike available Bayesian solutions, BSP provides a direct model for PASFRS, thus allowing to define the entire distribution of these quantities across all the ages, while characterizing its dynamic evolution over time.

5.2 Bayesian Skewed Process

A fertility curve defines the fertility rates at each age or age group y—i.e. the annual number of births to women of a specified age or age group y per woman in that age group. Following Hoem et al. (1981), such a function may be written as

$$g(y; R, \theta_2, \ldots, \theta_r) = R \cdot f(y; \theta_2, \ldots, \theta_r), \tag{5.1}$$

where R is the TFR, i.e. the expected number of children born per woman in her fertile window, and $f(\cdot; \theta_2, \ldots, \theta_r)$ is a density function characterizing the PASFRS. Such a choice ensures that for any set of valid parameters $(\theta_2, \ldots, \theta_r)$ the PASFRS are positive and integrate to one without further constraints on the $r - 1$ parameters and in the observed data (Bergeron-Boucher et al. 2017), thus facilitating estimation and inference. In this contribution, our main goal is to provide flexible, yet interpretable, models and inference procedures for $f(\cdot; \theta_2, \ldots, \theta_r)$ rather than $g(\cdot; R, \theta_2, \ldots, \theta_r)$. We shall, however, emphasize that when the interest is on learning the total fertility curve in equation (5.1), our approach can be easily combined with a Bayesian updating for the posterior distribution of R, thereby inducing a full posterior on $g(\cdot; R, \theta_2, \ldots, \theta_r)$.

Several specifications of $f(\cdot; \theta_2, \ldots, \theta_r)$ are illustrated in Hoem et al. (1981) leveraging the Hadwiger (inverse–Gaussian), Gamma, Beta, Coale–Trussell, Brass and Gompertz densities. Other formulations have been suggested by Peristera and Kostaki (2007), Bermúdez et al. (2012), Schmertmann (2003), and Chandola et al. (1999). More recently, Mazzuco and Scarpa (2015) proposed to use a generalization of the normal distribution, known as skew–normal, to fit age–specific fertility rates. Such a distribution is denoted as $y \sim \text{SN}(\xi, \omega, \alpha)$ and has density function equal to

$$f(y; \xi, \omega, \alpha) = 2\omega^{-1}\phi[\omega^{-1}(y - \xi)]\Phi[\alpha\omega^{-1}(y - \xi)], \tag{5.2}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and cumulative distribution function of the standard normal distribution, respectively, while $\xi \in \mathbb{R}$, $\omega \in \mathbb{R}_+$ and $\alpha \in \mathbb{R}$ represent the location, scale and skewness parameters. While direct interpretation of these parameters might be difficult, the first two moments of the skew-normal distribution have simple analytical expressions. In particular, the expectation of the random variable y is

$$\mathrm{E}(y) = \xi + \omega\delta\sqrt{2/\pi}, \tag{5.3}$$

whereas its variance is

$$\mathrm{var}(y) = \omega^2(1 - 2\delta^2/\pi), \tag{5.4}$$

with $\delta = \alpha(1+\alpha^2)^{-1/2}$ (Azzalini and Capitanio 2013). The properties of the skew–normal in equation (5.2) have been studied by Azzalini (1985) and other authors. One interesting feature is that, when $\alpha = 0$, equation (5.2) reduces to the density of a normal, thus allowing inclusion of both asymmetric ($\alpha \neq 0$) and symmetric ($\alpha = 0$) shapes in modeling the PASFRS via (5.2).[1] Indeed, Mazzuco and Scarpa (2015) have shown that in Italy the fertility schedule function has moved from a skewed to a symmetric shape.

Motivated by these considerations, we model PASFRS via a time–varying version of (5.2) and, taking a Bayesian approach, we allow flexible changes in this curve via suitable priors for its dynamic parameters ξ_t, ω_t and α_t. In this way, we define a new Bayesian skewed process, which allows forecasting of future PASFRS. As already mentioned, there is an abundance of models for forecasting of TFRS, while a coherent approach for PASFRS is still lacking. The method proposed in this chapter takes a first step toward addressing this important goal.

5.2.1 Model Specification

For every year $t = 1, \ldots, T$ and mother $i = 1, \ldots, n_t$, our data consist in artificial random samples of n_t women at the age of childbirth, where y_{it} represents the age of the i-th mother in year t. These artificial data are obtained by sampling, for each year t, a total of n_t age values from a discrete random variable with the proportionate age–specific fertility rates as probabilities, thereby obtaining a synthetic cohort generated by the dynamic PASFRS. As clarified in Sect. 5.3, the choice to rely on synthetic data is due to the computational intractability that would arise under BSP if the focus were on the full population. In fact, Bayesian inference under BSP requires sampling methods for multivariate truncated normals of dimension

---

[1]Common specifications, such as Hadwiger, Gamma, Gompertz, cannot assume a symmetric shape.

$\sum_{t=1}^{T} n_t$. Nonetheless, we will consider a sufficiently large n_t to allow efficient learning of the model parameters.

To further motivate the above construction, suppose that interest is on estimating how a fixed number of births n_t is distributed across the different ages, while the total intensity of fertility is kept fixed. This problem can be tackled via a multinomial distribution with cell counts corresponding to the number of mothers with a specific age, and a probability of falling in the k-th class (age equal to y_k) being proportional to $f(y_k; \xi, \omega, \alpha)$—the PASFR. Sampling from this hypothetical multinomial model is statistically equivalent to sampling from the discrete distribution mentioned above. Hence, the observed rates are effectively treated as data by our approach, and the uncertainty in the estimated parameters regulating the shape of PASFR will be fully incorporated via the posterior distribution, under our Bayesian approach to inference.

The aforementioned procedure allows to define a genuine likelihood based on a skew-normal specification. In fact, recalling the discussion in Sect. 5.2, we assume that each y_{it} has a skew–normal distribution with location ξ_t, scale ω_t and skewness parameter α_t, thereby obtaining

$$(y_{it} \mid \xi_t, \omega_t, \alpha_t) \sim \text{SN}(\xi_t, \omega_t, \alpha_t), \tag{5.5}$$

independently for each $i = 1, \ldots, n_t$ and $t = 1, \ldots, T$. Following a Bayesian approach to inference, we specify prior distributions for the parameters $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)^\mathsf{T} \in \mathbb{R}^T$, $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)^\mathsf{T} \in \mathbb{R}_+^T$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)^\mathsf{T} \in \mathbb{R}^T$ in (5.5) to incorporate temporal interdependence across the fertility rates observed in the different years. Such priors can be seen as distributions quantifying experts' uncertainty in the model parameters, and the goal of Bayesian learning is to update such quantities in the light of the observed data to obtain a posterior distribution which is used for inference.

To address the above goal, while maintaining computational tractability, we specify independent Gaussian process (GP) priors (Rasmussen and Williams 2006), with squared exponential covariance functions, for the location and skewness parameters, thus obtaining

$$\boldsymbol{\xi} = (\xi_1, \ldots, \xi_T)^\mathsf{T} \sim \text{N}_T(\boldsymbol{\mu_\xi}, \boldsymbol{\Sigma_\xi}) \text{ and } \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_T)^\mathsf{T} \sim \text{N}_T(\boldsymbol{\mu_\alpha}, \boldsymbol{\Sigma_\alpha}), \tag{5.6}$$

for any time grid $t = 1, \ldots, T$, where $[\boldsymbol{\mu_\xi}]_j = m_\xi(t_j)$, $[\boldsymbol{\Sigma_\xi}]_{jl} = \exp(-\kappa_\xi ||t_j - t_l||_2^2)$, $[\boldsymbol{\mu_\alpha}]_j = m_\alpha(t_j)$, and $[\boldsymbol{\Sigma_\alpha}]_{jl} = \exp(-\kappa_\alpha ||t_j - t_l||_2^2)$. Note also that $m_\xi(\cdot)$ and $m_\alpha(\cdot)$ denote pre–selected GP mean functions, whereas the covariances in $\boldsymbol{\Sigma_\xi}$ and $\boldsymbol{\Sigma_\alpha}$ are specified so as to decrease with the time lag. Refer to Rasmussen and Williams (2006) for additional details on Gaussian processes. The priors for the square of the scale parameters ω_t, $t = 1, \ldots, T$ are instead specified as independent Inverse–Gamma distributions

$$\omega_t^2 \sim \text{Inv–Gamma}(a_\omega, b_\omega), \quad t = 1, \ldots, T. \tag{5.7}$$

Although the prior in equation (5.7) does not allow for explicit temporal dependence across different values of the scale parameters, we stress that the skewness parameters α_t and the locations ξ_t have a central role in controlling the mean and the variance of the random variable y_{it}, as outlined in equations (5.3) and (5.4). Hence, the GP priors in (5.6) induce temporal dependence also in the expectation and in the variance of the variable y_{it}, and are arguably sufficient to characterize its main dynamic evolution.

5.2.2 Joint Likelihood and Posterior Distribution for α

Assume, for the moment, that the parameters ξ_t and ω_t are fixed at $\xi_t = 0$ and $\omega_t = 1$ for each $t = 1, \ldots, T$. The focus of this simplifying assumption is to illustrate the key steps to obtain the joint posterior distribution for the vector $\boldsymbol{\alpha}$ induced by a Gaussian prior and the model (5.5). Recently, Canale et al. (2016) showed that the posterior distribution from a Gaussian prior combined with a skew–normal likelihood is an unified skew–normal (SUN) distribution, which is a family of distributions that includes the skew–normal one (Arellano-Valle and Azzalini 2006). In the following paragraph, we illustrate the multivariate extension of such a result, focusing on the analytical form of the resulting posterior distribution and its associated parameters.

For simplicity of exposition suppose, without loss of generality, that $n_t = n$ for $t = 1, \ldots, T$ and let $\mathbf{y}_t = (y_{1t}, \ldots, y_{nt})^\mathsf{T}$. Then, incorporating the above assumptions, the likelihood for $\boldsymbol{\alpha}$ induced by model (5.5) is

$$L(\boldsymbol{\alpha}) = \prod_{t=1}^{T} \prod_{i=1}^{n} 2\phi(y_{it}) \Phi(\alpha_t y_{it}) \propto \prod_{t=1}^{T} \Phi_n(\alpha_t \mathbf{y}_t; \mathbf{I}_n) = \Phi_{nT}(\mathbf{Y}\boldsymbol{\alpha}; \mathbf{I}_{nT}), \quad (5.8)$$

where $\Phi_{nT}(\mathbf{Y}\boldsymbol{\alpha}; \mathbf{I}_{nT})$ is the cumulative distribution function of a nT–variate Gaussian with identity covariance matrix evaluated at $\mathbf{Y}\boldsymbol{\alpha}$. In (5.8), $\mathbf{Y}$ corresponds to a data matrix of dimension $nT \times T$ such that $\mathbf{Y}\boldsymbol{\alpha} = (y_{11}\alpha_1, y_{21}\alpha_1, \ldots, y_{it}\alpha_t, \ldots, y_{nT}\alpha_T)^\mathsf{T}$. Such a representation is useful to express the argument of $\Phi_{nT}(\cdot)$ in equation (5.8) as a linear term in $\boldsymbol{\alpha}$. The posterior distribution for $\boldsymbol{\alpha}$ is obtained combining the skew–normal likelihood in equation (5.8) with the Gaussian process prior. Formally, by applying the Bayes rule, we obtain $f(\boldsymbol{\alpha} \mid \mathbf{y}_1, \ldots, \mathbf{y}_T) \propto \phi_T(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}; \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}) \Phi_{nT}(\mathbf{Y}\boldsymbol{\alpha}; \mathbf{I}_{nT})$, with

$$\phi_T(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}; \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}) \Phi_{nT}(\mathbf{Y}\boldsymbol{\alpha}; \mathbf{I}_{nT})$$
$$= \phi_T(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}; \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}) \Phi_{nT}(\mathbf{s}^{-1}\mathbf{Y}\boldsymbol{\mu}_{\boldsymbol{\alpha}} + \mathbf{s}^{-1}\mathbf{Y}(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}); \mathbf{s}^{-2}), \quad (5.9)$$

where $\mathbf{s} = \mathrm{diag}[(\mathbf{Y}_1^\mathsf{T} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \mathbf{Y}_1 + 1)^{1/2}, \ldots, (\mathbf{Y}_{nT}^\mathsf{T} \boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \mathbf{Y}_{nT} + 1)^{1/2}]$. Recalling recent results in Durante (2019), equation (5.9) corresponds to the kernel of a SUN distribution. Specifically,

$$(\boldsymbol{\alpha} \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_T) \sim \text{SUN}_{T,nT}(\boldsymbol{\mu_\alpha}, \boldsymbol{\Sigma_\alpha}, \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} \boldsymbol{\sigma}_\alpha \mathbf{Y}^\mathsf{T} \mathbf{s}^{-1}, \mathbf{s}^{-1} \mathbf{Y} \boldsymbol{\mu_\alpha}, \mathbf{s}^{-1}(\mathbf{Y}\boldsymbol{\Sigma_\alpha}\mathbf{Y}^\mathsf{T} + \mathbf{I}_{nT})\mathbf{s}^{-1}),$$
$$(5.10)$$

with $\bar{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}}$ a full-rank correlation matrix such that $\boldsymbol{\Sigma_\alpha} = \boldsymbol{\sigma}_\alpha \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} \boldsymbol{\sigma}_\alpha$. Complete algebraic derivations to obtain the above result are extensively described in Durante (2019, Theorem 1).

5.3 Posterior Computation

In the general setting, where $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are unknown, the joint posterior for $(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\alpha})$ does not admit a closed–form expression, and, hence, it is necessary to rely on MCMC methods. Here, we propose a Metropolis–within–Gibbs algorithm which combines the results in the previous section and other SUN properties to iteratively sample values from the full–conditionals of $\boldsymbol{\xi}$, $\boldsymbol{\omega}$ and $\boldsymbol{\alpha}$. In doing so, MCMC builds on a Markov chain which produces realizations from the posterior distribution $f(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\alpha} \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_T)$ after convergence (Gelfand and Smith 1990). A sufficiently large sample of values simulated from the joint posterior distribution is then used to make inference on functionals of the parameters via standard Monte Carlo integration (Casella and George 1992).

Given the current values of $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$, the full–conditional for $\boldsymbol{\alpha}$ can be obtained via minor modifications of the results in the previous section. Indeed, if ξ_t and ω_t are known, the contribution of the generic y_{it} to the likelihood for $\boldsymbol{\alpha}$ is proportional to $\Phi[\alpha_t(y_{it} - \xi_t)/\omega_t] = \Phi(\alpha_t \bar{y}_{it})$. Hence, replacing each y_{it} with $\bar{y}_{it} = (y_{it} - \xi_t)/\omega_t$ in (5.8)–(5.9), the SUN full–conditional for $(\boldsymbol{\alpha} \,|\, \mathbf{y}_1, \ldots, \mathbf{y}_T, \boldsymbol{\xi}, \boldsymbol{\omega}) = (\boldsymbol{\alpha} \,|\, \bar{\mathbf{y}}_1, \ldots, \bar{\mathbf{y}}_T)$ has the same form of (5.10), with $\mathbf{Y}$ replaced by $\bar{\mathbf{Y}}$. To effectively use this result in a Metropolis–within–Gibbs algorithm, it is necessary to simulate from the distribution defined in equation (5.10). The following Lemma describes a constructive procedure for simulating from a SUN. See Azzalini and Capitanio (2013) and Durante (2019) for a formal proof.

Lemma 1 *If the full-conditional distribution for the skewness parameters comprising $\boldsymbol{\alpha}$ is $(\boldsymbol{\alpha} \,|\, -) \sim \text{SUN}_{T,nT}(\boldsymbol{\mu_\alpha}, \boldsymbol{\Sigma_\alpha}, \bar{\boldsymbol{\Sigma}}_{\boldsymbol{\alpha}} \boldsymbol{\sigma}_\alpha \bar{\mathbf{Y}}^\mathsf{T} \bar{\mathbf{s}}^{-1}, \bar{\mathbf{s}}^{-1} \bar{\mathbf{Y}} \boldsymbol{\mu_\alpha}, \bar{\mathbf{s}}^{-1}(\bar{\mathbf{Y}}\boldsymbol{\Sigma_\alpha}\bar{\mathbf{Y}}^\mathsf{T} + \mathbf{I}_{nT})\bar{\mathbf{s}}^{-1})$, then*

$$(\boldsymbol{\alpha} \,|\, -) \stackrel{\mathrm{d}}{=} \boldsymbol{\mu_\alpha} + \boldsymbol{\Sigma_\alpha}[\mathbf{V_0} + \bar{\mathbf{Y}}^\mathsf{T}(\bar{\mathbf{Y}}\boldsymbol{\Sigma_\alpha}\bar{\mathbf{Y}}^\mathsf{T} + \mathbf{I}_{nT})^{-1}\bar{\mathbf{s}}\mathbf{V_1}],$$

with $\mathbf{V_0} \sim \mathrm{N}_T(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}^{-1} - \bar{\mathbf{Y}}^\mathsf{T}(\bar{\mathbf{Y}}\boldsymbol{\Sigma_\alpha}\bar{\mathbf{Y}}^\mathsf{T} + \mathbf{I}_{nT})^{-1}\bar{\mathbf{Y}})$ denoting a multivariate Gaussian and $\mathbf{V_1} \sim \mathrm{TN}_{nT}[-\bar{\mathbf{s}}^{-1}\bar{\mathbf{Y}}\boldsymbol{\mu_\alpha}, \mathbf{0}, \bar{\mathbf{s}}^{-1}(\bar{\mathbf{Y}}\boldsymbol{\Sigma_\alpha}\bar{\mathbf{Y}}^\mathsf{T} + \mathbf{I}_{nT})\bar{\mathbf{s}}^{-1}]$ corresponding to a $nT-$ variate Gaussian distribution with zero mean, covariance matrix $\bar{\mathbf{s}}^{-1}(\bar{\mathbf{Y}}\boldsymbol{\Sigma_\alpha}\bar{\mathbf{Y}}^\mathsf{T} + \mathbf{I}_{nT})\bar{\mathbf{s}}^{-1}$, and truncation below $-\bar{\mathbf{s}}^{-1}\bar{\mathbf{Y}}\boldsymbol{\mu_\alpha}$.

Simulation from the SUN full–conditional distribution defined in Lemma 1 requires to sample from a nT–variate truncated Gaussian, which is very demanding for large values of nT. Recent developments in this direction involve slice sampling

(Liechty and Lu 2010) or Hamiltonian Monte Carlo (Pakman and Paninski 2014), with minimax tilting being the most efficient routine in moderate dimensions (Botev 2017). Despite these improved approaches, independent sampling from multivariate truncated Gaussian vectors is still unpractical when the dimension is greater than a few hundreds (Botev 2017). In these situations, Gibbs–sampling from sub–blocks of V_1 provides an appealing solution (Chopin 2011), since multivariate truncated Gaussians are closed under conditioning (Horrace 2005), and sampling of sub–blocks of moderate size—e.g., around 50—can be done efficiently via minimax tilting (Botev 2017).

To obtain conjugacy in the full–conditional for the locations ξ, we rely instead on the additive representation of the skew–normal distribution. Indeed, as a particular case of Lemma 1, we recall that if $z \sim N(0, 1)$ and $w \sim N(0, 1)$ independently, then $y = \xi + \omega[\delta|z| + (1 - \delta^2)^{1/2}w] \sim SN(\xi, \omega, \alpha)$, with $\alpha = \delta(1 - \delta^2)^{-1/2}$. Hence, it is possible to recast the skew–normal likelihood in terms of a conditional Gaussian likelihood, given a set of latent variables z_{it}. More specifically, if y_{it} is marginally distributed as a $SN(\xi_t, \omega_t, \alpha_t)$, by introducing latent observations z_{it}, we obtain

$$z_{it} \sim TN_1(0, 0, 1) \text{ and } (y_{it} \mid z_{it}) \sim N[\xi_t + \omega_t\delta_t z_{it}, \omega_t^2(1 - \delta_t^2)],$$

with $\delta_t = \alpha_t(1 + \alpha_t^2)^{-1/2}$, thereby allowing conditionally conjugate updates for ξ and a simple Metropolis step for ω. The complete Metropolis–within–Gibbs sampler algorithm for posterior computation iterates among the steps outlined below. Refer to the Appendix for detailed derivations.

[1] Latent variables z: Update every latent variable z_{it} from the truncated Gaussian full–conditional distribution

$$(z_{it} \mid -) \sim TN_1[0, \delta_t(y_{it} - \xi_t)/\omega_t, (1 - \delta_t^2)], \quad i = 1, \ldots, n, \quad t = 1, \ldots, T.$$

[2] Location vector ξ: Given the current value of the latent variables z_{it} and of the parameters α_t and ω_t, we can recast our formulation as a regression model for transformed Gaussian data $y_{it}^* = y_{it} - \omega_t\delta_t z_{it}$, $i = 1, \ldots, n$, $t = 1, \ldots, T$. Hence, letting $\bar{y}^* = (n^{-1}\sum_{i=1}^n y_{i1}^*, \ldots, n^{-1}\sum_{i=1}^n y_{iT}^*)^\mathsf{T}$, the full–conditional for ξ can be derived via Gaussian–Gaussian conjugacy and coincides with

$$(\xi \mid -) \sim N_T[(\Sigma_\xi^{-1} + nV_\xi)^{-1}(\Sigma_\xi^{-1}\mu_\xi + nV_\xi\bar{y}^*), (\Sigma_\xi^{-1} + nV_\xi)^{-1}],$$

where $V_\xi = \text{diag}[1/\omega_1^2(1 - \delta_1^2), \ldots, 1/\omega_T^2(1 - \delta_T^2)]$.

[3] Scale vector ω: For every time $t = 1, \ldots, T$, update ω_t independently with a Metropolis–Hasting step.

[4] Skewness vector α: Update α from the full–conditional SUN distribution, replacing y_{it} with the transformed value $(y_{it} - \xi_t)/\omega_t$ in (5.10) and using Lemma 1.

Coherently with a Bayesian specification, forecasts for years $T + 1, \ldots, T + q$ are obtained by treating the future observations $\mathbf{y}_{T+1}, \ldots, \mathbf{y}_{T+q}$ as missing data in the MCMC (Gelman et al. 2013). At each iteration, the parameters $(\xi_{T+1}, \omega_{T+1}, \alpha_{T+1}), \ldots, (\xi_{T+q}, \omega_{T+q}, \alpha_{T+q})$ are updated jointly with $(\boldsymbol{\xi}, \boldsymbol{\omega}, \boldsymbol{\alpha})$, after imputing the missing data $\mathbf{y}_{T+1}, \ldots, \mathbf{y}_{T+q}$ with values sampled from the conditional skew–normals in equation (5.5).

5.4 Forecasting Italian Fertility Rates

We apply the model defined in Sects. 5.2–5.3 to the proportionate age–specific Italian fertility rates from 1991 to 2014, creating an artificial population of $n = 500$ women for each year based on data at https://www.humanfertility.org/cgi-bin/main.php.

In performing posterior inference and forecasting, the GP priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\xi}$ have been centered around 0 and 30 respectively, setting $m_\alpha(t_j) = 0$ and $m_\xi(t_j) = 30$. These values define our prior guess on the shape of the curve and on the average age at childbirth. The prior GP covariance parameters κ_α and κ_ξ are instead fixed at 100 to induce modest dependence across years. Finally, we set $a_\omega = 10$ and $b_\omega = 300$ to obtain prior means and standard deviations for the scales around 30 and 10, respectively. These values were elicited by inspecting the variance of the historical data, and centering the priors around this value, while inducing sufficient variability to deviate from this assumption, if required. We also conducted sensitivity analyses obtaining similar results under many hyper–parameters' settings. Posterior inference relies on 5000 MCMC samples after a burn–in period of 2000. These choices were sufficient for convergence, whereas mixing was not perfect, but still satisfactory.

The focus of inference is on the time–varying mean $\xi_t + \omega_t \delta_t \sqrt{2/\pi}$, variance $\omega_t^2(1 - 2\delta_t^2/\pi)$ and skewness parameter α_t of the age at childbirth under (5.5)—with $\delta_t = \alpha_t(1 + \alpha_t^2)^{-1/2}$. The posteriors for these quantities can be easily computed from the MCMC samples of $(\xi_t, \omega_t, \alpha_t)$ and some key summaries are reported in Fig. 5.1. According to the upper panel, our empirical findings suggest that the average age at childbirth has increased in the last decades—a result which was expected and well investigated in the literature. This average age has moved from a minimum close to 28 years in 1991 to a maximum close to 31 years in 2010 and following years. The middle panel summarizes, instead, the posterior distribution for α_t, suggesting that the fertility rates have actually become symmetric in recent years and demonstrating the ability of the model to capture both symmetric and asymmetric shapes. Finally, the posterior distributions for the variance, reported in the bottom panel of Fig. 5.1, suggest a stable variability across the temporal window considered. Also these results are in line with the findings of Mazzuco and Scarpa (2015).

To validate the above results, Fig. 5.2 compares the histograms of the proportionate age–specific fertility rates, computed from the synthetic data, with the posterior distribution of $f(y_k; \xi_t, \omega_t, \alpha_t)$ in equation (5.2), for each age y_k, summarized via a pointwise posterior mean and the 95% credible intervals. Since the value of $f(y_k; \xi_t, \omega_t, \alpha_t)$ is a functional of model parameters, the posterior distribution for
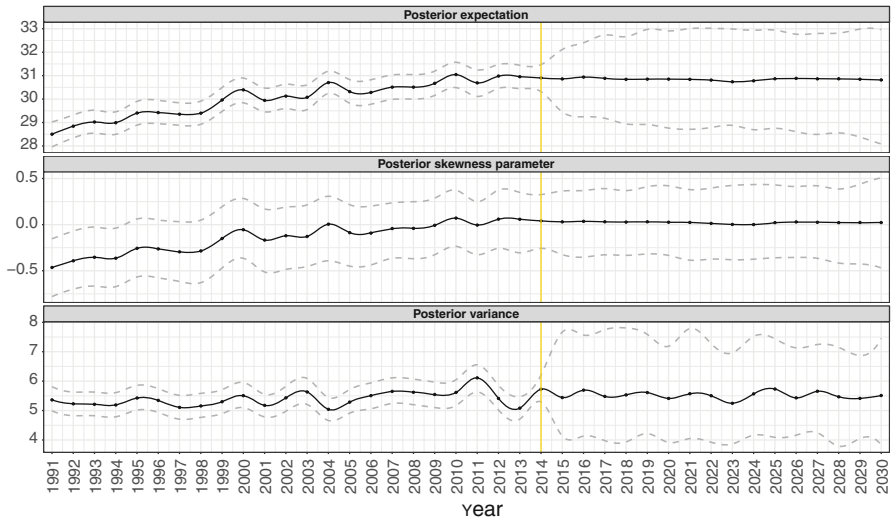
Fig. 5.1 Summaries of the posterior distribution for the mean, skewness parameter and variance of the skewed process for y_{it}. Dashed lines denote 95% credible intervals. Yellow vertical lines denote the last observed year
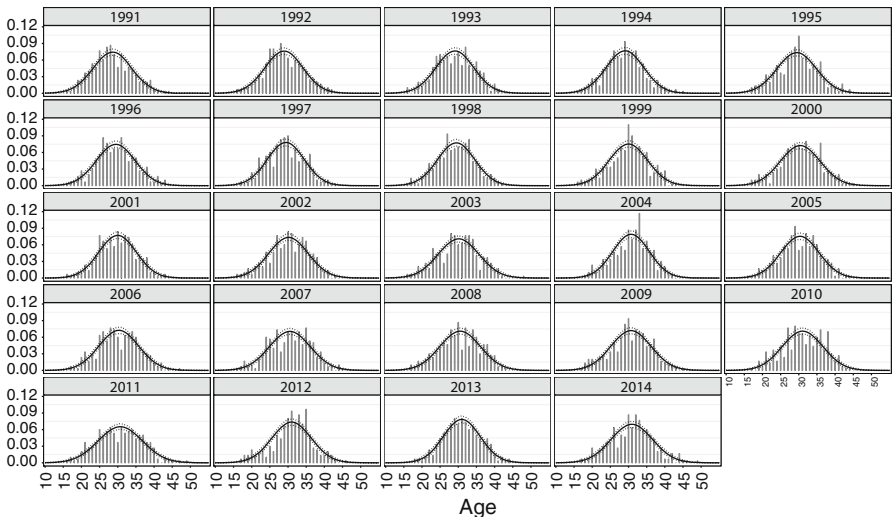

Fig. 5.2 For each year from 1991 to 2014, histograms of the proportionate age–specific fertility rates computed from the synthetic data, and summaries of the posterior distribution for $f(y_k; \xi_t, \omega_t, \alpha_t)$ in (5.2), for each age y_k. Black continuous line indicates pointwise posterior mean, while 95% credible intervals are denoted as dotted lines

$(\xi_t, \omega_t, \alpha_t)$ induces a posterior also for $f(y_k; \xi_t, \omega_t, \alpha_t)$, for each age y_k. Results suggest a satisfactory fit, with the rates arising from the artificial samples being close to the pointwise estimates. To summarize, posterior inference suggests that PASFRS have experienced a change in the last decade, which has impacted the location and shape of the curve while leaving variability stable. The goodness of fit of the proposed approach, in terms of adequacy with the empirical distribution of the artificial data, is satisfactory.

The results in terms of goodness of fit illustrated above motivate forecasts for the Italian PASFRS, producing these predictions for the 16 years after the last observed time. According to Fig. 5.1, forecasts for the posterior mean of the age at childbirth under the BSP model show a stable trend, which is coherent with the Italian fertility rates observed in the recent years. Also the forecasts for the variance and the skewness parameter of the age at childbirth are substantially stable.

We also compare our forecasting accuracy with the results from a default implementation of the approach proposed by Ševčíková et al. (2016) and available via the R library bayesPop (Ševčíková and Raftery 2016). The main routines of this library compute predictions for the TFR and life expectancies, and then obtain the cohort–specific fertility rates via post–processing of the MCMC output. We also highlight that the method available in bayesPop does not provide fertility rates for all the ages, but only for 5 years age groups. To compare these predictions with the results obtained from BSP, we represent the former as a step function with constant values within each age interval.

Results are reported in Fig. 5.3, with yellow curves referred to predictions from the BSP model and black step functions from bayesPop. The 90% credible intervals are illustrated as dotted lines. Direct comparison among the two approaches
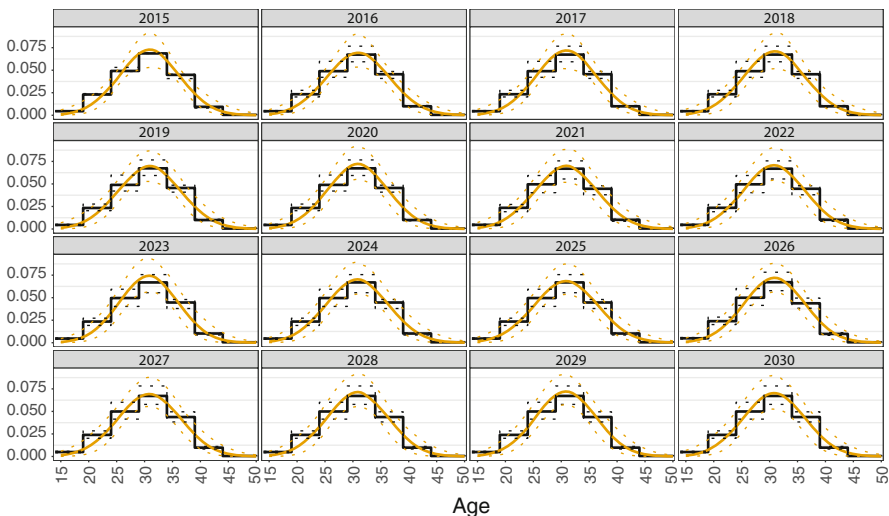
Fig. 5.3 Forecasted distribution for the BSP model (yellow) against those obtained under the package bayesPop. Dotted lines denote 90% credible intervals

suggests very similar results in terms of predicted probabilities, with both strategies assigning the highest probability of childbirth in the interval $(30 - 34]$. The credible intervals from BSP are wider than the competitor, likely due to the uncertainty in the dynamic components. This is not surprising, due to the assumptions made by Ševčíková et al. (2016) which may lead to under–coverage of the credible intervals when they are not met in practice.

5.5   Discussion

In this work we have proposed to model PASFRS via a Bayesian skewed process. Our specification incorporates symmetric and asymmetric shapes, while characterizing temporal dependence through the skew–normal parameters.

This approach takes a first step towards direct forecasting of PASFRS using Bayesian models. In facts, also Ševčíková et al. (2016) use a Bayesian framework to forecast PASFRS over time, but this is done within a hierarchical model applied to all countries which are further assumed to converge to a global pattern. The method proposed in this article provides, instead, single–country forecasts, borrowing information only from past PASFRS and not from other countries' patterns, nor from hypothetical global schedules. Results are comparable with Ševčíková et al. (2016), with a reasonably higher uncertainty of the forecasts.

Future extensions include methodological developments to allow joint modeling of multiple countries via a mixture of BSPs. This could also facilitate clustering of countries with respect to similarities in fertility patterns, thereby providing insights on important social aspects of developed countries. Also the inclusion of more complex dependence patterns among PASFRS and TFR could further improve predictions.

Another key improvement includes the reduction of the computational cost associated with posterior inference for BSP. The simulation of the nT–variate truncated Gaussian involved in the SUN can be demanding in high dimensions. An option to overcome this issue is to rely on approximate Bayesian inference.

Appendix

Here, we derive the key quantities involved in the algorithm described in Sect. 5.3.

Full conditional for z_{it}. Recall that $z_{it} \sim \text{TN}_1(0, 0, 1)$, and $(y_{it} \mid z_{it}) \sim \text{N}[\xi_t + \omega_t \delta_t z_{it}, \omega_t^2(1 - \delta_t^2)]$. Hence, the full conditional for z_{it} is proportional to

$$f(z_{it})f(y_{it}|z_{it}) \propto \mathbf{1}(z_{it} > 0) \exp(-0.5z_{it}^2) \exp[-(y_{it} - \xi_t - \omega_t \delta_t z_{it})^2/2\omega_t^2(1 - \delta_t^2)].$$

Focusing on the two terms in the exponents and applying classical Gaussian results, we obtain the kernel of a normal distribution with mean $\delta_t(y_{it} - \xi_t)/\omega_t$ and variance $(1 - \delta_t^2)$. Including the indicator function within such a kernel, we obtain

$$f(z_{it} \mid -) \propto \exp\left[-\frac{1}{2(1 - \delta_t^2)}\left(z_{it} - \frac{\delta_t(y_{it} - \xi_t)}{\omega_t}\right)^2\right] \mathbf{1}(z_{it} > 0).$$

Hence $(z_{it} \mid -) \sim \mathrm{TN}_1[0, \delta_t(y_{it} - \xi_t)/\omega_t, (1 - \delta_t^2)]$, for $i = 1, \dots, n, t = 1, \dots, T$.

Full conditional for $\boldsymbol{\xi}$. Recall that $y_{it}^* = y_{it} - \omega_t \delta_t z_{it}$ and let $\mathbf{y}_i^* = (y_{i1}^*, \dots, y_{iT}^*)^\mathsf{T}$ denote the T-dimensional vector of scaled observations. Since $(\mathbf{y}_i^* \mid -) \sim \mathrm{N}(\boldsymbol{\xi}, \boldsymbol{V}_{\boldsymbol{\xi}}^{-1})$, with $\boldsymbol{V}_{\boldsymbol{\xi}} = \mathrm{diag}[1/\omega_1^2(1 - \delta_1^2), \dots, 1/\omega_T^2(1 - \delta_T^2)]$, and $\boldsymbol{\xi} \sim \mathrm{N}_T(\boldsymbol{\mu_\xi}, \boldsymbol{\Sigma_\xi})$, by Gaussian–Gaussian conjugacy we obtain

$$(\boldsymbol{\xi} \mid -) \sim \mathrm{N}_T(\mathbf{S}_\xi^{-1}\mathbf{m}_\xi, \mathbf{S}_\xi^{-1}), \qquad \mathbf{S}_\xi = \boldsymbol{\Sigma_\xi}^{-1} + n\boldsymbol{V}_{\boldsymbol{\xi}}, \quad \mathbf{m}_\xi = \boldsymbol{\Sigma_\xi}^{-1}\boldsymbol{\mu_\xi} + n\boldsymbol{V}_{\boldsymbol{\xi}}\bar{\mathbf{y}}^*,$$

with $\bar{\mathbf{y}}^* = (n^{-1}\sum_{i=1}^n y_{i1}^*, \dots, n^{-1}\sum_{i=1}^n y_{iT}^*)^\mathsf{T}$.

References

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T., & Heilig, G.K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography, 48*(3), 815–839.

Arellano-Valle, R. B., & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics, 33*(3), 561–574.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics, 12*, 171–178.

Azzalini, A., & Capitanio, A. (2013). *The skew-normal and related families*. Cambridge: Cambridge University Press.

Bergeron-Boucher, M.-P., Canudas-Romo, V., Oeppen, J., & Vaupel, J. W. (2017). Coherent forecasts of mortality with compositional data analysis. *Demographic Research, 37*, 527–566.

Bermúdez, S., Blanquero, R., Hernández, J. A., & Planelles, J. (2012). A new parametric model for fitting fertility curves. *Population Studies, 66*(3), 297–310.

Billari, F., & Kohler, H. (2004). Patterns of low and lowest-low fertility in Europe. *Population Studies, 58*(2), 161–176.

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79*(1), 125–148.

Canale, A., & Scarpa, B. (2015). Age-specific probability of childbirth. Smoothing via Bayesian nonparametric mixture of rounded kernels. *Statistica, 75*(1), 101–110.

Canale, A., Kenne Pagui, C. E., & Scarpa, B. (2016). Bayesian modeling of university first-year students' grades after placement test. *Journal of Applied Statistics, 43*(16), 3015–3029.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*(3), 167–174.

Chandola, T., Coleman, D. A., & Hiorns, R. W. (1999). Recent European fertility patterns: Fitting curves to distorted distributions. *Population Studies, 53*(3), 317–329.

Chopin, N. (2011). Fast simulation of truncated Gaussian distributions. *Statistics and Computing, 21*(2), 275–288.

Durante, D. (2019). Conjugate Bayes for probit regression via unified skew-normal distributions. *Biometrika, 106*(4), 765–779.

Ediev, D. M. (2013). Comparative importance of the fertility model, the total fertility, the mean age and the standard deviation of age at childbearing in population projections. In *Meeting of the International Union for the Scientific Study of Population, Busan. Presented at the Meeting of the International Union for the Scientific Study of Population, Busan.*

Gelfand, A. E., & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*(410), 398–409.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. Boca Raton: Chapman and Hall/CRC.

Hadwiger, H. (1940). Eine analytische reproduktionsfunktion für biologische gesamtheiten. *Scandinavian Actuarial Journal, 1940*(3–4), 101–113.

Hoem, J. M., Madsen, D., Nielsen, J. L., Ohlsen, E., Hansen, H. O., & Rennermalm, B. (1981). Experiments in modelling recent Danish fertility curves. *Demography, 18*(2), 231–244.

Horrace, W. C. (2005). Some results on the multivariate truncated normal distribution. *Journal of Multivariate Analysis, 94*(1), 209–221.

Keilman, N., & Pham, D. Q. (2000). Predictive intervals for age-specific fertility. *European Journal of Population/Revue Européenne de Démographie, 16*(1), 41–65.

Kostaki, A., Moguerza, J. M., Olivares, A., & Psarakis, S. (2009). Graduating the age-specific fertility pattern using support vector machines. *Demographic Research, 20*, 599–622.

Liechty, M. W., & Lu, J. (2010). Multivariate normal slice sampling. *Journal of Computational and Graphical Statistics, 19*(2), 281–294.

Lutz, W., & Samir, K. C. (2010). Dimensions of global population projections: What do we know about future population trends and structures? *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1554), 2779–2791.

Mazzuco, S., & Scarpa, B. (2015). Fitting age-specific fertility rates by a flexible generalized skew normal probability density function. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 178*(1), 187–203.

Murphy, E. M., & Nagnur, D. N. (1972). A Gompertz fit that fits: Applications to Canadian fertility patterns. *Demography, 9*(1), 35–50.

Pakman, A., & Paninski, L. (2014). Exact Hamiltonian Monte Carlo for truncated multivariate Gaussians. *Journal of Computational and Graphical Statistics, 23*(2), 518–542.

Peristera, P., & Kostaki, A. (2007). Modeling fertility in modern populations. *Demographic Research, 16*, 141–194.

Raftery, A. E., Chunn, J. L., Gerland, P., & Ševčíková, H. (2013). Bayesian probabilistic projections of life expectancy for all countries. *Demography, 50*(3), 777–801.

Raftery, A. E., Alkema, L., & Gerland, P. (2014). Bayesian population projections for the United Nations. *Statistical Science, 29*(1), 58.

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian Processes for Machine Learning*. Cambridge: MIT.

Rindfuss, R. R., Morgan, P. S., & Offutt, K. (1996). Education and the changing age pattern of American fertility: 1963–1989. *Demography, 33*(3), 277–290.

Scarpa, B. (2014). Probabilistic and statistical models for conception. *Wiley StatsRef: Statistics Reference Online*.

Schmertmann, C. (2003). A system of model fertility schedules with graphically intuitive parameters. *Demographic Research, 9*, 81–110.

Ševčíková, H., & Raftery, A. E. (2016). Bayespop: Probabilistic population projections. *Journal of Statistical Software, 75*, 1–29.

Ševčíková, H., Li, N., Kantorová, V., Gerland, P., & Raftery, A. E. (2016). Age-specific mortality and fertility rates for probabilistic population projections. In *Dynamic demographic analysis* (pp. 285–310). Cham: Springer.