

Distr.: General
08 November 2019

English only

Economic Commission for Europe

Conference of European Statisticians

Work Session on Demographic Projections

Belgrade, 25–27 November 2019

Item No. 6 of the provisional agenda

Assumptions on fertility

Comparing fertility forecasting methods: how do parametric mixture models perform?¹

Note by Jason Hilton¹, Erengul Dodd¹, Jonathan J. Forster², Peter W.F. Smith¹, Jakub Bijak¹

¹ Centre for Population Change, University of Southampton, ² Department of Statistics, University of Warwick

Summary

Recent comprehensive work by Bohk-Ewald et al. (2018) has found that few fertility forecasting methods perform better than the naive ‘freeze rates’ method, whereby the last observed fertility rate for each age is taken as the forecast for all future time-points. However, that paper is primarily concerned with the ability of methods to forecasting completed cohort fertility; while this is an important objective, many important applications of fertility forecasts rely on accuracy at the level of age-specific fertility rates. Forecasts of the number of births for future years require the timing of each cohort’s fertility and not just their ultimate level. In practice, models that perform well in forecasting completed cohort fertility may also be accurate at the lower level of aggregation. However, empirical tests are needed to confirm this.

A number of best-performing models of fertility identified by Bohk-Ewald and colleagues are assessed by using the root mean-squared error and empirical coverage calculated over a range of countries, using data from the Human Fertility Database. A model developed by the authors that employs parametric mixture models to forecast fertility is also assessed. A range of different choices for the forms of the mixture components are compared, and it is found that this family of models is competitive with the best-performing competitor fertility forecasting models. This result contrasts with the results of Bohk-Ewald et al., who suggest that parametric models do not perform well in forecasting contexts. This contradictory finding can be attributed to the fact that our model allows dependency between adjacent cohorts through time series priors on the model parameters.

¹ This work was supported by the ESRC Centre for Population Change - phase II (grant ES/K007394/1) at the University of Southampton. The use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work is also acknowledged. All the views presented herein are those of the authors only. Related code is available at github.com/jasonhilton/fertmix. This conference paper is a development of the work in “Forecasting Fertility using Parametric Mixture Models” available at <https://arxiv.org/pdf/1909.09545.pdf>.

I. Introduction

1. Forecasts of fertility are vitally important in anticipating demand for a variety of products and services. In the short term, demand for maternity care and nursery school places and the aggregate cost of child benefit payments all depend on the size of birth cohorts. Over longer horizons, future fertility is a crucial determinant of overall population size and age structure, and consequently will affect national accounts and fiscal sustainability (Office for Budget Responsibility 2018).
2. The accuracy of fertility forecasts is therefore of vital importance, as it can make substantial differences to both the adequacy of service provision and the state's fiscal position. Thus, assessments of competing forecast methods provide guidance for statistical offices in their choice of techniques. Additionally, methodologies that accurately capture the degree of uncertainty surrounding forecasts may help manage the risks surrounding inevitable deviations from point forecasts of fertility (Bijak et al. 2015). A predictive forecast distribution can be combined with information about the costs of over- or under-prediction and the degree of risk aversion of the decision maker to allow a course of action to be decided upon, following the principles of statistical decision theory (Bijak et al. 2015).
3. A recent comprehensive assessment of methods for completing cohort fertility schedules established that few perform better than the naive 'freeze rate' technique (Bohk-Ewald, Li, and Myrskylä 2018a). Comparing the projected CFS against observed data for each method over a wide range of countries and forecast jump-off years, the authors found that only the methods developed by Myrskylä, Goldstein, and Cheng (2013), Schmertmann et al. (2014), Ševčíková et al. (2015), and De Beer (1985) consistently outperformed the naive forecast.
4. However, that paper is primarily concerned with the ability of methods to forecasting completed cohort fertility; while this is an important objective, many important applications of fertility forecasts rely on accuracy at the level of age-specific fertility rates. Forecasts of the number of births for future years require the timing of each cohort's fertility and not just their ultimate level. The focus on assessing the ability of models to complete cohort schedules also means their success in predicting new cohorts was not assessed; clearly fertility forecasts cannot be made without generating predictions for cohorts that have yet to start their childbearing, so this is also important when considering the merit of forecasting methods. In practice, models that perform well in forecasting completed cohort fertility may also be accurate at the lower level of aggregation and for future cohorts. However, empirical tests are needed to confirm this. In the results presented below, the two best-performing models of fertility identified by Bohk-Ewald and colleagues are assessed by using the root mean-squared error and empirical coverage calculated over a range of countries using data from the Human Fertility Database.
5. Parametric models also perform poorly in the assessment of Bohk-Ewald, Li, and Myrskylä (2018a). The parametric models of Chandola, Coleman, and Hiorns (1999) and Peristera and Kostaki (2007) are designed for graduation and not forecasting, and do not allow dependencies between successive cohorts, so it is unsurprising that they perform poorly in a forecasting context. A family of Bayesian parametric models that do include such dependencies is presented here, and the performance of its members is assessed against existing models. For context, the next section reviews existing fertility forecasting literature.

II. Fertility Forecasting Models

6. A wide range of approaches have been deployed for modelling and forecasting fertility, of which Booth (2006) and Bohk-Ewald, Li, and Myrskylä (2018a) provide comprehensive reviews. An instructive area of the literature focuses on merely modelling age-specific fertility rates f_x rather than forecasting them. One approach which has particularly inspired the current work was introduced by Hoem et al. (1981), who used a range of density

functions to approximate the shape of the fertility curve for a given period. Density functions are useful in such a context as they are parsimonious representations of uni-modal curves, and are guaranteed to integrate to one. A minimum of two parameters are needed to describe the location μ and scale τ of the curve, and a separate parameter θ representing the TFR is employed multiplicatively to determine the area under the curve and thus the overall level of fertility, so that

$$f_x = \theta g(x; \mu, \tau)$$

where g is a density function.

7. Chandola, Coleman, and Hiorns (1999) adapted this work to account for the changes in observed data in the 1990s. In particular, the emergence of bi-modal fertility curves in England and the United States necessitated the use of two parametric components in order to obtain a reasonable approximation of the data. Peristera and Kostaki (2007) and Bermúdez et al. (2012) went further, including new families of density functions. Peristera and Kostaki (2007) used split-normal distributions where the variance parameters take different values either side of the mean, while Bermúdez et al. (2012) employed Weibull density functions, which may be parameterised to have either positive or negative skews.
8. Parametric approaches have also been employed for forecasting fertility, as opposed to merely modelling its age structure, although these have tended to involve only single-component models rather than mixtures. The double-exponential nuptial model of Coale and McNeil (1972) was reparameterised and adapted by Knudsen, McNown, and Rogers (1993) to allow for forecasting using ARIMA models, while Congdon (1990) similarly employed the function of Hadwiger (1940) in the forecasting of fertility for London boroughs. More recently, De Iaco and Maggio (2016) used ARIMA methods to forecast forward parameters of a gamma function fitted to Italian fertility, and furthermore used a Markov field model to capture correlations in the error structure of this model. In keeping with the approach adopted in this paper, Mazzuco and Scarpa (2015) attempt to capture and forecast the bimodal structure of fertility using a Flexible Generalisable Skew-Normal Distribution, although the parameters in this model are hard to interpret demographically.
9. Myrskylä, Goldstein, and Cheng (2013) demonstrate a pragmatic method for completing cohort fertility whereby for each age, the average change in the fertility rate over the preceding 5-year period is assumed to persist for 5 years into the future. The authors describe this as the ‘freeze-slope’ approach, for obvious reasons, and this is contrasted with naive ‘freeze-rate’ approach, in which rates are held constant. Schmertmann et al. (2014) incorporate elements of this approach within a conjugate normal-normal Bayesian model that generates prior distributions that smooth fertility rates in the both the age and cohort direction. The approach extracts three principal components from a large collection of historical data taken from the Human Fertility Database. One element of the constructed prior penalises deviations from linear combinations of these principal components. An additional element of the prior penalises deviations from both ‘freeze-rate’ and ‘freeze-slope’ projections. The priors are calibrated so that the size of the penalty reflects the distribution of such deviations in historical data. This approach is developed by Ellison, Forster, and Dodd (2018), who construct a hierarchical Bayesian model maintaining the underlying assumptions of the model of Schmertmann et al. (2014).
10. The work of Ševčíková et al. (2015) is of particular interest because it has been employed by the United Nations for the World Population Prospects since 2015. This method uses a Bayesian hierarchical model to describe the global distribution of schedules of TFR evolution, noting that countries tend to follow a common pattern of initially high fertility, decline, and later stabilisation. Expert opinion is utilised to determine target patterns of fertility over age (the so-called Proportionate Age Specific Rates) for each country.
11. As noted above, Bohk-Ewald, Li, and Myrskylä (2018a) test the ability of a range of these models to complete cohort fertility schedules, and find that the models of Schmertmann et al.

(2014) and Myrskylä, Goldstein, and Cheng (2013) perform particularly well, while the parametric models of Peristera and Kostaki (2007) and Chandola, Coleman, and Hiorns (1999) do not. These latter parametric models form the basis of the family of models set out in the next section.

III. Forecasting using Bayesian parametric mixture models

12. This section describes in detail the different elements of the Bayesian parametric mixture model for fertility, and the justifications behind them. In several developed countries, fertility rates in recent decades can be observed to take on a bi-modal shape when viewed as a function of age (Chandola, Coleman, and Hiorns 1999). This suggests the possibility that the underlying population is heterogeneous with respect to their fertility behaviour (ibid), which may reflect, for example, the differential opportunity cost of childbearing to women of different educational groups (Van Bavel 2010; Billari and Philipov 2005). This section describes an approach to the forecasting of fertility that captures this feature of contemporary fertility patterns. The method focuses first on forecasting the overall level of childbearing, measured by the average number of births to mothers of particular cohorts (that is, mothers born in the same year). Following Chandola, Coleman, and Hiorns (1999), this summary measure of fertility is subsequently decomposed over the fertile age range using a mixture of smooth parametric curves. Extending their approach to a forecasting context, predictions of future values are produced by forecasting forward both the parameters of these curves and the summary measure of the level of fertility using time series methods. Posterior predictive distributions coherently incorporate both the underlying time series stochasticity and uncertainty about model parameters.
13. It is common to describe the level fertility for a given calendar year or cohort in terms of summary measures. For cohort fertility, the most natural summary is the area under the fertility hazard function for that cohort, which is the average number of children born to a member of this birth cohort over the course of their lifetime, assuming they survive to the highest possible age of childbearing X (Jasilioniene et al. 2015). This is generally termed Completed Family Size (CFS), and can be approximated by the sum of the individual age-specific rates:

$$\begin{aligned}\theta_c &= \int_{12}^{50} \frac{\lambda_{ac}}{R_{ac}} da \\ &\approx \sum_{x=12}^{49} f_{x,c}\end{aligned}$$

14. The approximation is exact under the assumption the hazard and exposure are constant within each age interval. The equivalent period-based measure is the well-known Total Fertility Rate (TFR), which is constructed in the same way from Age-Period data, substituting the c cohort index for a t . This measure approximates the average number of children born to a ‘synthetic’ cohort who experienced a given period’s age-specific fertility rates across their whole fertile lifespan.

IV. Model Description

A. Model Overview

15. Starting with the definition of the likelihood, births at each age are assumed to be distributed according to a negative binomial distribution (parameterised according to the mean), so that

$$B_{xc} \sim \text{Negative Binomial}(R_{xc}f_{xc}, \exp(\phi_x)),$$

where B_{xc} are the number of children born to women of cohort c between exact ages x and $x + 1$, R_{xc} is the exposure to risk between these ages for this cohort, f_{xc} is the corresponding age-specific cohort fertility rate, and ϕ_x is a parameter controlling the degree of over-dispersion relative to the Poisson distribution, which varies with age x .

16. The fertility rate f_{xc} is modelled as product of θ_c , a parameter describing the average number of births to cohort c by maximum childbearing age X , and ξ_{xc} , where the vector of ξ_c parameters for each time period describes the age distribution of the fertility rate schedule (termed the Proportionate Age-Specific Fertility Rate (PASFR) by Ševčíková et al. (2015)), so that:

$$f_{xc} = \theta_c \xi_{xc}, \quad \sum_{x=14}^X \xi_{xc} = 1, \text{ for all } c$$

17. The age distribution ξ_c is modelled as a mixture of two components, each parameterised by a location and a spread parameter. In particular, as with Hoem et al. (1981), probability density functions are used for the components as these have the convenient property that they integrate to one. Alternatively, other functions with this property could also be used. For a particular age x , an approximation of the mass of the distribution between x and $x+1$, weighted by the mixture parameter ψ , defines that component's contribution to ξ , so that

$$\begin{aligned} \tau_{x,c} &= \psi_c g_1(x + 1/2; \mu_c^{(1)}, \tau_c^{(1)}) + (1 - \psi_c) g_2(x + 1/2; \mu_c^{(2)}, \tau_c^{(2)}) \\ \psi &\in [0,1] \end{aligned} \quad (1)$$

where $g_i(x; \mu, \tau)$ is the value at x of the density function with location $\mu_c^{(i)}$ and spread $\tau_c^{(i)}$, and ψ_c is the mixture parameter. The superscripts (i) on the location and spread parameters distinguish between the two mixture components.

B. Modelling evolution over time

18. The specification in Equation (1) leaves us with six unique parameters for each time period: two location parameters, two scale parameters, a mixture-weight parameter, and a parameter describing the overall level of fertility. One practical problem is ensuring that the two components are distinct to avoid identifiability and label switching issues. One strategy to circumvent this is to force the location of one component be strictly greater than the other. This can be achieved by re-parameterisation to use the sum and difference of the locations rather than the raw values, combined with enforcing constraints on these new parameters so that they are positive and lie within reasonable bounds. Given the range of fertile ages (~14-50 year), the sum of the locations should lie above 35, and to allow identification of the two mixture components, the gap is constrained to be at least 2 years.
19. After this re-parameterisation, the resultant set of parameters η_c can then be forecast forwards using auto-regressive models,

$$\begin{aligned} \eta_c &= \{\theta_c, \psi_c, \mu_c^{(s)}, \mu_c^{(d)}, \tau_c^{(1)}, \tau_c^{(2)}\} \\ \eta_c &\sim N(A\eta_{c-1}, \Sigma_\eta), \end{aligned}$$

where superscripts s and d indicate the sum and difference of the location parameters respectively. A special case of this class of models is the simple independent random walk:

$$\begin{aligned} \eta_c &= \eta_{c-1} + \epsilon_c \\ \epsilon_c &\sim N(\mathbf{0}, \sigma_\eta^T I). \end{aligned}$$

Priors for the first elements of all the parameters modelled as time series are chosen to be weakly informative. Flexibility as to the exact forecasting model is possible; more

complicated model classes, including ARIMA models incorporating differencing and moving average elements, or stochastic volatility models, could also be considered.

C. Model Components

20. Building on the models discussed previously by Hoem et al. (1981), Chandola, Coleman, and Hiorns (1999), and Bermúdez et al. (2012), we examine two possible parametric functions for the mixture components; the gamma density and the Weibull density. We wish to adopt non-standard parameterisation for each of these functions so that they are defined in terms of a location and a spread parameter. In this way, elements of the model that are to be given time series priors have a meaningful interpretation that is relatively consistent regardless of the specific function used. For the gamma component, the mode and standard deviation are used, while for the Weibull density, the form of the density function necessitates other choices; the median and the distance from the median to the upper quartile are instead used for the location and spread parameters. The split normal distribution of Peristera and Kostaki (2007) is not used, as although the authors find it fits the data relatively well, the join in the function at the mode makes it conceptually unappealing.
21. Although Chandola, Coleman, and Hiorns (1999), Peristera and Kostaki (2007) and Bermúdez et al. (2012) use the same density function for both the components (g_1 and g_2) of their respective parametric mixtures, there are advantages to combining different parametric functions within a mixture, as the different functions used have different tail behaviour, and may be better suited to approximating fertility at older or younger ages. For example, the Weibull density function can skew in either direction, allowing a steeper decline than is possible with the positively-skewed gamma distribution. These distinctions may help provide for better identification of the two components. As a result, we examined all possible combinations of the two functions.

D. Overdispersion

22. The Poisson distribution, a natural choice for modelling count data, has variance equal to its mean. This assumption can be too restrictive in many cases, and the use of the negative binomial distribution allows for over-dispersion relative to the Poisson, which tends to lead to smoother estimates of rates. However, the effect of over-dispersion is more pronounced in absolute terms for larger counts, and so the use of a single over-dispersion parameter for fertility can lead to a greater emphasis on the fit in the tails of the age-specific rate distribution where counts are small than in the more fertile stretches of the age range. To address this concern, we specify a smooth, age-specific over-dispersion function using a penalised basis-spline (Wood 2006), and allow the data to determine where greater over-dispersion is needed. A random walk prior penalises deviations from smoothness in the basis function coefficients (Lang and Brezger 2001):

$$\begin{aligned}\phi_x &= \mathbf{\beta}\mathbf{B}(x) \\ \beta_i &\sim \text{Normal}(\beta_{i-1}, \sigma_\phi^2).\end{aligned}\quad (2)$$

where $\mathbf{B}(x)$ is a matrix of B-spline basis functions.

23. Fertility data is often aggregated over open age intervals at the beginning and end of the fertile age range. For example, in the UK, births to women of ages 49 and above are reported in aggregate. In contrast to the Poisson case, the sum of n negative binomial distributions with different means does not follow a standard distribution, so for simplicity the distribution of births in these aggregate groups is approximated by a negative binomial distribution with the appropriate mean and variance. Only a small proportion of births occur

to women in these age groups, so the approximation is unlikely to make a significant difference to results.

V. Estimation and Results

24. In order to assess the performance of both the model set out in the previous section and two of the models identified as the best performing by Bohk-Ewald, Li, and Myrskylä (2018a) (those of Myrskylä, Goldstein, and Cheng (2013) and Schmertmann et al. (2014)) assessments were carried out on data from the Human Fertility Database for 25 countries. Although the Human Fertility Database provides birth data from ages 12-55, these data are extrapolated at the youngest and oldest ages where empirical information is not available (Jasilioniene et al. 2015). As a result, HFD data on births below age 15 and above 49 are aggregated for the purposes of model fitting. The latest 10 years of data were held back for each country and not used in fitting, to provide for an out-of-sample assessment. The model of Myrskylä, Goldstein, and Cheng (2013) was fitted based on code generously provided in the supplemental material of Bohk-Ewald, Li, and Myrskylä (2018a), while routines for fitting the model of Schmertmann et al. (2014) were derived from code on the first authors' website.
25. The No-U-Turn Sampler (NUTS) variant of Hamiltonian Monte Carlo (Hoffman and Gelman 2014), available in the Stan software package, was used to obtain posterior samples of the parametric mixture model. (Stan Development Team 2018). Hamiltonian Monte Carlo simulates movement through the parameter space by analogy to a physical system where the potential energy is equal to negative log-posterior (Neal 2010). Four chains, each consisting of 10000 iterations, were run for each combination of model specification and country. The first half of each chain was used as a warm-up period during which the sampling parameters were allowed to adapt to the shape of the posterior. With respect to the adaptation parameters, a slightly higher target acceptance rate (0.95) and maximum NUTS tree-depth (12) was used than is selected by default, providing for smaller integrator step-sizes and longer NUTS trajectories (Stan Development Team 2019). The remaining samples, thinned by a factor of 4 to prevent excessive memory use, were used for posterior inference.
26. For the majority of model and country combinations, Gelman-Rubin split \hat{r} diagnostics are below the suggested 1.05 threshold and examination of trace-plots indicate sampler convergence to the target distribution (Gelman and Rubin 1992; Stan Development Team 2019). In a minority of cases (27 out of 100), diagnostics indicated problems were encountered during sampling. In some cases, this was due to multi-modality in the posterior distribution, while in other cases, diagnostics identified divergent transitions during the sampling process. Divergent transitions indicate that the value of Hamiltonian in the No-U-Turn sampler differs greatly over the course of an HMC trajectory due to integration errors (Stan Development Team 2019). In practice this means that the curvature of the posterior varies such that a single leap-frog integrator step-size is insufficient to explore all areas of the posterior (Betancourt 2016). For the sake of this exercise, results from these models were discarded. In theory, discarding some results could bias assessment results, although examining the set of countries for which no results were discarded indicates only small difference from the results described below.

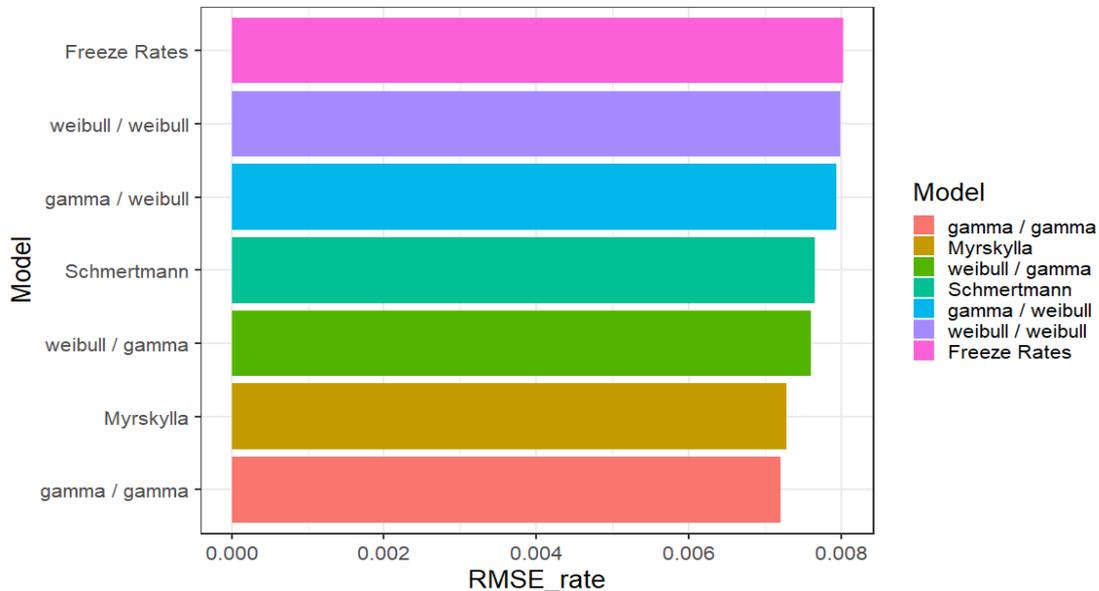
VI. Assessing Fertility Forecasting Models

27. In order to assess the forecasting performance of the models under study, we fit each model to HFD data for 25 countries, holding back the latest 10 years of data for each country. For the parametric mixture models, each possible configuration of functional components (for example, gamma-Weibull or gamma-gamma) the model was estimated in each case.

Forecasts of fertility rates are compared to actual observed forecasts for the ten most recent years (for cohort-based forecasts, the held-back sample is not strictly speaking for specific years, but for an index $t = c + a$, which will consist of data from two calendar years). The root mean squared error (RMSE) and the empirical coverage over all age-specific rates is calculated for each model to assess both the adequacy of the point forecasts and the calibration of the forecast uncertainty. With respect to coverage, for each forecast we calculate the proportion of corresponding observations of age-specific rates that fall within the specified probabilistic predictive intervals (90% and 50% in this instance). Because we are observing only one set of correlated outcomes for each country (age-specific rates for 2007-2016), rather than many independent repetitions, this metric does not correspond exactly to the frequentist concept of coverage, but it does however give an indication as to whether the uncertainty in forecasts appear well calibrated compared to actual observed deviations from predictions.

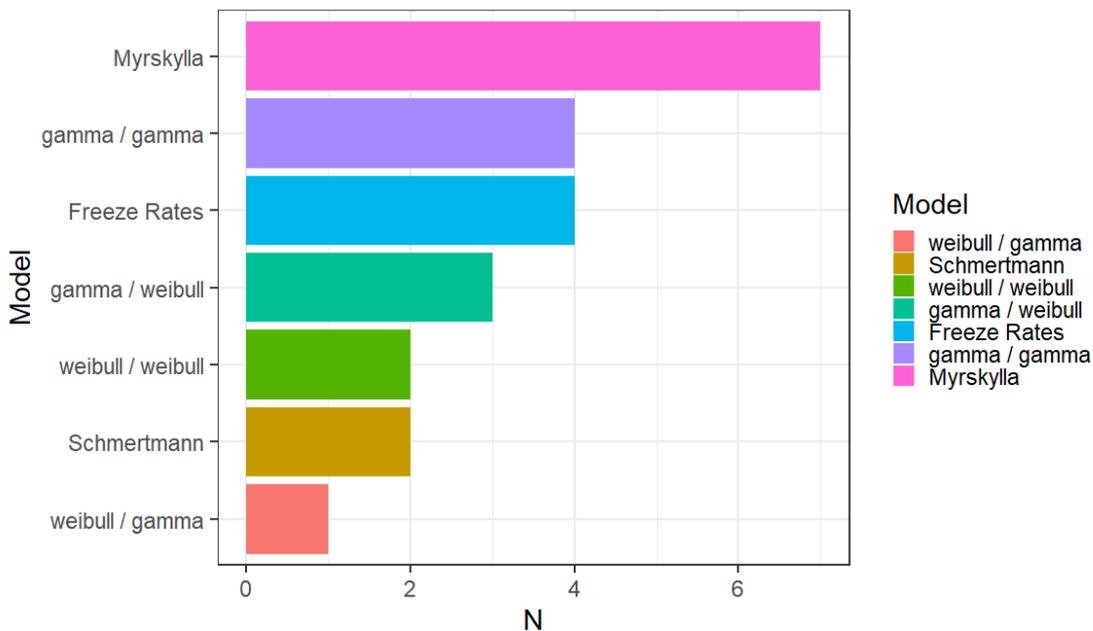
28. Figure 1 displays the average RMSE in rates for each model over all countries for which results are available (where lower values indicate better performance). In agreement with the Bohk-Ewald, Li, and Myrskylä (2018a) assessment of errors in completed family size, the models of both Myrskylä, Goldstein, and Cheng (2013) and Schmertmann et al. (2014) outperform the naive freeze-rate model. The four variants of the parametric mixture model also perform well, and in particular the gamma-gamma performs as well as the best competitor (Myrskylä, Goldstein, and Cheng (2013)). However, it should be noted that the assessment here only examines out-of-sample performance over the latest 10 years; the exercise conducted by Bohk-Ewald, Li, and Myrskylä (2018b) includes a far wider range of cohorts.

Figure 1: Average RMSE across all countries for forecast period, by model



29. As noted, the exclusion of some results from the set used to compute RMSEs because of sampling problems could potentially bias results. In particular, if sampling problems are more likely to occur for countries that are inherently more volatile and difficult to predict, this may lead to bias. A straight count of the number of times each model class ‘won’ is a more robust measure in this case. To compute this measure, for each country, the model which obtained the lowest RMSE was identified and these were summed to give a total for each model. Figure 2 shows that the model of Myrskylä, Goldstein, and Cheng (2013) outperforms all other models on this measure, although the gamma/gamma mixture again performs well.

Figure 2: Number of countries for countries for which a given model obtained the smallest RMSE, by model



30. The extent to which the quantification of uncertainty for a forecasting model reflects the distributions of outputs is also important for decision-making on the basis of forecasts. Figures 3 and 4 indicate the proportion of observations that fall within the 50% and 90% probabilistic interval for the models studied. From these charts it is evident that the model of Schmertmann et al. (2014) is particularly good at providing probabilistic intervals that reflect uncertainty in the underlying process. This is likely due to the calibration exercise that enables prior penalties to reflect historical variability in the data across a wide range of countries. The uncertainty intervals in parametric models are reasonable for the most part, however. The intervals associated with the forecasts produced by the Myrskylä, Goldstein, and Cheng (2013) model appear much too narrow. To a lesser extent, this was also a finding in Bohk-Ewald, Li, and Myrskylä (2018a).

Figure 3: Distribution over countries of proportion of observations within 50% forecast interval by model

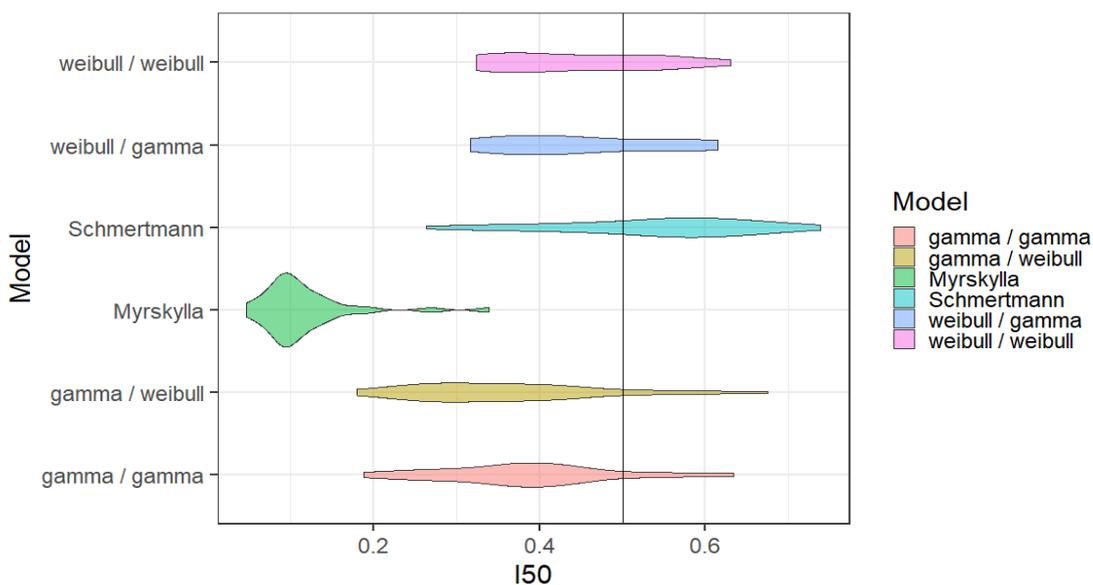
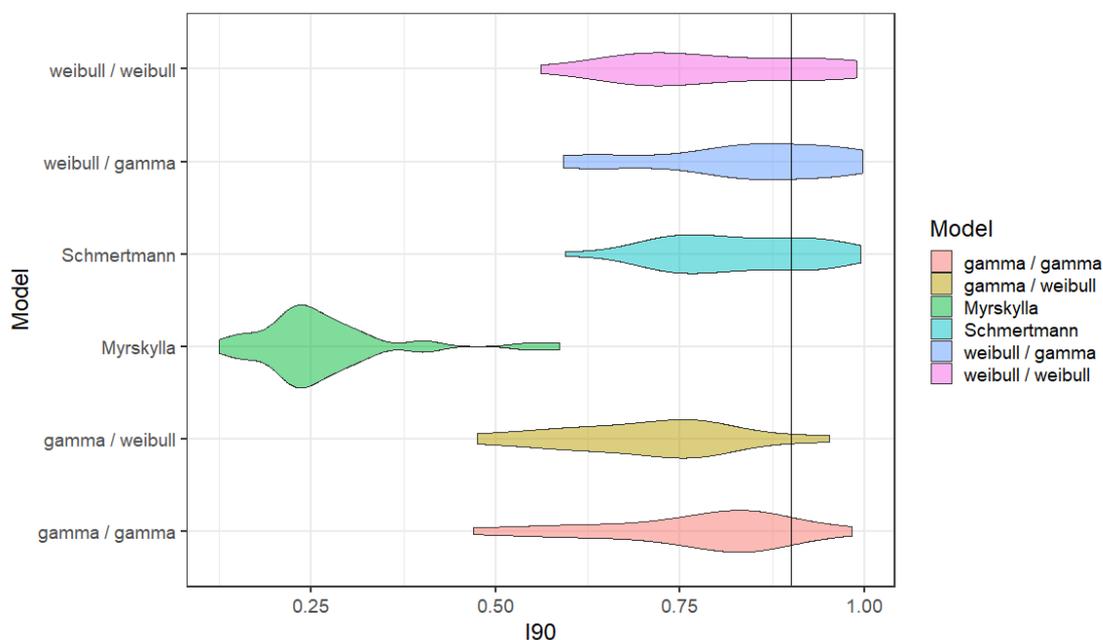


Figure 4: Distribution over countries of proportion of observations within 90% forecast interval by model



VII. Conclusion

31. This paper examined the performance of a small set of fertility forecasting models in predicting age-specific fertility rates over a 10 year horizon, complementing the work of Bohk-Ewald, Li, and Myrskylä (2018a), which looked at performance for completed family size. Looking at the combined RMSE and coverage assessments across the 25 countries studied, it is evident that parametric models perform reasonably well in comparison with competitor models, at least when looking at out-of-sample assessments over the last 10 years. Further work may examine expanding family of parametric models investigated to include more complex time series models, and include a range of forecast jump-off rates in the assessment process.

VIII. References

- Bermúdez, S., R. Blanquero, J.A. Hernández, and J. Planelles. 2012. “A new parametric model for fitting fertility curves A new parametric model for fitting fertility curves.” *Population Studies* 66 (3): 297–310.
- Betancourt, Michael. 2016. “Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo.”
- Bijak, Jakub, Isabel Alberts, Juha Alho, John Bryant, Thomas Buettner, Jane Falkingham, Jonathan J. Forster, et al. 2015. “Letter to the Editor: Probabilistic population forecasts for informed decision making.” *Journal of Official Statistics* 31 (4): 537–44. doi:[10.1037/emo0000122](https://doi.org/10.1037/emo0000122).Do.
- Billari, Francesco C., and Dimiter Philipov. 2005. “Women’s education and entry into a first union. A simultaneous-hazard comparative analysis of Central and Eastern Europe.” *Vienna Yearbook of Population Research* 1 (2004): 91–110. doi:[10.1553/populationyearbook2004s91](https://doi.org/10.1553/populationyearbook2004s91).
- Bohk-Ewald, Christina, Peng Li, and Mikko Myrskylä. 2018a. “Forecast accuracy hardly improves with method complexity when completing cohort fertility.” *Proceedings of the National Academy of Sciences* 115 (3): 630–81. doi:[10.1073/pnas.1722364115](https://doi.org/10.1073/pnas.1722364115).

- . 2018b. “Supplemental Material for ‘Forecast accuracy hardly improves with method complexity when completing cohort fertility.’” *Proceedings of the National Academy of Sciences* 115 (3): 630–81. www.pnas.org/cgi/doi/10.1073/pnas.1722364115.
- Bongaarts, John, and Griffith Feeney. 1998. “On the Quantum and Tempo of Fertility.” *Population and Development Review* 24 (2): 271–91.
- Booth, Heather. 2006. “Demographic Forecasting: 1980 to 2005 in review.” *International Journal of Forecasting* 22: 547–81. <http://www.sciencedirect.com/science/article/pii/S016920700600046X>.
- Chandola, T, D A Coleman, and R W Hiorns. 1999. “Recent European fertility patterns: fitting curves to ‘distorted’ distributions.” *Population Studies* 53 (3): 317–29. doi:[10.1080/00324720308089](https://doi.org/10.1080/00324720308089).
- Cheng, Roger, and Eric S. Lin. 2010. “Completing incomplete cohort fertility schedules.” *Demographic Research* 23: 223–56. doi:[10.4054/DemRes.2010.23.9](https://doi.org/10.4054/DemRes.2010.23.9).
- Coale, A. J., and D. R. McNeil. 1972. “The distribution by age of the frequency of first marriage in a female cohort.” *Journal of the American Statistical Association* 67 (340): 743–49. doi:[10.1080/01621459.1972.10481287](https://doi.org/10.1080/01621459.1972.10481287).
- Congdon, Peter. 1990. “Graduation of Fertility Schedules: An Analysis of Fertility Patterns in London in the 1980s and an Application to Fertility Forecasts.” *Regional Studies* 24 (4): 311–26. doi:[10.1080/00343409012331346014](https://doi.org/10.1080/00343409012331346014).
- . 1993. “Statistical Graduation in Local Demographic Analysis and Projection.” *Journal of the Royal Statistical Society (Series A)* 156 (2): 237–70.
- De Beer, J. 1985. “A time series model for cohort data.” *J Am Stat Assoc* 80 (391): 525–30. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve{&}db=PubMed{&}dopt=Citation{&}list{&}uids=12267306>.
- De Iaco, Sandra, and Sabrina Maggio. 2016. “A dynamic model for age-specific fertility rates in Italy.” *Spatial Statistics* 17. Elsevier Ltd: 105–20. doi:[10.1016/j.spasta.2016.05.002](https://doi.org/10.1016/j.spasta.2016.05.002).
- Ellison, Joanne, Jonathan J Forster, and Erenkul Dodd. 2018. “Forecasting of cohort fertility under a hierarchical Bayesian approach.” <https://eprints.soton.ac.uk/cgi/eprintbypureuuid?uuid=dead073b-5d4c-47af-b4bc-13c9dfdea9e6>.
- Evans, M. D. R. 1986. “American Fertility Patterns : A Comparison of White and Non-white Cohorts Born 1903-56.” *Population and Development Review* 12 (2): 269–93.
- Gelman, Andrew, and Donald B Rubin. 1992. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (4): 457–72.
- Hadwiger, H. 1940. “Eine analytische reproduktionsfunktion für biologische gesamtheiten.” *Scandinavian Actuarial Journal* 3-4: 101–13. doi:[10.1080/03461238.1940.10404802](https://doi.org/10.1080/03461238.1940.10404802).
- Hoem, Jan M., Dan Madsen, Jorgen Lovgreen Nielsen, Else-Marie Ohlsen, Hans Oluf Hansen, and Bo Rennermalm. 1981. “Experiments in Modelling Recent Danish Fertility Curves.” *Demography* 18 (2): 231–44.
- Hoffman, M D, and Andrew Gelman. 2014. “The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research* 15 (2008): 1–31. <http://arxiv.org/abs/1111.4246>.
- Hyndman, Rob J., and Md. Shahid Ullah. 2007. “Robust forecasting of mortality and fertility rates: A functional data approach.” *Computational Statistics and Data Analysis* 51 (10): 4942–56. doi:[10.1016/j.csda.2006.07.028](https://doi.org/10.1016/j.csda.2006.07.028).
- Jasilioniene, A., D. A. Jdanov, T Sobotka, E. M. Andreev, K. Zeman, and V. M. Shkolnikov. 2015. “Methods Protocol for the Human Mortality Database.” Human Fertility Database. <https://www.humanfertility.org/Docs/methods.pdf>.

- Knudsen, Christin, Robert McNown, and Andrei Rogers. 1993. "Knudsen, Christin McNown, Robert Rogers, Andrei." *Social Science Research* 22: 1–23.
- Lang, Stefan, and Andreas Brezger. 2001. "Bayesian P-Splines." Insitut Für Statistik Sonderforschungsbereich 386. Munich: Ludwig-Maximilians-Universität-München. <https://epub.ub.uni-muenchen.de/1617/>.
- Lee, R D. 1993. "Modeling and forecasting the time series of US fertility: age distribution, range, and ultimate level." *International Journal of Forecasting* 9 (2): 187–202. doi:[10.1016/0169-2070\(93\)90004-7](https://doi.org/10.1016/0169-2070(93)90004-7).
- Lee, Ronald D, and Lawrence R Carter. 1992. "Modeling and Forecasting U.S Mortality." *Journal of the American Statistical Association* 87 (419). American Statistical Association: pp. 659–71. <http://www.jstor.org/stable/2290201>.
- Li, Nan, and Zheng Wu. 2003. "Forecasting cohort incomplete fertility: A method and an application." *Population Studies* 57 (3): 303–20. doi:[10.1080/0032472032000137826](https://doi.org/10.1080/0032472032000137826).
- Mazzuco, Stefano, and Bruno Scarpa. 2015. "Fitting Age-Specific Fertility Rates By a Skew-Symmetric Probability Density Function." *Journal of the Royal Statistical Society (Series A)* 178 (1): 187–203.
- Myrskylä, M, Joshua R Goldstein, and Yen-Hsin Alice Cheng. 2013. "New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals." *Population and Development Review* 39 (1): 31–56. <http://onlinelibrary.wiley.com/doi/10.1111/j.1728-4457.2013.00572.x/abstract>.
- Neal, Radford. 2010. "MCMC using Hamiltonian Dynamics." In *Handbook of Markov Chain Monte Carlo*, edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng Meng. Chapman; Hall / CRC Press.
- Ní Bhrolcháin, Maire. 1992. "Period Paramount? A Critique of the Cohort Approach to Fertility." *Population and Development Review* 18 (4): 599. doi:[10.2307/1973757](https://doi.org/10.2307/1973757).
- Office for Budget Responsibility. 2018. "Fiscal Sustainability Report." July. Office for Budget Responsibility. doi:[10.1142/S021812749900095X](https://doi.org/10.1142/S021812749900095X).
- Peristera, Paraskevi, and Anastasia Kostaki. 2007. "Modeling fertility in modern populations." *Demographic Research* 16. doi:[10.4054/DemRes.2007.16.6](https://doi.org/10.4054/DemRes.2007.16.6).
- Ryder, N B. 1964. "The Process of Demographic Translation." *Demography* 1 (1): 74–82.
- Schmertmann, Carl, Emilio Zagheni, Joshua R. Goldstein, and Mikko Myrskylä. 2014. "Bayesian Forecasting of Cohort Fertility." *Journal of the American Statistical Association* 109 (506): 500–513. doi:[10.1080/01621459.2014.881738](https://doi.org/10.1080/01621459.2014.881738).
- Stan Development Team. 2018. "The Stan Core Library." <http://mc-stan.org>.
- . 2019. "Stan Modeling Language Users Guide and Reference Manual." <http://mc-stan.org/index.html>.
- Ševčíková, Hana, Nan Li, Vladimíra Kantorová, Patrick Gerland, and Adrian E. Raftery. 2015. "Age-Specific Mortality and Fertility Rates for Probabilistic Population Projections." doi:[10.1007/978-3-319-26603-9_15](https://doi.org/10.1007/978-3-319-26603-9_15).
- Van Bavel, Jan. 2010. "Choice of study discipline and the postponement of motherhood in europe: The impact of expected earnings, gender composition, and family attitudes." *Demography* 47 (2): 439–58.
- Wiśniowski, Arkadiusz, Peter W. F. Smith, Jakub Bijak, James Raymer, and Jonathan J. Forster. 2015. "Bayesian Population Forecasting: Extending the Lee-Carter Method." *Demography* 52: 1035–59. doi:[10.1007/s13524-015-0389-y](https://doi.org/10.1007/s13524-015-0389-y).
- Wood, Simon N. 2006. *Generalised Additive Models: An Introduction with R*. Boca Raton: Chapman; Hall / CRC Press.