Incorporating model uncertainty into fertility schedule estimates for population forecasting

Jonathan J. Forster, Jakub Bijak, Peter W.F. Smith

Centre for Population Change, University of Southampton, UK
Email: {J.J.Forster}{J.Bijak}{P.W.Smith}@soton.ac.uk

Extended abstract

Population estimation and forecasting via the cohort-component approach requires, as a key input, an estimate or projection of the age-specific fertility rates over the corresponding estimation or forecast period. A wide range of models have been proposed for estimating fertility schedules; see, for example, Bermúdez et al (2012), Coale and Trussell (1974), Congdon(1990), De Beer (1989), Hoem et al (1981), Lee (1993), Li and Wu (2003), Peristera and Kostaki (2007) and Rogers (1986). Having been chosen, the favoured model for age-specific fertility rates is typically fitted against a suitable dataset to obtain a fertility schedule estimate and, ideally, an estimate of the associated uncertainty. However, different models not only yield different best estimates but also generate different prediction intervals, that is, the uncertainty is model-dependent. In this paper, we describe a statistical approach to the quantification of fertility schedule uncertainty that incorporates model uncertainty, that is, an approach that explicitly accounts for the fact that different models are available.

Bayesian statistical inference can provide a coherent fully probabilistic approach where inference and prediction statements are based on the posterior probability distribution of the unknown quantities given the observed data. Commonly these unknown quantities are parameters of a single statistical model, but the approach is sufficiently flexible to include uncertainty about the form of the model itself. Crucially, this allows predictive probability statements to be made which incorporate model uncertainty, and hence we consider such an approach to be well-suited to population projection under model uncertainty. For example, Abel et al (2013) illustrate how Bayesian methods can be used to provide more realistic projection uncertainty in an application to population forecasting using simple time series models. Here, we consider a more realistic Bayesian projection methodology based on the cohort-component approach, focussing on the fertility schedules which are required as projection model inputs.

Briefly, the Bayesian approach under model uncertainty updates a prior probability distribution over the models (in the form of probabilities or weights assigned to models) to a posterior distribution, in light of observed data. The posterior distribution accounts for how well the various models fit the observed historical data, and is used explicitly in weighting the models in projections. This approach is sometimes referred to as Bayesian model averaging – see Hoeting et al (1999) for details. The final projections fully integrate both model uncertainty and uncertainty about the parameters of the constituent models. Furthermore, as the approach is fully probabilistic, it is straightforward to also allow additional assumptions about future changes in key parameters (structural breaks) and other aspects to be input into the projections, with the corresponding uncertainty being integrated in an entirely coherent fashion.

Although the principles of Bayesian inference (including under model uncertainty) are straightforward, practical methodology for incorporating probabilistic model uncertainty into estimates and forecasts of fertility schedules is currently underdeveloped. In this paper, we provide such a methodology. Initially, we focus on individual models, and develop Bayesian methodology for computing the predictive (forecast) distributions for various models. This is in the same spirit as the Bayesian approach of Schmertmann et al (2013) for forecasting fertility for incomplete cohorts. The models we consider range from parsimonious models with few parameters through more complex mixture-type models, which account for bifurcation in fertility behaviour, to highly flexible semi-parametric models, and include models with and without explicit cohort dynamics. The main contribution of our work is that we then demonstrate how to effectively compute probabilistic projections *across* fertility schedule models. This enables more robust projections to be obtained, with more realistic quantification of uncertainty.

Our approach is illustrated on data from England and Wales. Using a selection of different plausible models, we present the estimated fertility schedules provided by each model, and illustrate how the posterior model probabilities are computed, together with the resulting forecast arising from integrating over the models to account for model uncertainty. The integrated projection uncertainty provides a coherent and more realistic assessment of uncertainty than any corresponding analysis based upon a single model. We also discuss how 'model-averaged' fertility schedules can be combined with similarly integrated mortality forecasts in an overall probabilistic population projection.

References

Abel, G.J., Bijak, J., Forster, J.J., Raymer, J., and Smith, P.W.F. (2013). What do Bayesian methods offer population forecasters? *Demographic Research* to appear.

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

Coale, A. J. and Trussell, T. J. (1974). Model fertility schedules: variations in the age structure of childbearing in human populations. *Population Index*, **40**, 185-258.

Congdon, P. (1990). Graduation of fertility schedules: an analysis of fertility patterns in London in the 1980s and an application to fertility forecasts. *Regional Studies*, **24**, 311-326.

De Beer, J. (1989). Projecting age-specific fertility rates by using time-series methods. *European Journal of Population*, **5**, 315-346.

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

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. (1999). Bayesian Model Averaging: A Tutorial. *Statistical Science*, Vol. 14, No. 4, 382-417.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: age distribution, range, and ultimate level. *International Journal of Forecasting*, **9**, 187-202.

Li, N., and Wu, Z. (2003). Forecasting cohort incomplete fertility: a method and an application. *Population Studies*, **57**, 303-320.

Peristera, P. and Kostaki, A. (2007). Modelling fertility in modern populations. *Demographic Research*, **16** 141-194.

Rogers, A. (1986). Parameterized multistage population dynamics and projections. *Journal of the American Statistical Association*, 81, 48-61.

Schmertmann, C., Zagheni, E., Goldstein. J. and Myrskylä, M. (2013) Bayesian forecasting of cohort fertility *MPIDR Working Paper WP*, **2012-003**.