Bayesian Demography 250 Years after Bayes

Short title: Bayesian Demography

Jakub Bijak[1*] and John Bryant[2]

[1] University of Southampton, Southampton, United Kingdom

[2] Statistics New Zealand, Christchurch, New Zealand, and the University of Waikato, Hamilton, New Zealand

* Address for correspondence: Department of Social Statistics and Demography, University of Southampton, Southampton, SO17 1BJ, United Kingdom, email: j.bijak@soton.ac.uk

Abstract

Bayesian statistics offers an alternative to classical (frequentist) statistics. It is distinguished by its use of probability distributions to describe uncertain quantities, what leads to elegant solutions to many difficult statistical problems. Although Bayesian demography, like Bayesian statistics more generally, is around 250 years old, only recently has it begun to flourish. The aim of this paper is to review the achievements of Bayesian demography, address some misconceptions, and make the case for wider use of Bayesian methods in population studies. We focus on three applications: demographic forecasts, limited data, and highly-structured or complex models. In general, the ability to integrate information from multiple sources, and to describe uncertainty coherently, is the key advantage of Bayesian methods. Bayesian methods also allow for including additional (prior) information next to the data sample. As such, Bayesian approaches are complementary to many traditional methods, which can be productively re-expressed in Bayesian terms.

Bayesian Demography 250 Years after Bayes

Keywords:

Bayesian demography, Bayesian statistics, demographic methodology, population estimates and forecasts, statistical methods

1. Introduction

The original paper of Thomas Bayes (1763) establishing the theorem that bears his name was presented to the Royal Society just over 250 years ago, two years after Bayes's death, by his friend Richard Price (Courgeau 2012). Not long afterwards, Bayesian demography was born: in 1778, Pierre-Simon de Laplace, who later extended Bayes's Theorem to a more general case, applied this method of inference to estimating sex ratios at birth for Paris and London (Laplace, 1781; after Courgeau 2012). However, for about two centuries, Bayesian demography remained largely dormant. Only in recent decades there has been a revival of demographers' interest in Bayesian methods, following the methodological and computational developments of Bayesian statistics. The area is currently growing fast, especially with the UN population projections becoming probabilistic – and Bayesian (Gerland et al. 2014).

The aim of this paper is to review the achievements of Bayesian demography, address some misconceptions about Bayesian approaches, and to make the case for more widespread use of Bayesian statistical methods, especially in currently underexplored areas of population studies. We review, synthesise and evaluate three especially promising areas of application: demographic forecasts, problems with limited data, as well as highly-structured and complex models. However, this paper is not only pedagogical in scope, but also aims to contribute to the literature by offering suggestions for new, uncharted applications of Bayesian methods in demography. Another common theme appearing throughout this paper is: how traditional demographic methods can be re-expressed in Bayesian terms – and how fruitful this approach can be.

For each of the three areas mentioned above, we discuss the main arguments put forward for the use of Bayesian methods, illustrated by selected examples from the demographic literature. Whilst the current review has inevitably omitted some particular pieces of work, we have nonetheless attempted to cover all the main applications of contemporary Bayesian demography.

This paper is structured as follows. After the current Introduction, in Section 2 we present the basic tenets of Bayesian statistics, with particular focus on the practical areas of application that are relevant for contemporary demography, and critically evaluate some related misconceptions. We further explore three of these areas in more detail in Section 3, where we look at population forecasting; population modelling for small samples, sparse or incomplete data; and the use of highly-structured and complex models in demography. Finally, in Section 4, we discuss the future of Bayesian demography, its prospects and challenges.

2. Bayesian statistics in a nutshell

2.1 A brief history of the Bayesian approach

In the Bayesian approach, probability is typically interpreted as a representation of the subjective beliefs of the person making the inference. In contrast, the objectivist interpretation of probability, grounded on different philosophical premises, has been the dominant paradigm in statistics for most part of the 20th century. The most prominent school of thought, associated with Ronald A. Fisher, Jerzy Spława-Neyman, Egon S. Pearson, Karl Pearson, and others, linked probability with the frequencies of events under study (see Courgeau 2012 for details). The “frequentist” school is mainly associated with methods for estimation and inference that are purely based on the *likelihood*: probability that a set of data was generated by a model with given parameters, treated as fixed, yet unknown. From this point of view, the Bayesian perspective was criticised, mainly for its expressed recourse to the notion of subjectivity.

Philosophical differences aside, computational demands also for a long time held back the practical application of Bayesian statistics. The Bayesian paradigm has made a slow, yet steady comeback to the statistical mainstream only since the 1970s, at first owing to theoretical achievements (Savage 1954/1972; Lindley 1972; for an overview, see Bernardo and Smith 2000). This process has accelerated in the 1980s thanks to rapid developments in fast computing (Ntzoufras 2009). Due to increase in computing speed, and the development of Markov chain Monte Carlo (MCMC) methods (Geman and Geman 1984; Gelfand and Smith 1990), computations that were once difficult or impossible are now routine.

Currently, even though the frequentist statistics still dominates in introductory statistical curricula, Bayesian inference and methods are becoming increasingly common in applied research in various fields, including demography and other related population sciences – from epidemiology (Broemeling 2013), to paleodemography (Caussinus and Courgeau 2010), to phylogenetics and historical linguistics (Greenhill and Gray 2009). An excellent introduction to a range of social science applications of Bayesian methods has been provided by Lynch (2007), and an important general textbook, also including many examples of application in social sciences, is Hoff (2009). In demographic literature, a brief introduction to Bayesian statistics for population sciences, with focus on migration, has been offered by Bijak (2010).

Some of these developments are visible in trends in relative frequencies of phrases ("*n*-grams") used in publications – in this case, in the digital Google books collection. Figure 1 shows English-language examples of such trends, smoothed with five-term moving averages, for Bayesian statistics, estimation and methods (black lines), which are compared with classical/frequentist/likelihood statistics, estimation and methods (grey lines). Thus, within the digitised books, Bayesian statistics has been mentioned more frequently than classical/frequentist statistics since the 1980s. Even though the classical/likelihood estimation is currently mentioned over four times more frequently in the Google books content, its relative frequency is stagnant, as opposed to its steadily-increasing Bayesian counterpart. With respect to more practically-oriented "methods", the trends are even more favourable to the Bayesian approaches.

[Figure 1 here]

Figure 1: Frequencies for Bayesian and frequentist / likelihood search terms in Google books

Notes: For caveats regarding the use of Google Ngrams in the context of demography, see Bijak et al. (2014)

Source: Google books Ngram Viewer http://books.google.com/ngrams, English corpus, queried on 20.01.2014

2.2 Bayesian theory: Key aspects

The foundation of Bayesian analysis is Bayes's theorem (1763), which can be stated as follows:

$$(1) \qquad p(Unknowns \mid Data) = \frac{p(Data \mid Unknowns) \times p(Unknowns)}{p(Data)}.$$

The 'p's in the equation denote probability distributions, that is, probabilities or probability density functions. The 'unknowns' may be abstract quantities such as regression coefficients, but may also be potentially observable quantities such as births that were missed by a registration system. The outcome from a Bayesian analysis is a *posterior probability distribution* for the unknowns, conditional on the data. The posterior distribution is obtained by multiplying the likelihood, $p(Data \mid Unknowns)$, by the prior probability distribution, $p(Unknowns)$, all divided by $p(Data)$, the marginal likelihood, that is, the probability of obtaining a particular sample.

The fact that the outcome of an analysis is a probability distribution is an important advantage of Bayesian statistics. A probability distribution contains a great deal more information than the point estimate and standard error typically produced by a frequentist analysis. For instance, examination of the full distribution may reveal multiple modes, or show the probability mass to be concentrated in a certain area. But even more importantly, probability distributions can be summarised in ways that are more intuitive both to lay people and fellow scientists. It is, for instance, legitimate to interpret a Bayesian 95% credible interval as having a 95% probability of containing the true value. This is *not* the correct interpretation of a frequentist 95% confidence interval. The '95%' from such a confidence interval refers to the proportion of hypothetical intervals that would contain the true value *if* the study were replicated many times. Converting a confidence interval into a statement about the data at hand requires further assumptions, like those described by a prior (Jaynes 1976; Howson and Urbach 2006).

In addition, it is straightforward to derive substantively meaningful quantities from a full probability distribution. It is, for instance, easy and natural to turn a joint probability distribution for age-specific mortality rates into a probability distribution for life expectancy. In contrast, turning point estimates and standard errors for age-specific mortality rates into a point estimate and standard error for life expectancy is technically more complicated and requires additional assumptions. Moreover, standard frequentist methods do not always produce the desired results when the numbers of deaths are small (Eayres and Williams 2004).

The likelihood component $p(Data \mid Unknowns)$ in a Bayesian analysis is essentially the same as the likelihood in a frequentist analysis. In both cases it is a model for how the data were generated, given a set of parameters. A likelihood function might, for instance, state that births follow a Poisson distribution, with expected values that vary with age and population size. The marginal likelihood $p(Data)$, is fixed for any given dataset, and plays a minor role in most applications.

The prior, $p(Unknowns)$, is distinctively Bayesian. It captures information about unknowns that is not contained in the data. An example of a prior for might be a normal distribution with mean 1.8 and standard deviation 0.2, $N(1.8, 0.2^2)$, used to represent prior beliefs about a total fertility rate in a given developed country. Priors of this type, expressing quantitative statements about a parameter, are an important feature of applied Bayesian statistics. However, when data are abundant, it is more common for priors to have much less information content, and to be restricted to qualitative features of the data. For instance, a prior might state that a mortality profile is expected to rise monotonically with age; that neighbouring regions are likely to be more similar than distant ones; or that household incomes can be treated as independent draws from the same distribution (Carlin 1992; Congdon 2010, Chapter 4). Indeed, priors may be even weaker than this, to the point where they are ‘non-informative’, and are dominated by the likelihood (Gelman et al. 2014: 51-55). For such cases, the results of the analysis are often numerically close to their classical (frequentist) equivalents.

In modern Bayesian analyses, priors are often hierarchical. In other words, the parameters governing the priors are themselves given ‘hyper-priors’ governed by ‘hyper-parameters’. For instance, let y_{asr} represent deaths of people of age a and sex s in region r; and let x_{asr} be the corresponding population at risk. Then, the age-sex-region-specific mortality rates θ_{asr} might be estimated using the following model, where the tilde ‘~’ denotes ‘follows a distribution’, and M stands for an arbitrary large number:

(2)

$$y_{asr} \sim \text{Poisson}(\theta_{asr} x_{asr})$$

$$\log \theta_{asr} \sim \text{Normal}(\beta_0 + \beta_s + \beta_a + \beta_r, \sigma^2)$$

$$\beta_0 \sim \text{Normal}(0, M^2)$$

$$\beta_s \sim \text{Normal}(0, M^2)$$

$$\beta_a \sim \text{Normal}(\beta_{a-1}, \tau_{\mathrm{A}}^2)$$

$$\beta_r \sim \text{Normal}\left(0, \tau_{\mathrm{R}}^2\right),$$

For the hyper-parameters, the following distributions can be assumed: $\sigma, \tau_{\mathrm{A}}, \tau_{\mathrm{R}} \sim \text{Uniform}(0, M)$.

The first equation in (2) is the likelihood. The second equation gives the prior for the mortality rates. The log-transformed rates are expected to follow normal distributions, with means varying by age, sex, and region. Each of the age, sex, and region effects in turn has a prior: for instance, the means for the age effects are expected to follow a random walk without drift. In this example, the intercept and sex terms, as well as all the standard deviations, are given vague (hardly informative) priors.

As discussed further in Section 3, hierarchical models are an attractive way to model many population processes, and lead to sensible estimates when data are sparse. Frequentist versions of Bayesian hierarchical models exist, such as multilevel models or random effects models. However, hierarchical models are particularly natural within the Bayesian framework because of the blurring of the distinction between observations and parameters, with both being treated as draws from probability distributions (Rubin 1984, p. 1154).

A computational shortcut when fitting a hierarchical model is to obtain point estimates for hyper-parameters using methods such as maximum likelihood, and to plug these estimates back into the model, treating them as known with certainty. This is referred to as an empirical Bayesian approach (e.g. Maritz 1970). However, treating hyper-parameters as known when they are actually estimated, and using the same data twice – once to estimate hyper-parameters, and again to estimate the remaining parameters – can lead to estimates that are spuriously precise. Refinements of empirical Bayesian methods that address these concerns do exist (Carlin and Louis 2009, Chapter 5). However, with advances in computing power, the need to avoid estimating hyper-parameters within a full probabilistic model, and hence the motivation for empirical Bayesian methods, is declining.

Finally, there is one set of applications where the marginal likelihood *p*(*Data*) does play an important role: Bayesian model selection and model averaging. Here, the model specification itself is treated as an unknown (e.g. Raftery 1995). A set of alternative models is considered, and the posterior probability of selecting a particular model from this set, *p*(*Model* | *Data*), is also calculated by using Bayes's theorem. This approach requires setting prior probabilities for all the elements of the model space, *p*(*Model*), and providing the likelihood function, *p*(*Data* | *Model*), which is simply the marginal likelihood for the more standard single-model case. Calculating the marginal likelihood can be non-trivial and usually involves the use of sophisticated numerical algorithms, an overview of which can be found in Dellaportas, Forster and Ntzoufras (2002). Once the posterior model probabilities are estimated, they can be used either to select the model with the highest data support, or to average the outcomes of different models, weighting them by these posterior probabilities. In the latter case, model uncertainty is explicitly included in the results.

2.3 Bayesian practice: Computations and decisions

In a sense, once a Bayesian has specified a prior and a likelihood for a given dataset, the modelling is finished. The result of the analysis – the posterior distribution – is now completely determined via Bayes's theorem. In practice, describing the implied posterior distribution can be challenging. Except in textbooks, the posterior distribution rarely takes a standard form such as a gamma or normal distribution, or other priors from the so-called ‘conjugate families’, for which convenient analytical solutions exist (see Gelman et al. 2014 for examples). More typically, the posterior can be obtained as a marginal distribution from the (multidimensional) joint probability distribution. Determining the shape of the posterior distribution, or even deriving some summary measures, was once the greatest practical obstacle to applied Bayesian analysis.

Since the 1980s, many new methods for describing the posterior distribution have been developed. The dominant approach has been Markov chain Monte Carlo (MCMC) (Gelfand and Smith 1990; Gilks et al. 1996; Gelman et al. 2014). MCMC yields an approximate sample from the posterior distribution. The analyst calculates summary statistics, such as means or intervals, for the approximate sample, and treats them as summary statistics for the true posterior distribution.

There are two key reasons why MCMC methods make otherwise intractable modelling challenges tractable. The first is that they allow the difficult problem of drawing a value from $p(\theta)$ to be replaced by the easier problem of generating a candidate value θ^* and calculating $p(\theta^*)$. The second is that they permit a divide-and-conquer approach: samples for the whole model can be constructed by successively sampling from each of the components. The main disadvantage of MCMC methods is that they can be computationally-intensive and slow.

One area where Bayesian statistics lags behind frequentist statistics is user-friendly software for fitting standard models (for an overview, see Wiśniowski’s contribution in Bijak 2010). Fitting a basic Bayesian model typically requires more programming than an equivalent frequentist model. BUGS (Bayesian Inference Using Gibbs Sampling) is the most mature general-purpose user-friendly package for Bayesian modelling (Lunn et al. 2009; Ntzoufras 2009), but it can struggle with large datasets or complex models. Bayesian computation is, however, an active field of research. New packages implementing general-purpose algorithms or specific techniques relevant to demography appear regularly on the “Bayesian Inference” home page for the *R* programming language. The programming effort required for fitting a *non-standard* model, with features such as constraints or nonstandard distributions, is often smaller for Bayesian models than it is for frequentist ones, thanks to the flexibility of MCMC (Gilks, Richardson, and Spiegelhalter 1996). Moreover, there are now many books such as Congdon (2010, 2014), Gelman and Hill (2006), Kruschke (2010), and Marin (2014) that use social science examples to teach Bayesian modelling and MCMC.

In coming years, it seems likely that existing MCMC methods will cede ground to newer, faster techniques. Gelman et al (2014), for instance, promote the software package STAN as a successor to BUGS. STAN combines standard MCMC with a related technique called Hamiltonian Monte Carlo (Neal 2011). Similarly, the statistical package INLA produces extremely fast and accurate approximations for a general class of Bayesian models (Rue et al. 2007).

In practical applications, the outcomes of Bayesian analysis have also a potential to serve as a basis for a formal decision support. Here, it is worth noting that the axiomatic foundations of Bayesian analysis are firmly rooted in statistical decision theory, and in the notions of utility or loss functions (see DeGroot 1970/2004, and Bernardo and Smith 2000 for details; and Courgeau 2012 for a discussion). If the results of a Bayesian analysis can be combined with such utility or a loss functions, reflecting the gains or losses from making particular decisions based on the unknown (estimated) quantities, the result is an elegant statistical system for decision-making (e.g. DeGroot 1970/2004).

In brief, an optimal Bayesian decision for a given unknown quantity, described by a probability distribution, is one that maximises the expected utility, or minimises the expected loss. This approach allows the analyst to reduce whole probability distributions to point estimates, taking into account the respective costs or gains from over- and under-estimation, which need not be symmetric. Standard examples of such point estimates include: the median of the probability distribution for symmetric linear loss (utility) functions; quantiles for asymmetric linear functions; and the mean for quadratic function. (DeGroot 1970/2004). In practical applications, such utility or loss functions could measure the various outcomes – for instance in monetary terms – of decisions based on uncertain demographic parameters (for examples related to fiscal and other macroeconomic implications, sustainability of pension systems, or health care expenditure, see Alho, Hougaard Jensen and Lassila 2008).

2.4 Criticisms and misconceptions

Critics of Bayesian statistics have traditionally portrayed the need to specify a prior as a weakness of the Bayesian approach, on the grounds that it introduces an unacceptable element of subjectivity. This is an influential objection. It has, for instance, deterred the use of Bayesian methods by national statistical agencies, which are sensitive to any charge of subjectivity (Fienberg 2011; Pfeffermann 2013). Bayesians typically respond to such charges with three overlapping arguments: (i) some degree of subjectivity is unavoidable in any non-trivial statistical analysis; (ii) in many analyses, a wide range of sensible priors will lead to similar results; and (iii) in applications where it is impossible to avoid a substantial degree of subjectivity, documenting the subjective choices in the form of priors is a healthy and transparent habit.

A non-trivial statistical analysis – that is, an analysis involving imperfect data and complex relationships among variables – typically requires many choices that cannot be decided from the data alone. For instance, an analyst may have to choose between several possible error distributions, several plausible methods for dealing with missing values, or several possible sets of explanatory variables. These sorts of choices typically require knowledge about the process generating the data beyond what is contained in the data themselves. Such choices are necessary whether the analysis is Bayesian or frequentist, and arise in the specification of the likelihood as well as the prior. It is therefore misleading to say that priors make Bayesian analyses uniquely subjective.

A fundamental property of most Bayesian models is that, as the amount of data increases, the influence of the prior on the results declines. Moreover, with large, high-quality datasets and well-identified models, it becomes feasible to use vague or non-informative priors. Because demographers often use large, high-quality datasets, they can often avail themselves of such priors, and be confident that their results are not overly sensitive to the prior selection. This approach – “objective Bayesianism” – allows the data to dominate the estimation process, and is argued to have desirable epistemological properties, especially from the point of view of the links with formal logic (Williamson 2013).

Besides the criticism of subjectivity, there are several examples of misconceptions related to particular demographic applications of Bayesian methods. The recent Bayesian population forecasts of the UN (Gerland et al. 2014) have been criticised as offering not much beyond simple extrapolations from the past into the future, devoid of expert input (see for example the voices cited by Stukenberg 2014). This criticism is misplaced: Bayesian methods allow for including expert knowledge formally and consistently in the models, through prior distributions of the parameters. This has been done in the UN work, by including informative priors intended to offer guidance on long-term asymptotic conditions in the absence of reliable empirical evidence (e.g. Raftery et al. 2013). The prior information has been additionally augmented by calibration of forecasting uncertainty based on the performance of the same model used on various shortened data series from the past (see also the next section).

On the other hand, purely expert-based approaches either ignore other sources of information and uncertainty in the forecasts, such as past data series, or are overly reliant on the knowledge of experts, which on its own – without data – may be problematic (for a general discussion, see Lawrence et al. 2006; demographic examples of problems with purely expert-based assumptions and the related biases are discussed for example in Oeppen and Vaupel 2002 or Keilman 2008). Such methods also usually do not attempt a formal calibration of uncertainty.

In short, Bayesian analysis indeed contains explicit subjective elements, but presents the subjectivities in a transparent way, and enables combining different types of information – data and expert knowledge – in a coherent manner. Moreover, the results are described by whole probability distributions, rather than just mean or median points, and as such convey much more information. An example of a practical misconception related to this aspect can be found in Brücker and Siliverstovs (2006), where Bayesian methods have been simply used as an alternative method of estimating point values of model parameters, which are presumably posterior means, although the prior and posterior distributions are not discussed explicitly.

Despite the subjectivity of some elements of the Bayesian approach, its transparency and coherence are increasingly appealing to statistical demographers in many areas of application, examples of which are reviewed and discussed next.

3. Key areas of demographic application

Despite the existence of Bayesian demographic studies in the 18th century (Laplace 1781), demography has largely remained a deterministic exercise until the middle of the 20th century. As argued by Courgeau (2012), this was largely due to the proliferation of census-based information. In particular, availability of population-level figures have led to the problems of variability being largely ignored. Only after the Second World War did the challenges of uncertainty and probability began to make a comeback within mainstream demography (Courgeau 2012). As mentioned by Alho (1999), there are examples of pioneering quasi-Bayesian population research from that period, notably the work of Leo Törnqvist's group on Finnish population forecasts (Hyppölä, Tunkelo, and Törnqvist 1949). A landmark paper by Alho and Spencer (1985) provided an overview of probabilistic population studies until early 1980s, and also made a case for Bayesian analysis, labelled as a "more elegant" solution to the population forecasting problem (Alho and Spencer 1985, p. 314).

Since the 1990s Bayesian demography has been undergoing a revival, with a near-exponential growth in related research output, especially in the past decade. In this section, we examine the underlying driving forces behind this trend, for the three areas of application: forecasting, limited data, and highly-structured and complex models. For these three areas, we review the motivations behind the use of Bayesian methods, and suggest some possible ways through which their applications may expand into new demographic territories over the course of the 21st century.

3.1. Forecasting

Population forecasts, and their less accountable cousins, population projections, are crucial for many areas of public policy, and a variety of planning applications, in the public, private and third sectors of the economy. Booth (2006, p. 548) refers to forecasting as "the public face of the [demographic] profession". Partly in response to the demand for forecasts, and partly thanks to the adoption of methods from other disciplines of science, the past 30 years have witnessed very rapid methodological developments in this area (see Booth 2006 for an excellent overview).

As for successes in predicting populations, Xie (2000, p. 670) observes that demography has been "fundamental in forecasting future states of human societies … with a high degree of confidence". This can be attributed largely to the persistent regularities of some demographic processes, as well as to the deterministic nature of the underlying mechanism of population renewal. The latter manifests itself through population accounting and population momentum, with a lot of information being already embedded in the age structures. Its presence marks a clear contrast to other social science areas, and increases the predictability of populations, especially in the short term (Keyfitz 1981). Besides, as pointed out by Morgan and Lynch (2001), demography is heavily reliant on empirical data, which further strengthens the knowledge base of produced forecasts.

However, the issues with forecast errors are also well known, and have been extensively discussed, together with the need to communicate the errors to the forecast users (e.g. Keyfitz 1981; Keilman 1990; Alho and Spencer 2005). There are many sources of forecast uncertainty, and from the point of view of the forecasting process, six of these come especially to the fore – uncertainties in baseline data, model specification, model parameters, expert judgement on the assumptions, length of the forecast horizon, and the inherent randomness of the processes being forecast. This diversity calls for an integrated and coherent treatment of different uncertainty types, ideally within a single approach.

With regard to forecast uncertainty, a natural advantage of Bayesian methods is that all different sources of error can be potentially embedded in a joint probabilistic forecasting model. Parameter uncertainty is reflected through prior distributions, which themselves can include the formally elicited expert uncertainty. Specification of the inherent randomness of the process forms a part of the model design, which can also include additional terms for baseline data errors. The issue of model specification can be addressed by adopting the Bayesian model selection and averaging (Raftery 1995), as discussed in Section 2, and this has been done in several demographic applications to estimation (Murphy and Wang 2001) and forecasting (Bijak 2010; Abel et al. 2013ab).

This ability of Bayesian modelling to combine different uncertainties in a coherent manner also allows for other approaches to population forecasting to be expressed in a Bayesian form, such as those based purely on expert opinion (Lutz et al. 2004), or purely data-driven time series models (for an overview, see e.g. Alho and Spencer 2005). Such forecasts can be then interpreted as being conditional on the lack of uncertainty in the remaining aspects of the modelling process, such as parameters or models. In other words, such models could be seen as special cases of a fully Bayesian approach. Recently, an attempt to reconcile the purely expert-based and Bayesian approaches has been undertaken by Billari, Graziani and Melilli (2014), who proposed to give Bayesian interpretation to expert-based predictions, and to treat expert inputs as data. Still, this approach is subject to the same caveats as other expert-based propositions, as it does not formally incorporate the information carried by the time series of data.

Irrespective of a particular modelling approach, the outcome of the whole forecasting process in the Bayesian framework – a set of predictive distributions – not only offers a natural description of the overall uncertainty in the language of probabilities, but also follows directly from the joint statistical model and is very easy to derive and interpret (see e.g. Bernardo and Smith 2000). The forecast uncertainty can be evaluated by assessing *ex post*, for example based on truncated series or against external data, how well-calibrated the probability distributions are. In its basic form, this exercise looks at whether the *ex post* distributions of forecast errors – obtained by comparing the forecasts with actual observations for a subset of data, such as a truncated time series – at least roughly match the ones predicted *ex ante* from the statistical model. In some cases, additional prior information gives an advantage in that respect: for example, assuming *a priori* low predictability of migration can often lead to better calibrated forecasts (Bijak 2010). At a more general level, Gneiting and Raftery (2007) have proposed several scoring rules which can be applied to produce a numerical summary of the calibration, as well as precision (*sharpness*) of the error distribution.

Furthermore, the joint probabilistic model can serve as a basis for deriving other conditional forecasts in a natural and straightforward manner, by assuming that certain parameters or processes are known without error. This can be interpreted as a probabilistic equivalent of “what-if” scenarios (Bijak 2010). Moreover, under the Bayesian framework, predictions can be sequentially updated as soon as new information becomes available (Dawid 1984). Finally, probabilistic forecasts can serve as input for a formal decision analysis aimed at supporting the forecast users in their planning decisions, assuming that the loss functions can be at least approximately elicited from the decision makers (Alho and Spencer 2005; Bijak 2010). This possibility remains still largely unexplored in demographic practice.

A further example where the Bayesian approach is useful is age-period-cohort (APC) modelling, with clear forecasting applications, e.g. by using models from the extended Lee and Carter (1992) family. The problem of the lack of identifiability of individual age, period and cohort factors in a linear setting has been long known (see Fienberg 2013 for a recent overview and critique of some of the proposed approaches). Given the need for "resolving the APC dilemma using substantive judgment and knowledge" (Fienberg 2013, p. 1982), Bayesian methods offer a transparent and flexible alternative, by including strong prior information on particular factors. Examples of Bayesian APC models are given by Nakamura (1986) and Berzuini, Clayton, and Bernardinelli (1993), with a dedicated software "BAMP" designed by Schmid and Knorr-Held (2007). It is worth noting that the APC analysis is useful both for forecasting and reconstructing past populations, as discussed in Section 3.3.

In terms of examples of forecasting, besides the pioneering work of Törnqvist's group in Finland (Hyppölä, Tunkelo, and Törnqvist 1949), the main developments of Bayesian population predictions have begun in the 1990s, with the seminal papers by Bernardo and Muñoz (1993), on component-level population forecasts for the Valencia, Spain, and by Daponte, Kadane and Wolfson (1997) on reconstructing the Iraqi Kurdish population under Saddam Hussein's regime. For Bayesian demographic techniques applied to non-human populations, notable examples include bowhead whales (*Balaena mysticetus*, Raftery, Givens, and Zeh 1995) and northern spotted owls (*Strix occidentalis caurina*, Clark 2003). Since then, both the methodological developments and the areas of application have been gaining pace, with many publications recently, especially since 2010. In particular, all three components of population change have received due attention from demographers applying Bayesian methods.

Fertility forecasts have been presented for example by Tuljapurkar and Boe (1999) for the United States, and by Alkema et al. (2011) for the whole world, the latter in the context of the ongoing work on the United Nations World Population Prospects (UN WPP). Recently, Schmertmann et al. (2014) proposed a method for predicting cohort fertility for countries featuring in the Human Fertility Database.

Bayesian mortality forecasts have been chiefly based on variants of the Lee and Carter (1992) bi-linear model, examples of which include Czado, Delwarde and Denuit (2005) for France, and Girosi and King (2006) for the United States, and Li (2014) for China and Taiwan. In particular, Girosi and King's book – despite a somewhat misleading title, being entirely devoted to mortality – addresses some of the shortcomings of the Lee and Carter model by analysing mortality by cause of death and incorporating covariates into the models. As alternatives, Lynch and Brown (2001) have compared three Bayesian models for compression and deceleration of mortality rates: based on the classical Gompertz model, logistic function and a trigonometric transformation (arctangent). Lynch and Brown (2010), in turn, have extended Sullivan's method for reconstructing multistate life tables from period data, with potential direct applications in multistate projections or forecasts. Finally, Raftery et al. (2013) have produced forecasts of life expectancy for the whole world, also in the context of the UN WPP work; recently extended to a two-sex framework (Raftery, Lalic, and Gerland, 2014).

For forecasting migration, a range of Bayesian time-series methods has been proposed, firstly by Gorbey, James and Poot (1999) for flows between Australia and New Zealand; subsequently for a selection of European flows by Bijak (2010), and Bijak and Wiśniowski (2010); and then by Abel et al. (2013a) for environmental migration to the United Kingdom. Recently, Wiśniowski, Bijak, and Shang (2014) have forecasted Scottish migration after the 2014 referendum on independence, by using a probabilistic mixture of two sets of forecasts, conditional on the referendum outcome. Given the high level of uncertainty and paucity of data on migration flows, many of these forecasts were making full use of informative priors, often based on explicitly expressed expert opinion. Finally, for other forms of mobility, examples include Congdon's (2000) Bayesian forecasts of patient flows to hospitals.

Recent years have seen several other examples of coherent Bayesian forecasts of whole populations, combining the predictions made for individual demographic components. Abel et al. (2013b) have provided a tutorial for overall time series of fertility, mortality, and migration for the United

Kingdom, without age (as in Bernardo and Muñoz 1993), but with model uncertainty. The model has subsequently been extended by Wiśniowski et al. (2015) to include age by applying a common framework, based on the Lee and Carter (1992) approach.

Finally, the existing component forecasts related to UN WPP – Alkema et al. (2011) and Raftery et al. (2013) – as well as the (so far) deterministic assumptions on migration have been combined in the prototype Bayesian population forecasts for the whole world (Raftery et al. 2012; Gerland et al. 2014). These models retain the use of double logistic curves, which the previous work by the UN Population Division has shown to work well in capturing mortality and fertility transitions, but they reduce the need for expert judgment by using formal statistical models to estimate parameters and synthesise multiple data sources. The methods use commonalities between countries to help estimate trends in countries with unreliable or missing data. Moreover, the new methods account for uncertainty in the parameter estimates and future rates, and all estimates and forecasts come with measures of uncertainty. Validation tests suggest that the models are well calibrated and the confidence intervals produced by the models accurately reflect the true level of uncertainty. The computer to code to implement the new methods is available through a set of R packages, such as bayesPop, bayesTFR, bayesLife and bayesDem (for links, see Raftery et al. 2012).

Development of methods for Bayesian forecasting of migration rates for the whole world is currently in progress (Azose and Raftery 2014). The picture of future uncertainties that emerges from the Bayesian methods so far has been quite different from that suggested by the traditional scenario-based projections issued by the UN Population Division. Once the work related to the UN World Population Prospects is completed, with well-calibrated assessment of uncertainty stemming from all three components of population change, this will become a very significant step towards including Bayesian modelling in the methodological state of the art related to population forecasting.

3.2 Limited data

Besides forecasting, Bayesian methods are also very well suited to two types of limited data frequently encountered in demography: data that are sparse, and data that are unreliable or incomplete.

Demographers often work with datasets, such as censuses or vital registration data, which are orders of magnitude larger than the datasets used by other social scientists. However, the ever-increasing demand for disaggregated statistics, such as life tables for small areas, or fertility schedules for ethnic groups, means that demographers often deal with sparse data. In other words, once events have been cross-classified along multiple dimensions such as age, sex, region, time and population size, the number of events within each cell or the number of person-years of exposure can be small. In a small country, it is common to encounter small cell sizes even when working at the national level. In New Zealand, for instance, it is common for annual deaths of five-year-olds to equal zero. Small numbers would pose no methodological problems if, like survey statisticians, demographers were mainly interested in ‘finite population’ quantities, such as the number of deaths that actually occurred. However, demographers are typically interested in ‘super-population’ quantities such as “current mortality conditions” (Vaupel 2002, p. 366). For instance, New Zealand demographers do not conclude that five-year-olds have no chance of dying in years when no five-year-olds in fact die.

One approach traditionally used in demography to analyse sparse data is to combine cells until the number of events within each cell is large enough that random variation no longer dominates. An alternative is to retain detailed classifications but to draw a curve through the observed rates that tries to pick out genuine changes in the underlying rates while filtering out random fluctuations. A curve can be obtained by fitting a low-dimensional parametric model, such as a model life table, or a parametric fertility, mortality or migration schedule, by ‘graduating’ the rates, or by fitting some sort of general-purpose smoother, such as a spline (Preston et al. 2000; Keyfitz and Caswell 2005). The challenge with all these approaches is finding an appropriate balance between robustness and sensitivity. Aggregations that are appropriate for a small region may be unnecessarily coarse for a large region, for instance, and a spline that works well over most ages may be inappropriately smooth in the young adult groups.

Bayesian hierarchical models provide an elegant solution to the problem of balancing robustness and sensitivity. A hierarchical model is set up with likelihood and prior that pull in opposite directions. The likelihood pulls towards the ‘direct’ estimate, that is, the estimate obtained by simply dividing the observed number of events by the population at risk. The prior pulls towards the predictions from a model for the underlying rates, which, in demographic applications, typically contains a smooth function of age. The posterior for the hierarchical model is, as always, a compromise between likelihood and prior, with the likelihood receiving more weight in cells where there are relatively more observations, and the prior receiving more weight in cells where there are fewer observations. The result is that the hierarchical model is sensitive where it can be, and robust where it needs to be (Gelman and Hill 2007).

[Figure 2 here]

Figure 2: Estimates of annual emigration rates for females aged 30–34 in three selected regions of New Zealand, 1992–2014

Notes: The light grey bands are 95% credible intervals, the dark grey bands are 50% credible intervals, the pale lines are posterior medians, and the black lines are direct estimates

Source: Customised tabulations from Statistics New Zealand

Figure 2 illustrates these points with estimates of emigration rates from a Bayesian hierarchical model for three New Zealand regions with varying population sizes. The data on numbers of emigrants come from departure cards filled out by everyone entering the country. In 2014, the female populations of Auckland, Porirua, and Mackenzie were 776,900, 27,300, and 2,000. The model gives the most weight to the prior, and the least weight to the direct estimates, in Mackenzie, where the number of observations is smallest and the direct estimates are most erratic. In Auckland, in contrast, the model estimates are almost indistinguishable from the direct estimates.

In recent years, the number of applications of hierarchical Bayes models to demographic questions has been growing rapidly. Representative examples include estimation of marriage rates, including the probability of never marrying, for seventeenth century Italy (Rosina 2006); a study of geographical variation in mortality in modern Italy (Divino et al. 2009); a model of age at first birth in Nigeria (Gayawan and Adebayo 2013); estimation of mortality rates by migration status in New Zealand (Richardson et al. 2013); estimation of fertility rates in over 5,000 municipalities in Brazil (Schmertmann et al. 2013); and an analysis of intergenerational 'transmission' of fertility patterns (Osiewalska 2013). Some Bayesian principles, including heavy reliance on prior information, have been also utilised by Schmertmann (2012) in designing a calibrated spline method for estimating the age patterns of fertility.

Finally, if existing data are incomplete or unreliable, the analysts have little choice but to bring in extra information not contained in the data themselves. One example is the IMEM (*Integrated Model of European Migration*) project, which produced a migration flow matrix for 31 European countries (Raymer et al. 2013). The input data varied substantially in their accuracy, coverage, and definitions. Only by incorporating information on accuracy, coverage, and definitions was it possible to produce sensible estimates. Such information did not generally exist in quantitative form. Instead it was elicited from experts as prior distributions (Wiśniowski et al. 2013). Techniques for eliciting prior distributions that accurately reflect the beliefs of subject matter experts have received substantial research, reviewed in O'Hagan et al. (2006). Having a standard method for incorporating extra information of this type is an important strength of Bayesian analysis. It becomes easier to incorporate such information, and increases the transparency of the process, facilitating criticism and replication.

Other applications of Bayesian modelling to problematic data involve, for example, models with constraints on parameters, such as in the proportional hazards model applied to breastfeeding durations (McDonald and Prevost 1997); detecting underreporting of births in China by using discrete-time hazard models and change-point regressions (Merli and Raftery 2000); and adjusting published fertility rates for specific subpopulations, such as the US Hispanics (Rendall, Handcock, and Jonsson 2009). In historical demography, Kasakoff, Lawson and Van Meter (2014) have explored genealogical data from the 19th century to disentangle different correlates of individual wealth in north-eastern United States. As was the case with forecasting, applications of Bayesian methods to estimating demographic parameters and features of different populations are currently developing rapidly in many areas of population sciences.

3.3. Highly-structured and complex models

The majority of models traditionally used for estimation and prediction are either static, or are characterised by simple, one-directional dynamics, such as in time series analyses. Real demographic systems, however, typically include feedback loops, constraints, and rates that change over age, time, and space. Data on different components of the system may have to be assembled from different sources, each with its own biases and different levels of completeness. Moreover, demographic systems typically give rise to many types of uncertainty, arising from incomplete knowledge of historical trends or causal mechanisms, or from random variation in disaggregated counts. No model can ever capture all these complexities. However, Bayesian methods, and particularly numerical techniques such as MCMC, allow demographers to build models that would be intractable using traditional statistical methods.

One example of a highly-structured model made possible by Bayesian methods is the framework for subnational population estimation presented in Bryant and Graham (2013). The framework is summarised in Figure 3. The dark rectangles in the figure represent known quantities, and the light rectangles represent unknown quantities. Arrows represent probabilistic relationships.

The core of the model is a demographic account (Rees 1979; Stone 1984). The account describes all the demographic stocks and flows of interest, linked by accounting identities, and disaggregated by variables such as age, sex, region, and time. In a typical application, the main aim of modelling is to infer values for the demographic account.

[Figure 3 here]

Figure 3: An example of a Bayesian framework for subnational population estimation

Source: Adapted from Bryant and Graham (2013, p. 594)

Entries within an account typically exhibit strong regularities: for instance, mortality rates have distinctive age profiles, and populous regions tend to stay populous. The overarching 'demographic model' captures these regularities. Including the demographic model in the framework means that values within the demographic account that are more demographically plausible receive higher implicit weights. In turn, the individual data models 1, 2, …, K capture the relationships between the corresponding K datasets and the demographic account. For instance, a data model might state, in mathematical form, that data from the deaths registration system capture 90% of deaths on average, with this relationship varying over age and region.

This framework has some important advantages over more traditional approaches to population estimation. All outputs from the model come with measure of uncertainty. These measures include uncertainty from random variation in demographic events or reporting, uncertainty about demographic rates, and uncertainty about the reliability of the data sources. Because the model 'predicts' the contents of each datasets from the contents of the demographic account and the corresponding data model, it is easy to deal with missing data, or data that are less detailed than the account: the relevant parts of the demographic account are simply omitted or aggregated before they are supplied to the data model. Some of the most *ad hoc* and time-consuming parts of the population estimation process are thereby avoided. Because the approach uses statistical models to carry out tasks such as data evaluation that are traditionally accomplished through expert judgment, it is more transparent, and more amenable to replication and automation.

Similar ideas of using Bayesian methods to build complex demographic models for the purpose of reconstructing populations, have been discussed by Wheldon et al. (2013). Their work, illustrated on the example of the population of Burkina Faso, is generic and combines the elements of population reconstruction, and dealing with missing data, which is a major issue, especially for many developing countries.

A second example of highly-structured demographic modelling via Bayesian methods is the statistical analysis of fecundity and conception, which involves several structural challenges. Measures such as the number of fertile days per cycle vary from woman to woman, as well as varying over time for the same woman. Such variability may be viewed as a noise to be smoothed away, or as an object of interest in itself. Predictors such as frequency of intercourse or daily temperature contain substantial measurement error. Some data are ‘censored’, so that, for instance, the lengths of birth intervals that were still open when recording finished are not known. Fecundity itself follows a complicated non-linear trajectory. Using Bayesian tools such as hierarchical models and MCMC, scholars have made considerable progress in all these areas (e.g. Dunson and Weinberg 2000; Dunson, Colombo, and Baird 2002; Dunson and Colombo 2008; McDonald et al. 2011).

A third example of demographic treatment of complexity is related to computational simulation models that are not tractable analytically due to their complex structures, presence of non-linear relationships, and possible feedback loops. Here, the associations between model inputs and outputs need to be unravelled. Several approaches have been proposed, including *Bayesian melding* by Poole and Raftery (2000), later extended by Ševčíková, Raftery, and Waddell (2007), and statistical *emulators* – a special class of *meta-models* of the underlying complex computational models. The emulators are usually based on Gaussian processes, which are also typically analysed within a full Bayesian framework (Kennedy and O’Hagan 2001; Oakley and O’Hagan 2002). A simplified approach also exists – so-called *Bayes linear* – whereby uncertainty beliefs are reflected by the measures of mean and dispersion, rather than by the whole probability distributions (e.g. Vernon, Goldstein, and Bower, 2010).

There are trade-offs between Bayesian melding on the one hand, and emulator-based or Bayes linear approaches on the other: the two last-mentioned methods provide only approximate solutions, but may be computationally less expensive to run, especially in the Bayes linear case. Interesting arguments for both approaches can be found for example, in D. Poole’s discussion of Vernon, Goldstein, and Bower (2010), and the authors’ subsequent rejoinder.

Demographic examples of Bayesian studies of complex computational models include the applications of Bayesian melding in analysing the dynamics of HIV epidemics (Alkema, Raftery, and Clark 2007; Alkema, Raftery, and Brown 2009; Clark, Thomas, and Bao 2012; Sharrow et al. 2013), and in agent-based models of transport networks (Ševčíková, Raftery, and Waddell 2007). Gaussian process emulators have been also used to study an agent-based model of marriage formation (Bijak et al. 2013). Agent-based models are micro-level computational simulation models, whereby the individual units of analysis (*agents*) interact with each other and with their environment according to some rules driving their behaviour. These interactions yield macro-level patterns that can be then compared with the empirical observations (for details, see Billari and Prskawetz 2003). Given the increasing recognition of complexity of the population processes, agent-based models and their analysis through meta-models constitute a very promising path for further enquiries involving the application of Bayesian methods.

4. Bayesian demography in the 21st century

There are several distinct features of demography that make it especially suited for Bayesian modelling. First of all, the revived interest in uncertainty, and the gradual shift from deterministic to probabilistic perspectives (Alho and Spencer 2005; Courgeau 2012), naturally point to Bayesian methods because of its ability to combine many different uncertainties via probability distributions. Second, as discussed above, in applied demography and population statistics there is often a need to combine several data sources, incorporate additional information and constraints, include expert knowledge, and deal with sparse or messy data, all in a coherent manner. Third, as argued by Courgeau (2012), the increasingly popular multilevel paradigm, combining the analysis at the levels of individuals, groups, and whole populations, also naturally lends itself to the use of Bayesian methods. This is even truer for a statistical analysis of complex computational models, discussed in the previous section.

On the other hand, demography also has a lot to offer to the methodology of Bayesian statistics. First, its strong empirical orientation (Xie 2000; Morgan and Lynch 2001), combined with a uniquely detailed knowledge of some of the underlying mechanisms under study, such as population renewal, can offer a unique testing ground for many Bayesian methods. Second, demographers have already come up with solutions to some specific estimation problems. An example here may be the Lee-Carter (1992) model of mortality surfaces, and its various extensions and revisions, such as those proposed by Girosi and King (2008), who have emphasised the need to preserve good ideas from demography which may be applicable in wider contexts. Third, given its policy relevance, demography can offer statisticians a unique applied area for experimenting with the user engagement, communication of uncertainty, and public understanding of statistics. Here, the only coherent framework covering different aspects of such user engagement – from uncertain models, estimates and forecasts, to the informed support of policy decisions, and analysis of their possible consequences – is Bayesian.

So, why should more demographers use Bayesian methods, and what would be the value added offered by such approaches to the 21st century population scientists? First, we believe that the notions of risk and uncertainty will likely gain more ground in social science, as an honest and scientifically sound way of describing the social reality. A coherent description of uncertainty will become crucial: end users will no longer remain satisfied with point estimates or purely qualitative indications of reliability. Second, even in the age of Big Data, such established techniques as Bayes's theorem will facilitate the analysis by helping to avoid making false positive claims, which are one of the dangers of using large-scale data mining techniques. Third, as argued in Section 3, the analysis of complex social phenomena, for example through computational computer models, will also need to recourse to a formal language of describing the underlying mechanisms in a coherent, probabilistic fashion. This work is still in its infancy although some promising ideas have already emerged, such as the use of recursive Bayesian networks as formalism for causal mechanistic modelling (Casini et al. 2011).

Recent rapid developments in Bayesian demography provide grounds for optimism. As argued by Courgeau (2012), the history of demography and population studies in the past have been largely cumulative, with new paradigms or perspectives, such as longitudinal, event-history, or multi-level modelling supplementing old ones, rather than replacing them completely. We believe there are several important reasons why this cumulativity constitutes a very strong case for furthering the Bayesian perspective in the 21st century demography.

Most importantly, as discussed above, the Bayesian platform can be seen as a more general framework encompassing some other probabilistic approaches, especially the purely data-based and purely expert-based ones, in a coherent manner. As such, Bayesian approaches are complementary to many traditional demographic methods, rather than being in direct competition with them, since these traditional methods can be productively re-expressed in Bayesian terms. The very essence of Bayesian inference is based on the notion of continuity – constant updating of beliefs in the light of new evidence, in line with the main tenets of the scientific method. It allows incorporating new insights – quantitative data, as well as expert views, some of which may be qualitative – rather than reinventing the existing knowledge base or, worse, ignoring it. The theory, methods, as well as the logical and philosophical underpinnings of Bayesian statistics are also continually developing, partially in response to various contemporary scientific challenges (for a discussion, see e.g. Williamson and Corfield 2002).

There are several challenges for the further developments of Bayesian demography, and for its practical applications. In our view, three of them are of special importance. Firstly, there is a lack of training in Bayesian methods at the undergraduate and postgraduate levels in the social science and statistics curricula. For Bayesian demography to gain further momentum, more training opportunities – some of which already exist, especially in the form of elective courses – should be offered to practising demographers and social statisticians. Second, current computational methods are mainly targeted at academic users, rather than practitioners: although, as discussed above, there are few general-purpose platforms, many Bayesian solutions are bespoke. Until recently there was a distinct lack of Bayesian modules in mainstream software, with a current exception of R (Park 2015) and SAS (Sas Inc. n.d.).

A separate, important challenge is the communication of uncertainty to the users so that they can make the most of the information provided to them via probability distributions (Bijak et al. 2015). This problem is not limited to demographic or social science applications, and there is already considerable work in this area more broadly (e.g. Spiegelhalter, Pearson, and Short 2011). A related issue is related to the practical utility of the probabilistic outcomes of population estimates and forecasts. A promising practical extension of Bayesian estimation and forecasting consists of a formal decision analysis, introduced briefly in Section 2.3, which could be utilised for policy and planning purposes to mitigate the expected uncertain outcomes.

Contemporary applied Bayesian statistics does not emphasise the connection with the decision analysis, and one does not have to commit to utility theory to be a Bayesian – for example prominent orthodox Bayesians such as Lindley (1992) argue that the main outcomes of any Bayesian analysis are whole distributions rather than point estimates. However, the applications of the Bayesian decision theory have been suggested, also in the context of demography, to solve practical problems. In particular, this approach could help select appropriate values from the probability distributions, which could be then used for policy or planning purposes (see Alho and Spencer 2005; Bijak 2010). The main methodological challenge here consists of the elicitation of the utility or loss function from the users of the analysis. Hence, despite the scarcity of concrete applications so far, the framework is there if it is needed, and in our view this is one of the important directions of the practical of applied Bayesian demography in the future.

We think that these features of Bayesian statistics – and Bayesian demography – are really remarkable in a 250-year-old invention, and that they bear promise of many further exciting developments in the applied area of population sciences throughout the 21st century.

Acknowledgements

This work was supported by the Economic and Social Research Council (ESRC) under the Centre for Population Change Grant ES/K007394/1. We are grateful to Daniel Courgeau, Jonathan J. Forster, Patrick Graham, Peter W.F. Smith, Arkadiusz Wiśniowski, and three anonymous reviewers, and the Editors of Population Studies, for all discussions, comments and suggestions. All the views and interpretations reported in this paper are those of the authors, and should not be attributed to any institution with which we are or were affiliated. All the remaining errors are ours.

References

Abel, Guy J., Jakub Bijak, Allan M. Findlay, David McCollum, and Arkadiusz Wiśniowski. 2013a. Forecasting environmental migration to the United Kingdom: An exploration using Bayesian models, *Population and Environment* 35(2): 183–203.

Abel, Guy J., Jakub Bijak, Jonathan J. Forster, James Raymer, Peter W. F. Smith, and Jackie S. T. Wong. 2013b. Integrating uncertainty in time series population forecasts: An illustration using a simple projection model, *Demographic Research* 29(43): 1187–1226.

Alho, Juha. 1999. *On Probabilistic Forecasts of Population and Their Uses*. Paper for the 52nd Session of the International Statistical Institute, Helsinki, 10–18 August.

Alho, Juha M., and Bruce D. Spencer. 1985. Uncertain population forecasting, *Journal of the American Statistical Association* 80(390): 306–314.

Alho, Juha M., and Bruce D. Spencer. 2005. *Statistical Demography and Forecasting*. Springer Series in Statistics. Berlin-Heidelberg: Springer.

Alho, Juha M., Svend E. Hougaard Jensen, and Jukka Lassila (eds.). 2008. *Uncertain Demographics and Fiscal Sustainability*. Cambridge: Cambridge University Press.

Alkema, Leontine, Adrian E. Raftery, and Tim Brown. 2009. Bayesian melding for estimating uncertainties in national HIV prevalence estimates, *Sexually Transmitted Infections* 84(Suppl. I): i11–i16.

Alkema, Leontine, Adrian E. Raftery, and Samuel J. Clark. 2007. Probabilistic projections of HIV prevalence using Bayesian melding, *Annals of Applied Statistics* 1(1): 229–248.

Alkema, Leontine, Adrian E. Raftery, Patrick Gerland, Samuel J. Clark, François Pelletier, Thomas Buettner, and Gerhard K. Heilig. 2011. Probabilistic projections of the total fertility rate for all countries, *Demography* 48(3): 815–839.

Azose, Jonathan J., and Adrian E. Raftery. 2014. *Bayesian Probabilistic Projection of International Migration Rates*. Paper for the Annual Meeting of the Population Association of America, Boston, MA, 1–3 May 2014.

Bayes, Thomas. 1763. An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society* 53: 370–418. Via: https://archive.org/details/philtrans09948070 (accessed on 20 August 2014).

Bernardo, José M., and Javier Muñoz. 1993. Bayesian analysis of population evolution, *The Statistician* 42(4): 541–550.

Bernardo, José M., and Adrian F. M. Smith. 2000. *Bayesian Theory*. Wiley Series in Probability and Statistics. Chichester: Wiley.

Berzuini, Carlo, David Clayton, and Luisa Bernardinelli. 1993. Bayesian inference on the Lexis diagram, *Bulletin of the International Statistical Institute* 55(1): 149–165.

Bijak, Jakub. 2010. *Forecasting International Migration in Europe: A Bayesian View*. Springer Series on Demographic Methods and Population Analysis. vol. 24. Dordrecht: Springer.

Bijak, Jakub, Isabel Alberts, Juha Alho, John Bryant, Thomas Buettner, Jane Falkingham, Jonathan J. Forster, Patrick Gerland, Thomas King, Luca Onorante, Nico Keilman, Anthony O'Hagan, Darragh Owens, Adrian Raftery, Hana Ševčíková, and Peter W.F. Smith. 2015. Probabilistic population forecasts for informed decision-making, *Journal of Official Statistics*, forthcoming.

Bijak, Jakub, Daniel Courgeau, Eric Silverman, and Robert Franck. 2014. Quantifying paradigm change for demography, *Demographic Research* 30(32): 911–924.

Bijak, Jakub, Jason Hilton, Eric Silverman, and Viet Dung Cao. 2013. Reforging the wedding ring: Exploring a semi-artificial model of population for the United Kingdom with Gaussian process emulators, *Demographic Research* 29(27): 729–766.

Bijak, Jakub, and Arkadiusz Wiśniowski. 2010. Bayesian forecasting of immigration to selected European countries by using expert knowledge, *Journal of the Royal Statistical Society A* 173(4): 775–796.

Billari, Francesco C., Rebecca Graziani, and Eugenio Melilli. 2014. Stochastic population forecasting based on combinations of expert evaluations within the Bayesian paradigm, *Demography* 51(5): 1933–1954.

Billari, Francesco C., and Alexia Prskawetz, eds. 2003. *Agent Based Computational Demography. Using Simulation to Improve Our Understanding of Demographic Behaviour*. Heidelberg: Physica Verlag.

Lyle D. Broemeling. 2013. *Bayesian Methods in Epidemiology*. Chapman & Hall/CRC Biostatistics Series. Boca Raton, FL: Chapman and Hall/CRC.

Brücker, Herbert, and Boriss Siliverstovs. 2006. On the estimation and forecasting of international migration: How relevant is heterogeneity across countries, *Empirical Economics* 31(3): 735–754.

Bryant, John R., and Patrick J. Graham. 2013. Bayesian demographic accounts: Subnational population estimation using multiple data sources, *Bayesian Analysis* 8(3): 591–622.

Carlin, Bradley P. 1992. A simple Monte Carlo approach to Bayesian graduation, *Transactions of the Society of Actuaries* 44: 55–76.

Carlin, Bradley P., and Thomas A. Louis. 2009. *Bayesian Methods for Data Analysis: Third Edition*. Boca Raton: CRC Press.

Casini, Lorenzo, Phyllis McKay Illari, Federica Russo, and Jon Williamson. 2011. Models for prediction, explanation and control: Recursive Bayesian networks, *Theoria* 26(1)70: 5–33.

Caussinus, Henri, and Daniel Courgeau. 2010. Estimating age without measuring it: A new method in paleodemography, *Population–E* 65(1): 117–144.

Clark, James S. 2003. Uncertainty and variability in demography and population growth: A hierarchical approach, *Ecology* 84(6): 1370–1381.

Clark, Samuel J., Jason Thomas, and Le Bao. 2012. Estimates of age-specific reductions in HIV prevalence in Uganda: Bayesian melding estimation and probabilistic population forecast with an HIV-enabled cohort component projection model, *Demographic Research* 27(26): 743–774.

Congdon, Peter. 2000. A Bayesian approach to prediction using the gravity model, with an application to patient flow modelling, *Geographical Analysis* 32(3): 205–224.

Congdon, Peter. 2010. *Applied Bayesian Hierarchical Methods*. Boca Raton: CRC Press.

Congdon, Peter. 2014. *Applied Bayesian Modelling (Second Edition)*. Chichester: John Wiley & Sons.

Courgeau, Daniel. 2012. *Probability and Social Science: Methodological Relationships Between the Two Approaches*. Methods Series vol. 10. Dordrecht: Springer.

Czado, Claudia, Antoine Delwarde, and Michel Denuit. 2005. Bayesian Poisson log-bilinear mortality projections, *Insurance: Mathematics & Economics* 36(3): 260–284.

Daponte, Beth O., Joseph B. Kadane, and Lara J. Wolfson. 1997. Bayesian demography: Projecting the Iraqi Kurdish population, 1977–1990, *Journal of the American Statistical Association* 92(440): 1256–1267.

Dawid, A.P. 1984. Statistical theory: The prequential approach (with discussion), *Journal of the Royal Statistical Society A* 147(2), 278–292.

DeGroot, Morris H. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill (2004 edition: Wiley Classics Library, New York: Wiley-Blackwell).

Dellaportas, Petros, Jonathan J. Forster, and Ioannis Ntzoufras. 2002. On Bayesian model and variable selection using MCMC, *Statistics and Computing* 12(1): 27–36.

Divino, Fabio, Viviana Egidi, and Michele Antonio Salvatore. 2009. Geographical mortality patterns in Italy: A Bayesian analysis, *Demographic Research* 20(18): 435–466.

Dunson, David B., and Bernardo Colombo. 2003. Bayesian modeling of markers of day-specific fertility, *Journal of the American Statistical Association* 98(461): 28–37.

Dunson, David B., and Clarice R. Weinberg. 2000. Modeling human fertility in the presence of measurement error, *Biometrics* 56(1): 288–292.

Dunson, David B., Bernardo Colombo, and Donna D. Baird. 2002. Changes with age in the level and duration of fertility in the menstrual cycle, *Human Reproduction* 17(5): 1399–1403.

Eayres, Daniel, and E. S. Williams. 2004. Evaluation of methodologies for small area life expectancy estimation, *Journal of Epidemiology and Community Health* 59(3): 243–249.

Fienberg, Stephen E. 2011. Bayesian models and methods in public policy and government settings, *Statistical Science* 26(2): 212–226.

Fienberg, Stephen E. 2013. Cohort analysis' unholy quest: A discussion, *Demography* 50(6): 1981–1984.

Gayawan, Ezra, and Samson B. Adebayo. 2013. A Bayesian semiparametric multilevel survival modelling of age at first birth in Nigeria, *Demographic Research* 28(45): 1339–1372.

Gelfand, Alan E., and Adrian F.M. Smith. 1990. Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85:398–409.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis. Third Edition*. Boca Raton, FL: Chapman and Hall/CRC.

Geman, Stuart, and Donald Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6(6): 721–741.

Gerland, Patrick, Adrian E. Raftery, Hana Ševčíková, Nan Li, Danan Gu, Thomas Spoorenberg, Leontine Alkema, Bailey K. Fosdick, Jennifer Chunn, Nevena Lalic, Guiomar Bay, Thomas Buettner, Gerhard K. Heilig, and John Wilmoth. 2014. World population stabilization unlikely this century, *Science* 346: 234–237.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman and Hall/CRC.

Girosi, Federico, and Gary King. 2008. *Demographic Forecasting*. Princeton University Press, Princeton, NJ.

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

Gorbey, Susi, Doug James, and Jacques Poot. 1999. Population forecasting with endogenous migration: An application to trans-Tasman migration, *International Regional Science Review* 22(1): 69–101.

Greenhill, Simon J., and Russell D. Gray. 2009. Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods. In: Alexander Adelaar and Andrew Pawley (eds.), *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*. Canberra: Pacific Linguistics, pp. 1–23.

Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics. New York: Springer.

Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach.* Third Edition. Peru: Carus Publishing Company.

Hyppölä, Jorma, Aarre Tunkelo, and Leo Törnqvist, eds. 1949. *Suomen väestöä, sen uusiutumista ja tulevaa kehitystä koskevia laskelmia* [*Calculations Concerning the Population of Finland, Its Renewal and Future Development*]. Tilastollisia tiedonantoja 38. Statistics Finland, Helsinki.

Jaynes, Edwin T. 1976. Confidence intervals vs. Bayesian intervals. In: William L. Harper, and Cliff A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science 2*. Dordrecht: Reidel, pp. 175–257 (with discussion).

Kasakoff, Alice, Andrew Lawson, and Emily Van Meter. 2014. A Bayesian analysis of the spatial concentration of individual wealth in the US North during the nineteenth century, *Demographic Research* 30(36): 1035–1074.

Keilman, Nico. 1990. *Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations*. Amsterdam and Rockland, MA: Swets and Zeitlinger Publishers.

Keilman, Nico. 2008. European demographic forecasts have not become more accurate over the past 25 years, *Population and Development Review* 34(1): 137–153.

Kennedy, Marc C., and Anthony O'Hagan. 2001. Bayesian calibration of computer models, *Journal of the Royal Statistical Society B* 63(3): 425–464.

Keyfitz, Nathan. 1981. The limits of population forecasting, *Population and Development Review* 7(4): 579–593.

Keyfitz, Nathan, and Hal Caswell. 2005. *Applied Mathematical Demography*. New York: Springer.

Kruschke, John. 2010. *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Burlington: Academic Press.

Laplace, Pierre-Simon de. 1781. Mémoire sur les probabilités, *Mémoires de l'Académie Royale des Sciences de Paris, Année 1778*: 227–332. Via: http://gallica.bnf.fr/ark:/12148/bpt6k77597p/f386 (accessed on 20 August 2014).

Lawrence, Michael, Paul Goodwin, Marcus O'Connor, and Dilek Önkal. 2006. Judgmental forecasting: A review of progress over the last 25 years, *International Journal of Forecasting* 22(3): 493–518.

Lee, Ronald D., and Lawrence R. Carter. 1992. Modeling and forecasting the time series of U.S. mortality, *Journal of the American Statistical Association* 87(419): 659–671.

Li, Jackie. 2014. An application of MCMC simulation in mortality projection for populations with limited data, *Demographic Research* 30(1): 1–48.

Lindley, Dennis V. 1972. *Bayesian Statistics: A Review*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Lindley, Dennis V. 1992. Is our view of Bayesian statistics too narrow? (with discussion) In: José M. Bernardo, James O. Berger, A.P. Dawid, and Adrian F.M. Smith (eds.), *Bayesian Statistics 4*. Oxford: Oxford University Press, pp. 1–15.

Lunn, David, David Spiegelhalter, Andrew Thomas, and Nicky Best. 2009. The BUGS project: Evolution, critique and future directions (with discussion), *Statistics in Medicine* 28(25): 3049–3082.

Lutz, Wolfgang, Warren C. Sanderson, and Sergei Scherbov. 2004. *The End of World Population Growth in the 21st Century: New Challenges for Human Capital Formation and Sustainable Development*. London: Earthscan.

Lynch, Scott M. (2007) *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Statistics for Social and Behavioral Sciences series. New York: Springer.

Lynch, Scott M., and J. Scott Brown. 2001. Reconsidering mortality compression and deceleration: An alternative model of mortality rates, *Demography* 38(1): 79–95.

Lynch, Scott M., and J. Scott Brown. 2010. Obtaining multistate life table distributions for highly refined subpopulations from cross-sectional data: A Bayesian extension of Sullivan’s method, *Demography* 47(4): 1053–1077.

McDonald, John W., and Andrew Toby Prevost. 1997. The fitting of parameter-constrained demographic models, *Mathematical and Computer Modelling* 26(6): 79–88.

McDonald, John W., Alessandro Rosina, Ester Rizzi, and Bernardo Colombo. 2011. Age and fertility: Can women wait until their early thirties to try for a first birth?, *Journal of Biosocial Science* 43(6): 685–700.

Marin, Jean-Michel, and Christian P. Robert. 2014. *Bayesian Essentials with R*. New York: Springer.

Maritz, J. S. (1970). *Empirical Bayes Methods*. London: Methuen and Co.

Merli, M. Giovanna, and Adrian E. Raftery. 2000. Are births underreported in rural China? Manipulation of statistical records in response to China’s population policies, *Demography* 37(1): 109–126.

Morgan, S. Philip, and Scott M. Lynch (2001) Success and future of demography: The role of data and methods, *Annals of the New York Academy of Sciences* 954(1): 35–51.

Murphy, Michael, and Duolao Wang. 2001. Do previous birth interval and mother's education influence infant survival? A Bayesian model averaging analysis of Chinese data, *Population Studies* 55(1): 37–47.

Nakamura, Takashi. 1986. Bayesian cohort models for general cohort table analyses, *Annals of the Institute of Statistical Mathematics* 38: 353–370.

Ntzoufras, Ioanis. 2009. *Bayesian Modeling Using WinBUGS*. Wiley Series in Computational Statistics. Chichester: Wiley.

Oakley, Jeremy, and Anthony O’Hagan. 2002. Bayesian inference for the uncertainty distribution of computer model outputs, *Biometrika* 89(4): 769–784.

Oeppen, Jim, and James W. Vaupel. 2002. Broken limits to life expectancy, *Science* 296(5570): 1029–1031.

O’Hagan, Anthony, Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Expert Probabilities*. Chichester: Wiley.

Osiewalska, Beata. 2013. Transmission of fertility pattern in mother-daughter relation – Bayesian view (A case study of Austria), *Studia Demograficzne* 163(1): 3–35.

Park, Jong Hee. 2015. CRAN Task View: Bayesian Inference. Version: 2015-07-01. R online resource. Via: https://cran.r-project.org/web/views/Bayesian.html (accessed on 29 July 2015).

Pfeffermann, Danny. 2013. New important developments in small area estimation, *Statistical Science* 28(1): 40-68.

Poole, David, and Adrian E. Raftery. 2000. Inference for deterministic simulation models: The Bayesian melding approach, *Journal of the American Statistical Association* 95(452): 1244–1255.

Preston, Samuel, Patrick Heuveline, and Michel Guillot. 2000. *Demography: Measuring and Modeling Population Processes*. Oxford: Wiley-Blackwell.

Raftery, Adrian E. 1995. Bayesian model selection in social research, *Sociological Methodology* 25: 111–163.

Raftery, Adrian E., Jennifer L. Chunn, Patrick Gerland, and Hana Ševčíková. 2013. Bayesian probabilistic projections of life expectancy for all countries, *Demography* 50(3): 777–801.

Raftery, Adrian E., Geof H. Givens, and Judith E. Zeh. 1995. Inference from a deterministic population dynamics model for bowhead whales (with discussion), *Journal of the American Statistical Association* 90(430): 402–430.

Raftery, Adrian E., Nevena Lalic, and Patrick Gerland. 2014. Joint probabilistic projection of female and male life expectancy, *Demographic Research* 30(27): 795–822.

Raftery, Adrian E., Nan Li, Hana Ševčíková, Patrick Gerland, and Gerhard K. Heilig. 2012. Bayesian probabilistic population projections for all countries, *Proceedings of the National Academy of Sciences USA* 109(35): 13915–13921.

Raymer, James, Arkadiusz Wiśniowski, Jonathan J. Forster, Peter W.F. Smith, and Jakub Bijak. 2013. Integrated modeling of European migration, *Journal of the American Statistical Association* 108(503): 801–819.

Rendall, Michael S., Mark S. Handcock, and Stefan H. Jonsson. 2009. Bayesian estimation of Hispanic fertility hazards from survey and population data, *Demography* 46(1): 65–83.

Richardson, Ken, Santosh Jatrana, Martin Tobias, and Tony Blakely. 2013. Migration and Pacific mortality: Estimating migration effects on Pacific mortality rates using Bayesian models, *Demography* 50(6): 2053–2073.

Rosina, Alessandro. 2006. A model with long-term survivors for the analysis of current-status nuptiality data, *Population Studies* 60(1): 73-81.

Rubin, Donald B. 1984. Bayesianly justifiable and relevant frequency calculations for the applied statistician, *The Annals of Statistics* 12(4): 1151–1172.

Rue, Håvard, Sara Martino, and Nicolas Chopin. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society, Series B* 71(2): 319–392.

SAS Inc. n.d. Bayesian Analysis Using SAS/STAT® Software. Online resource. Via: http://support.sas.com/rnd/app/da/Bayesian/ (accessed on 29 July 2015).

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: John Wiley and Sons (1972: revised edition, New York: Dover Publications).

Schmertmann, Carl P. 2012. *Calibrated Spline Estimation of Detailed Fertility Schedules from Abridged Data*. Rostock: Max Planck Institute for Demographic Research, and Tallahassee, FL: Center for Demography and Population Health, Florida State University.

Schmertmann, Carl P., Suzana M. Cavenaghi, Renato M. Assunção, and Joseph E. Potter. 2013. Bayes plus Brass: Estimating total fertility for many small areas from sparse census data, *Population Studies* 67(3): 255–273.

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.

Schmid, Volker J., and Leonard Knorr-Held. 2007. Bayesian age-period-cohort modeling and prediction – BAMP, *Journal of Statistical Software* 21(8), http://www.jstatsoft.org/v21/i08.

Ševčíková, Hana, Adrian E. Raftery, and Paul A. Waddell. 2007. Assessing uncertainty in urban simulations using Bayesian melding, *Transportation Research B* 41(6): 652–669.

Sharrow, David, Samuel J. Clark, Mark Collinson, Kathleen Kahn, and Stephen Tollman. 2013. The age pattern of increases in mortality affected by HIV: Bayesian fit of the Heligman-Pollard Model to data from the Agincourt HDSS field site in rural northeast South Africa, *Demographic Research* 29(39): 1039–1096.

Spiegelhalter, David, Mike Pearson, and Ian Short. 2011. Visualising uncertainty about the future, *Science* 333: 1393–1400.

Stukenberg, Timo. 2014. Weltbevölkerung im Jahr 2100: Wir werden bis zu 13 Milliarden sein, *Spiegel Online* 18 September 2014. Via: http://www.spiegel.de/wissenschaft/mensch/weltbevoelkerung-vereinte-nationen-sagen-wachstum-bis-2100-voraus-a-991874.html (as of 15 November 2014).

Tuljapurkar, Shripad, and Carl Boe. 1999. Validation, probability-weighted priors and information in stochastic forecasts, *International Journal of Forecasting* 15(3): 259–271.

Vaupel, James W. 2002. Life expectancy at current rates vs. current conditions: A reflexion stimulated by Bongaarts and Feeney's "How long do we live?", *Demographic Research* 7(8): 365–378.

Vernon, Ian, Michael Goldstein, and Richard G. Bower. 2010. Galaxy formation: A Bayesian uncertainty analysis (with discussion), *Bayesian Analysis* 5(4): 619–670.

Wheldon, Mark C., Adrian E. Raftery, Samuel J. Clark, and Patrick Gerland. 2013. Reconstructing past populations with uncertainty from fragmentary data, *Journal of the American Statistical Association* 108(501): 96–110.

Williamson, Jon. 2013. From Bayesian epistemology to inductive logic, *Journal of Applied Logic* 11(4): 468–486.

Williamson, Jon, and David Corfield. 2002. Introduction: Bayesianism into the 21st Century. In: Jon Williamson and David Corfield (eds.), *Foundations of Bayesianism*. Dordrecht: Kluwer Academic Publishers, pp. 7–22.

Wiśniowski, Arkadiusz, Jakub Bijak, Solveig Christiansen, Jonathan J. Forster, Nico Keilman, James Raymer, and Peter W. F. Smith. 2013. Utilising expert opinion to improve the measurement of international migration in Europe, *Journal of Official Statistics* 29(4): 583–607.

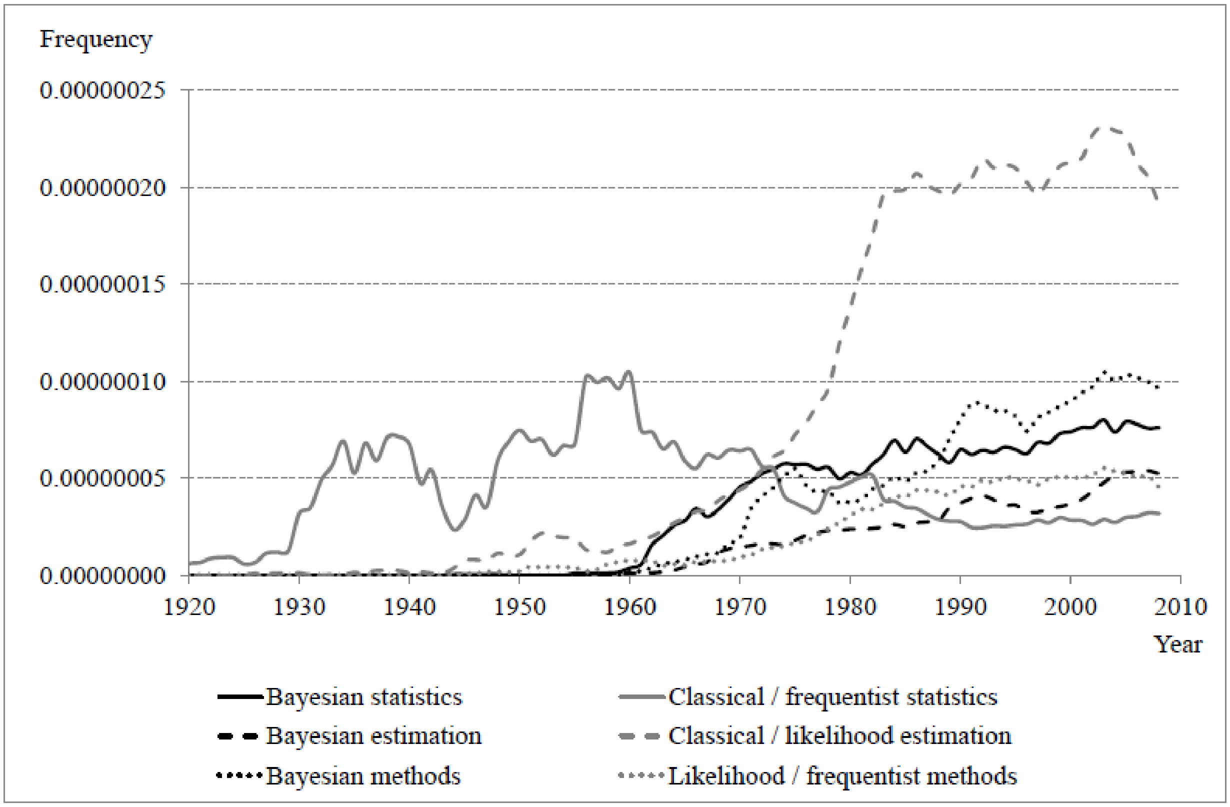
Wiśniowski, Arkadiusz, Jakub Bijak, and Han Lin Shang. 2014. Forecasting Scottish migration in the context of the 2014 constitutional change debate, *Population, Space and Place* 20(5): 455–464.

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(3): 1035–1059.

Xie, Yu. 2000. Demography: Past, present and future, *Journal of the American Statistical Association* 95(450): 670–673.

Figure captions

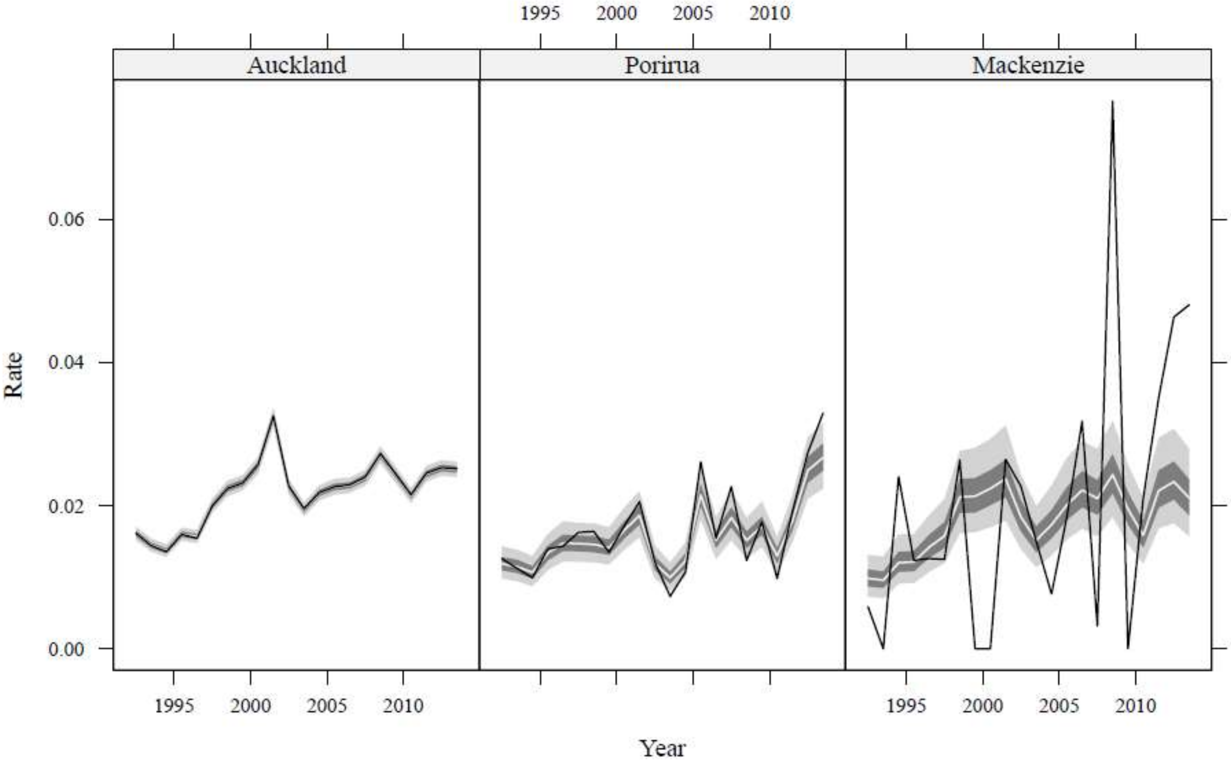
Figure 1: Frequencies for Bayesian and frequentist / likelihood search terms in Google books



Notes: For caveats regarding the use of Google Ngrams in the context of demography, see Bijak et al. (2014)

Source: Google books Ngram Viewer http://books.google.com/ngrams, English corpus, queried on 20.01.2014

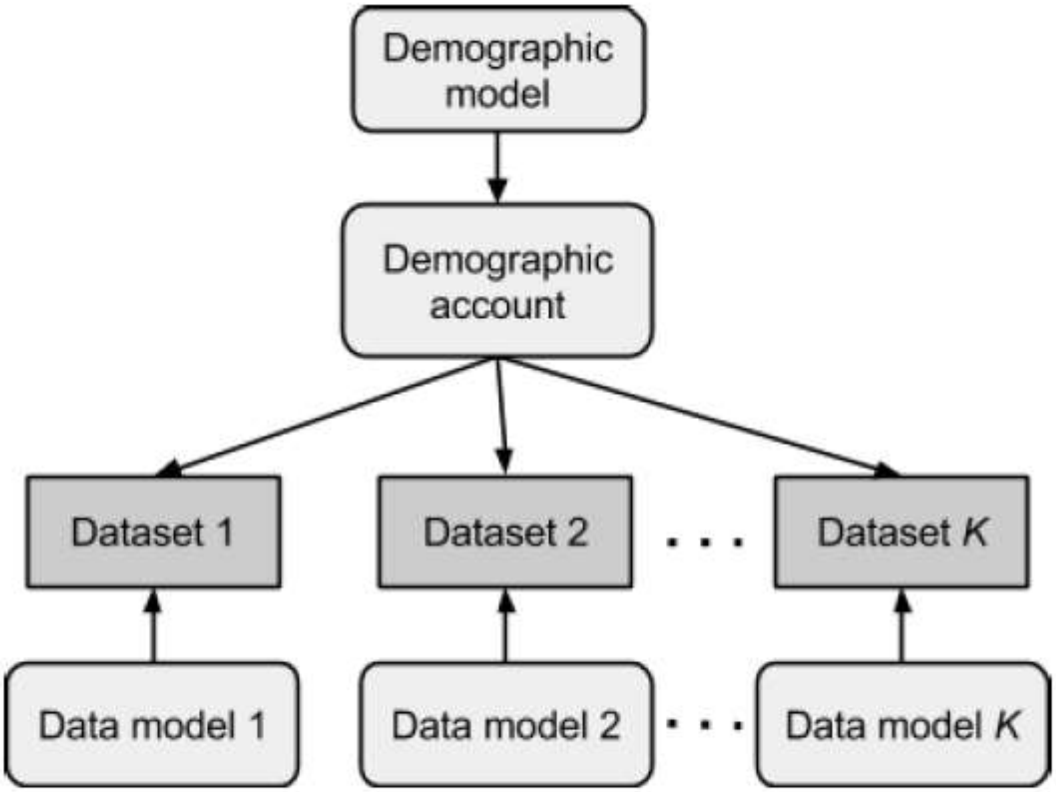
Figure 2: Estimates of annual emigration rates for females aged 30–34 in three selected regions of New Zealand, 1992–2014



Notes: The light grey bands are 95% credible intervals, the dark grey bands are 50% credible intervals, the pale lines are posterior medians, and the black lines are direct estimates

Source: Customised tabulations from Statistics New Zealand

Figure 3: An example of a Bayesian framework for subnational population estimation



Source: Adapted from Bryant and Graham (2013, p. 594)