A comparison of official population projections with Bayesian time series forecasts for England and Wales

Guy J. Abel, Jakub Bijak and James Raymer

University of Southampton

Abstract

We compare official population projections with Bayesian time series forecasts for England and Wales. The Bayesian approach allows the integration of uncertainty in the data, models and model parameters in a coherent and consistent manner. Bayesian methodology for time-series forecasting is introduced, including autoregressive (AR) and stochastic volatility (SV) models. These models are then fitted to a historical time series of data from 1841 to 2007 and used to predict future population totals to 2033. These results are compared to the most recent projections produced by the Office for National Statistics. Sensitivity analyses are then performed to test the effect of changes in the prior uncertainty for a single parameter. Finally, in-sample forecasts are compared with actual population and previous official projections. The article ends with some conclusions and recommendations for future work.

Contents

A comparison of official population projections with Bayesian time series forecasts for England and Wales 95

Abstract 95

Introduction 97

A review of population projection approaches 98

The uncertainty of the future UK population 99

Comparisons of forecasts 104

Conclusion 111

Acknowledgements 112

References 112

Introduction

In recent years there has been an increasing emphasis by national statistical offices to include uncertainty in their official population projections so that the user community has a more realistic sense of what the future might hold. For most national statistical offices this has involved the inclusion of several plausible (deterministic) projection variants based on assumptions regarding future fertility, mortality and migration in a cohort-component population projection framework. In this article, we focus on the issues and practicalities of including uncertainty from a probabilistic viewpoint.

In the 1990s there were several convincing papers arguing for the need to move away from variant-style projections to probabilistic ones. See, for example, Ahlburg and Land 1992[1]; Lee and Tuljapurkar 1994[2]; Lutz 1996[3]; Bongaarts and Bulatao 2000[4]. The advantages are clear: probabilistic projections specify the likelihood that a particular future population value will occur given a set of assumptions about the underlying probability distributions. With variant projections, on the other hand, the user has no idea how likely they are. Here, the users have to trust that the experts have provided them with plausible scenarios representing the “most likely” (the principal projection) and the extremes (the high and low population projections). Of course, in both cases, the quality of the forecasts depends on the input data, projection model and assumptions made.

Despite the advantages of a probabilistic approach and the abundance of applications[5], nearly all national statistical offices in the world still rely on deterministic variant projections to provide uncertainty[6]. However, progress is being made; the Office for National Statistics (ONS), for example, has recently been testing probabilistic models for use in its official projections[7], although their framework for including uncertainty has yet to be fully defined.

Uncertainty in population projections come from four main sources: the projection model(s), parameter estimates, expert judgments and historical data[8]. Uncertainty can also be based on the results of past projections[9] [10]. In this article we show how historical observations and model assumptions influence uncertainty, as well as the inclusion of expert beliefs regarding future patterns. We do this by applying various autoregressive time series models to population growth rates in England and Wales. Population forecasts are based on past patterns, where a long time series of data are very valuable for assessing our uncertainty for the future.

In nearly all of the probabilistic literature on population forecasting the approach has been from a frequentist (classical) perspective. We introduce a Bayesian approach, which offers population forecasters the most flexibility in terms of specifying uncertainty. Unlike frequentist models, Bayesian models allow for the integration of uncertainty expressed in prior distributions, empirical data and expert judgements. However, these models have yet to be widely applied in the population forecasting literature (see the next section in this article).

As this work is written for a general audience, we have left out the technical details of the models used to produce the Bayesian time series forecasts. For those interested in the specification of these models, refer to Abel *et al*. (2010)[11]. Also, note that this work represents some of the early efforts carried out by a team of researchers in the ESRC Research Centre for Population Change (CPC). In the future we plan to expand these ideas to more complex population models that include, for example, age, sex and state transitions that a population experiences (for example, residential, marriage, and employment).

In terms of structure, we first provide a review of standard population projection approaches and describe the current approach of ONS. This is followed by a section outlining the Bayesian approach to time series forecasting. We then compare our forecasts with official projections by ONS and to alternative forecasts based on a different prior assumption and shortened time series. Finally, we end the article with some conclusions and recommendations for future work.

A review of population projection approaches

Various typologies of macro-level population projection methods can be obtained by applying some simple criteria. In this brief review we focus on three of them: dimensionality of the problem under study (simple extrapolations of population size or growth rates, single region cohort-component models and multiregional models), the approach to uncertainty (deterministic versus stochastic) and methodology (data-driven versus expert-driven). For simplicity we assume that expert-driven methods encompass projections based on theories and expectations about the future. In many cases, however, projections combine aspects of data-driven methods and expert judgements[12]. More detailed typologies can be found in Willekens (1990)[13], de Beer (2000)[14], O'Neill *et al*. (2001)[15], Wilson and Rees (2005)[5], Booth (2006)[16] and Bijak[17].

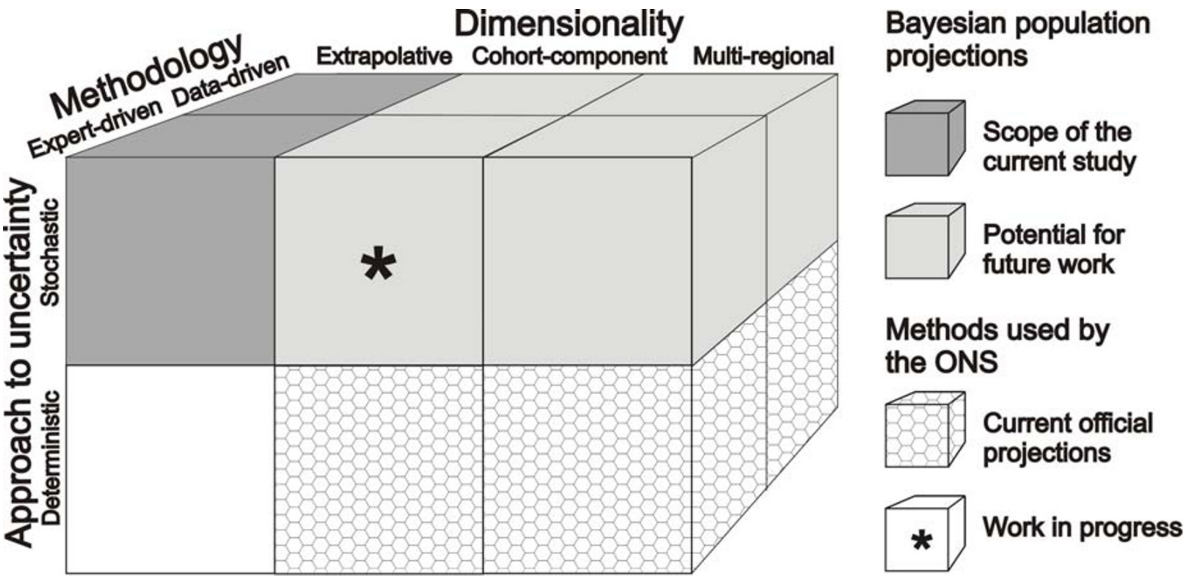
With respect to the dimensionality of population projections, the simplest models rely on the extrapolations of population size, population growth rates or crude rates related to particular components of demographic change (fertility, mortality and migration). The adding of age and sex leads to the cohort-component framework of population accounting developed by Leslie[18]. Cohort-component models are extendable by adding additional dimensions, such as spatial regions, as suggested by Rogers in his seminal work on multiregional demography[19], or subgroups, such as ethnicity[20]. Here 'multiregional' refers to all multidimensional extensions of the cohort-component model, including other multistate models. See: Land and Rogers[21]; Schoen[22] [23] and Rogers[24].

Another feature characterising any method of population projection is the approach to uncertainty. Uncertainty in projections can be ignored, described using various plausible scenarios or quantified using probabilities[14]. The deterministic scenarios can be data-driven, that is, based on simple mathematical extrapolations of past trends, or expert-driven, that is, relying mainly on expert judgement[17]. Similarly, stochastic (probabilistic) projections can be based on time series analysis or extrapolation of past projection errors[8], or based on expert opinion used to assess the future uncertainty[3]. The Bayesian methodology, advocated throughout this article, allows for combining both features in a coherent and consistent way. So far, only a handful of population forecasts have been prepared within the Bayesian framework[25] [26].

The current official population projections for England and Wales produced by ONS represent results from a deterministic model with uncertainty, not quantified in terms of probabilities, but denoted by various plausible scenarios[12]. Recently, promising attempts were undertaken to produce expert-based stochastic population projections for the United Kingdom[7]. Both the current and probabilistic work of ONS are indicated in the 'methodology cube' in **Box one** using patterns and an asterisk, respectively.

The philosophy of Bayesian statistics enables the combining of data- and expert-based approaches within a common, stochastic framework. Results are presented in this article for a simple, extrapolative example (darker shading in Box one), but our approach can be extended to include cohort-components, and eventually multiregional cases (lighter shading). In this way we believe that the Bayesian approach can complement the methodological developments currently undertaken within ONS[7].

Box one Proposed Bayesian methods and the practice of the ONS population projections



Source: Own elaboration.

For more detailed typologies, see Willekens (1990)[13], de Beer (2000)[14], O'Neill et al. (2001)[15], Booth (2006)[16] and Bijak (forthcoming)[17].

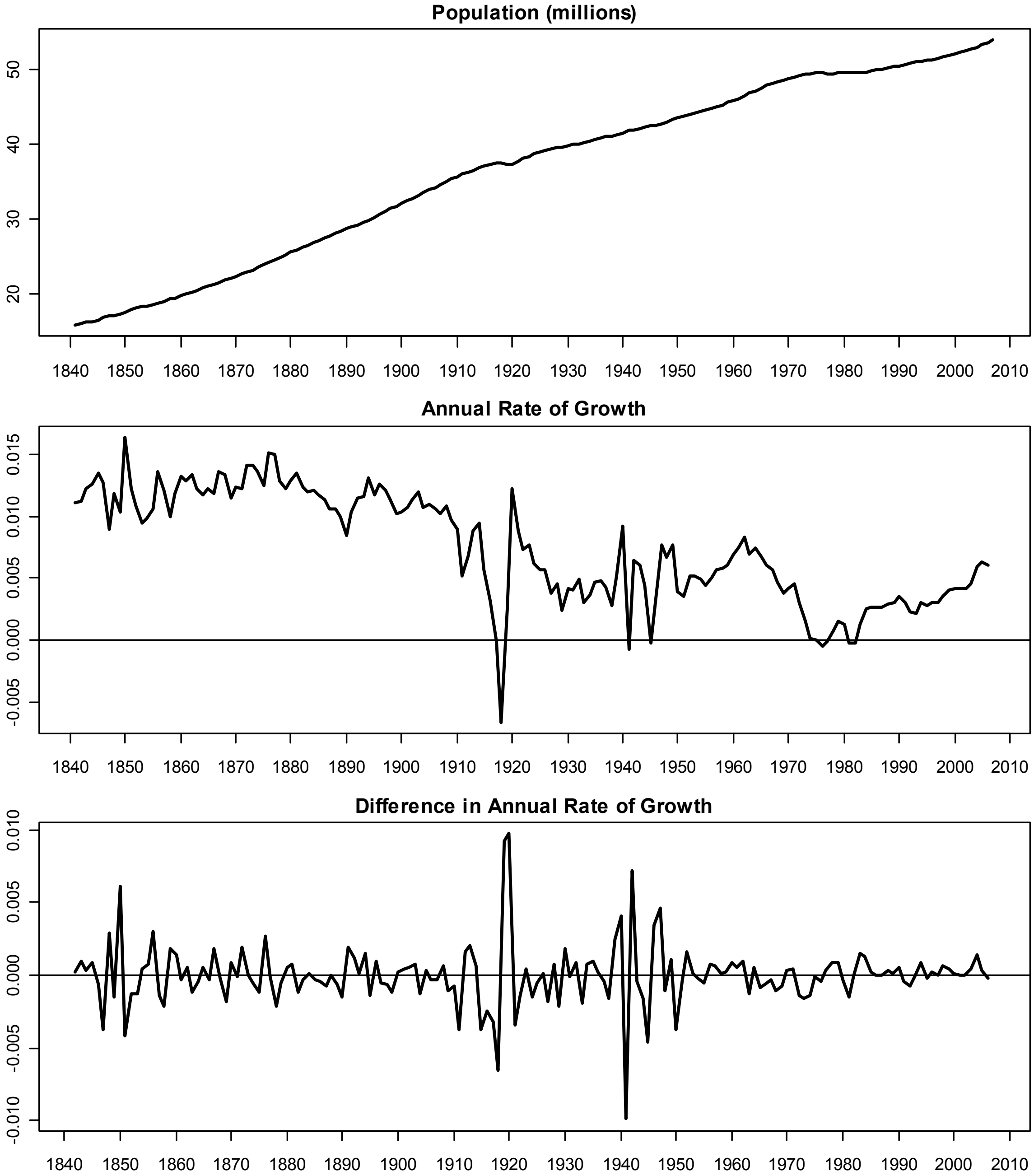
The uncertainty of the future UK population

Bayesian time series models for population forecasting are introduced in this section. First, we present the methodology for Box-Jenkins time series models. This includes transformations to data, autoregressive models for the mean process and stochastic variance models for observed data with a non-constant variance over time. Second, Bayesian methods for estimating parameter values are discussed. Here, the specification of prior distributions and model uncertainty represent the main focus. More detailed information about the models and parameter estimation can be found in Abel *et al*. (2010)[11].

Time series modelling of annual population series

Annual time series of population totals often display some form of trend or fluctuations over long time periods. To illustrate, consider the mid-year population estimates (including military personnel) obtained from the Human Mortality Database[27] for England and Wales from 1841 to 2007, as presented in the top panel of **Figure 1**. The graph clearly illustrates an increasing trend with the population rising from 15.8 million in 1841 to 53.9 million in 2007.

Figure 1 England and Wales population data, 1841-2007



Time series models for population forecasting usually concentrate on the rates of population growth over time, r_t, provided in the second panel of Figure 1. For this work the growth rates are calculated as:

$$r_t = \frac{p_{t+1}}{p_t} - 1 \qquad (1)$$

where p_t is the population total at time t. A standard requirement for fitting time series models is that the data must exhibit (weak) stationarity. This implies that both the mean and the variance of the data are constant over time. These properties are not present in the historical series of r_t shown in the middle panel of Figure 1. Instead, the series exhibits a downward trend, caused predominantly by falls in mortality and fertility rates from pre-industrial levels.

Experience suggests that, if we are to use time series models which assume stationarity, transformations of the data may be required[28]. One such transformation is to take the differences in r_t, that is:

$$y_t = r_t - r_{t-1} \quad (2)$$

and to model them instead. A plot of y_t is provided in the bottom panel of Figure 1, where a constant mean level, close to zero, is clearly illustrated. The plot also demonstrates peaks during some noticeable historical events, such as the two world wars and the 1918 influenza pandemic, which had dramatic effects on the change in the annual rate of growth. These events may lead to the conclusion that, although the series of y_t has a constant mean, it cannot be considered to be completely stationary as the variance appears non-constant over time. Models to account for this feature are outlined later in this section.

Autoregressive (AR) models have a long history of being used to forecast populations. See for example, Saboia, 1974[29]; Ahlburg, 1987[30]; Pflaumer, 1992[31]; Alho and Spencer, 2005[8]. The key feature of AR models is the inclusion of parameters for the regression of variables such as y_t, on previous values of itself, y_{t-j}, where j represents the time lag. This is commonly known as *autocorrelation*. AR models can include multiple parameters for autoregression at different time lags. For example, an AR model of order 3 is denoted as AR(3) and has autoregressive terms at lag 1, 2 and 3. Time series models also tend to have a parameter for the mean level of the process, represented by μ.

Once fitted, AR models can be used to forecast future values of the time series process. If the process considered is the change in population growth rates, y_t, (as in this article), future values of the original population growth rates, r_t , can be derived by re-arranging Equation (2). In our case the last observed population total is p_{2007} (Figure 1). Based on these data, we can derive a series of population growth rates up to r_{2006} and changes in population growth rates up to y_{2006}. Thus the first step-ahead forecast from an AR model, y_{2007}, can then be used to obtain

$$r_{2007} = y_{2007} + r_{2006} \,. \quad (3)$$

From this, we can derive the forecast of p_{2007} by re-arranging Equation (1) as

$$p_{2008} = (1 + r_{2007})\, p_{2007} \,. \quad (4)$$

Subsequent values of r_t and p_t may be calculated in the same manner, using the forecasted y_t estimated from the model.

As noted previously, historical time series of demographic data often exhibit some volatility due to events such as epidemics, wars or baby booms. This is certainly true for the data presented in Figure 1. Stochastic Volatility (SV) models allow for a non-constant variance when modelling time series data. This is done by specifying a time-dependent model for the variance, as well as the mean. Consequently, SV models can account for heterogeneity found in the demographic data, allowing forecasts to be adjusted according to the level of volatility estimated at the time the projection is made.

Bayesian time series methods

The estimation of parameters in time series models can be undertaken using a number of different methodologies. In this article we use a Bayesian methodology because both expert opinion and uncertainty in model choice can be included. See **Box two** for an introduction to Bayesian inference.

The incorporation of expert opinion has become an increasingly important input into the prediction of future populations[7]. Bayesian methods allow these opinions to be fully incorporated into the estimation procedure by specifying prior distributions in relation to the model parameters. The distributions can be set to 'flat' if the expert does not have any notions about what the parameter values should be. This results in parameter estimates that are very similar to those fitted by using classical statistical methods. On the other hand, if the expert does have some ideas about what particular parameter values should be, then that person can specify a distribution centred on these values and incorporate them directly into the estimation procedure. The result is parameter estimates that reflect the combination of the expert's prior beliefs and the empirical data.

Bayesian methods allow uncertainties in model choice to be incorporated using probability distributions representing the likelihood for each model. This allows models to be averaged across a set of plausible models, rather than selected, as is the common practice in classical statistics. This is advantageous as it is unrealistic to be sure that any particular model is the right one on which to base forecasts. In addition, Bayesian model averaging can operate across models that are non-nested, such as between AR models and SV models. Finally, Bayesian methods allow the incorporation of model uncertainty to be directly integrated with parameter uncertainty. The result is probabilistic forecasts that are likely to be more realistic than those produced with classical methods.

Box two Bayesian inference in a nutshell

The Bayesian approach to statistical inference dates back to the seminal work of an English nonconformist clergyman, Rev. Thomas Bayes (1763)[32]. The essence of Bayesian inference consists of updating *prior distributions* about the model parameters, θ, in the light of some empirical data, x. The combination of the two results in a *posterior distribution*.

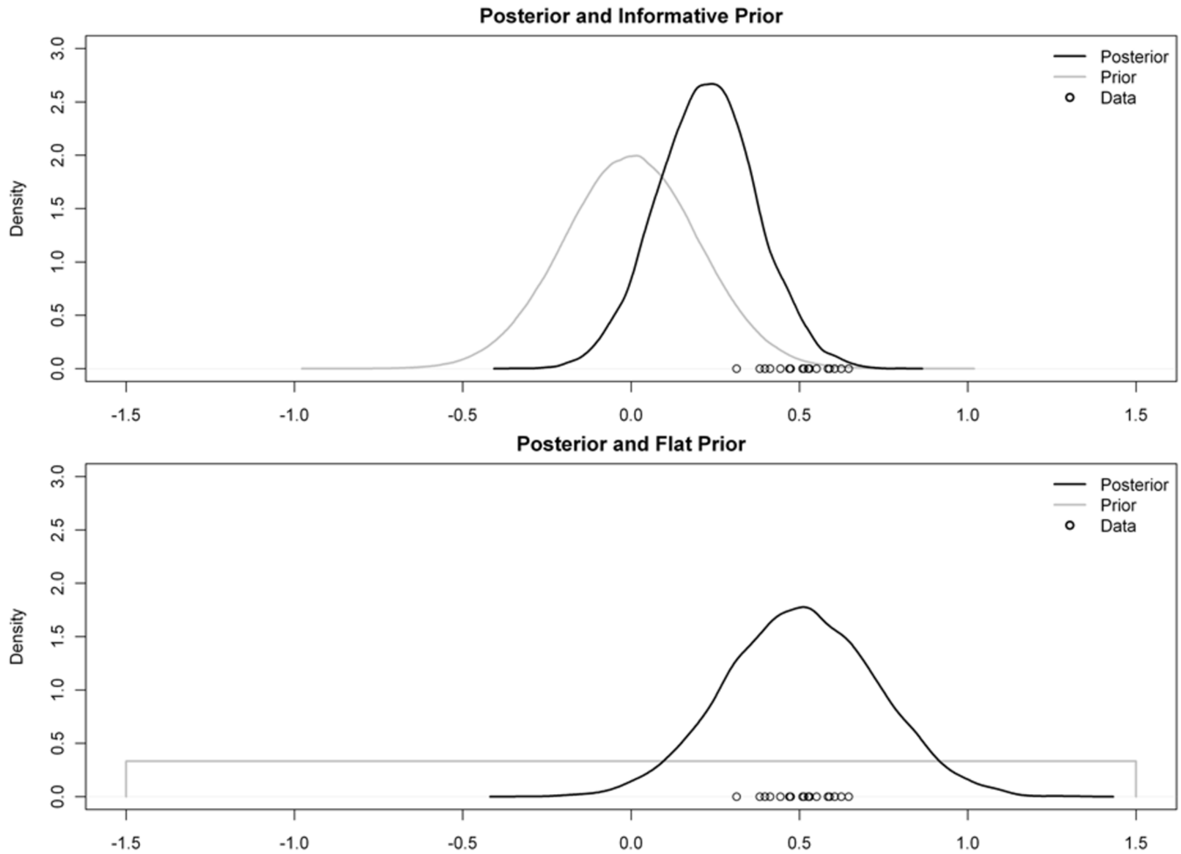
The prior distributions reflect the knowledge or belief of a researcher in different values of θ, without taking the data into account. The prior distributions can be either informative, or rather vague, as it is the case for 'flat' distributions (see the stylised example below). Formally, the Bayes theorem can be written using probability distributions, as:

$$p(\theta \mid x) = \frac{p(\theta) \cdot p(x \mid \theta)}{p(x)}.$$

Box two Bayesian inference in a nutshell

Here, $p(\theta \mid x)$ is the posterior distribution, $p(\theta)$ is the prior, and $p(x \mid \theta)$ denotes the likelihood of data.

Example: In order to illustrate the effect of alternative prior assumptions we simulated 20 observations from a Normal distribution with mean 0.5 and standard deviation of 0.1. We then re-estimated the mean using two alternative prior assumptions. In the first case, a normal distribution with mean 0 and standard deviation of 0.2 was used. The resulting posterior distribution is shown in the top panel of the graph below. In the second case, a Uniform distribution with a lower bound of -1.5 and an upper bound of 1.5 was assumed. The resulting posterior distribution is shown in the bottom panel of the plot. The two plots demonstrate how a posterior distribution can become narrower and alter in central tendency when an informative prior is included.



The combining of prior information and empirical data can be of great benefit when forecasting populations. For example, the future mean in a time series model based on past data may suggest an annual growth rate of 0.5 percent. This may be different from that expected by demographic experts (who may for example expect a future mean annual growth rate of zero). Hence, the inclusion of their opinions as an informative prior can help direct the parameter estimate of the mean level, on which model forecasts are based, away from an estimate that is based on the data and with an uninformative flat prior.

Recently, computation of a Bayesian model has become relatively easier as computational power has become more readily available, including introduction of the WinBUGS software. The latter has allowed users to estimate posterior distributions easily and quickly, without having to programme complex Markov chain Monte Carlo (MCMC) routines. For example, only a few lines of code are required to set up an AR model and state the prior distributions of the model parameters. Posterior distributions of parameters from a converged sample of an MCMC chain can be obtained speedily on a standard desktop computer. One such application of these advances has been in the estimation of parameters in SV models, which typically use Bayesian methods. See for example, Meyer and Yu, 2000[33]; Congdon, 2001[34]; Jacquier, 2003[35]. Bayesian estimation is typically preferred over classical parameter estimation because of the intractable form of the likelihood function[32].

Comparisons of forecasts

In this section results of forecasts from Bayesian time series models fitted to the historical data in Figure 1 are presented and compared with several official population projections. First, we compare our model averaged forecasts to the latest ONS scenario-based projections. We then revise our model averaged forecasts by adding expert opinion for a single parameter to understand better the effect of changing from a flat prior distribution to an informative one (see Box two). In the last section we compare our Bayesian time series forecasts on several shortened data series against past official projections, and to the actual observations.

Model averaged forecasts

For the Bayesian time series forecasts we consider 18 models for the differenced population growth rate, y_t, which are the same as those described in Abel *et al*.[11]. These consist of an independent normal (IN) model (with just a mean parameter and no autoregressive terms) and eight AR models (with non-zero means) that increase in order from AR(1) to AR(8). Nine more models with additional terms to control for stochastic volatility in y_t were also considered. This range of models was selected in order to represent all possible autoregressive processes that might adequately describe the differences in the overall growth rate series. As we had no previous knowledge about the nature of the parameters in each model, we assigned non-informative prior distributions. Also, note that the priors used for each model were the same.

The posterior distributions of the Bayesian time series forecasts can be summarised in a number of ways. In this section, we focus on summaries of the posterior distributions at two levels: the model probabilities and the joint predicted posterior distributions for future values of r_t and p_t. The posterior model probabilities of the 18 Bayesian time series forecasts fitted to several series of y_t with different end-points (see below under 'In sample forecasts') are provided in **Table 1**. The two last columns refer to two forecasts relying on the entire data set, that is, 1841-2007: one with 'vague' (flat) prior assumptions and the other with information assumed *a priori* (see the next section). In both cases the results indicate strong support, with a model probability of between 0.75 and 0.80, for the independent normal with stochastic volatility term (IN-SV). This model has only a single term for the mean level of change in the population growth rates (with no autoregressive terms) alongside parameters to control for the volatility shown in the data. The next most likely model is the AR(3)-SV model, followed by the AR(1)-SV. These models indicate that there is a small degree of support for models that include terms for autoregression at lags 3 or 1. The SV models with higher order AR terms, in addition to the models with constant variance terms, have very low model probabilities under 0.01.

Table 1 Posterior model probabilities for 18 models fitted for data series with different end points

Model	Posterior model probabilities: 1957	1967	1977	1987	1997	2007: Flat prior	2007: Informative prior
IN	0.00054	0	0	0	0	0	0
AR(1)	0.00021	0	0	0	0	0	0
AR(2)	0.00067	0	0	0	0	0	0
AR(3)	0.00147	0.00001	0	0	0	0	0
AR(4)	0.00051	0	0	0	0	0	0
AR(5)	0.00036	0	0	0	0	0	0
AR(6)	0.00006	0	0	0	0	0	0
AR(7)	0.00001	0	0	0	0	0	0
AR(8)	0	0	0	0	0	0	0
IN-SV	0.29967	0.49045	0.74872	0.79542	0.67155	0.79833	0.74962
AR(1)–SV	0.23621	0.18083	0.18004	0.11968	0.12038	0.07126	0.05314
AR(2)–SV	0.03650	0.03367	0.01731	0.01656	0.03123	0.01762	0.01431
AR(3)–SV	0.39972	0.27773	0.05032	0.06229	0.16113	0.10025	0.16832
AR(4)–SV	0.02240	0.01570	0.00321	0.00551	0.01436	0.01127	0.01332
AR(5)–SV	0.00152	0.00148	0.00038	0.00049	0.00123	0.00117	0.00118
AR(6)–SV	0.00014	0.00011	0.00003	0.00003	0.00011	0.00008	0.00011
AR(7)–SV	0.00001	0	0	0	0.00001	0.00001	0.00001
AR(8)–SV	0	0	0	0	0	0	0

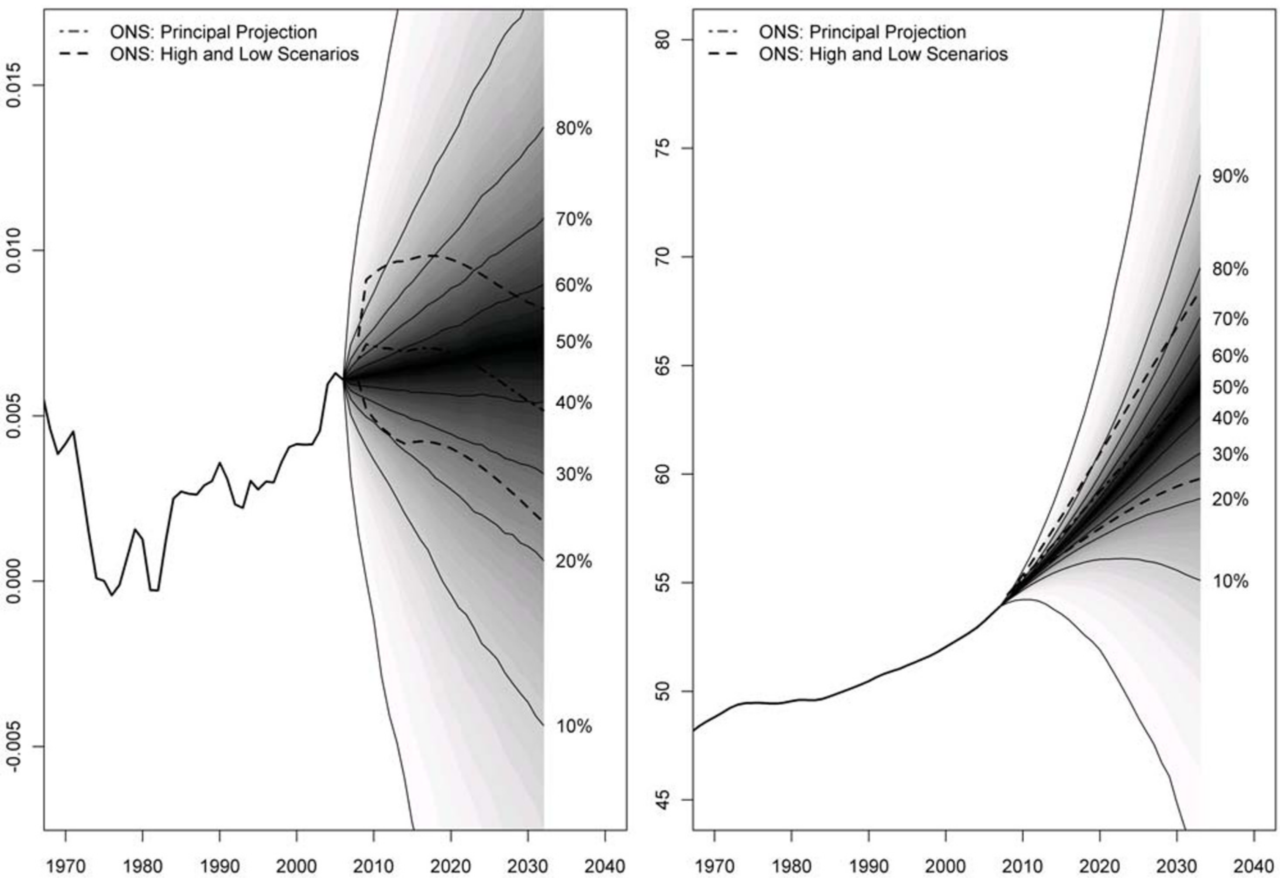
Given the posterior model probabilities from all 18 models, the joint predictive posterior distribution for future y_t up to 2032 was estimated. This provided a sample of 10,000 observations of future y_t values. These were then transformed to obtain the joint predicted distributions of future r_t and p_t using Equations (3) and (4), updated for each subsequent year.

The results are presented in the left and right panels of **Figure 2**, respectively. Each shade of the forecasted fan in these plots represents a single percentile of the estimated posterior density, where the darkest shades correspond to most central values and the lighter shades to the tails of the distribution. Contour lines are also plotted at each decile and at the 1st and 99th percentiles.

From these forecasts the median predictive population in 2033 was 64.0 million. Numerous measures of uncertainty are also available. For example, in 2033 the 20th percentile is 59.0 million persons and the 80th percentile is 69.4 million persons. In other words, our forecasts predict a 60 per cent probability that the 2033 population will fall between these two numbers.

Summaries of the predictive probability distributions can be compared with national projections. In the United Kingdom, ONS regularly prepares a set of projected total populations based on cohort component methodology under a range of deterministic scenarios. For this study we compare our results with approximations for three variants (principal projection, high and low population) published in the latest set of projections for England and Wales[36]. The principal variant relies on assumptions considered to reflect best the demographic patterns at the time they were adopted. The high (or low) population variant assumes a combination of high (or low) fertility, high (or low) life expectancy and high (or low) net migration, and is intended to provide users with a better sense of the plausible future extremes in population change. All three variants of population totals are displayed on the right hand panel in Figure 2. In the left panel, the derived values of r_t, calculated using Equation (1), and the future values of p_t are shown. The central, dot-dashed lines represent the principal projections, while the upper and lower dashed lines represent the high and low population variants, respectively.

Figure 2 **Joint predictive probability distribution of the model averaged growth rates (left) and resulting population forecast in millions (right)**



The panels in Figure 2 illustrate a number of differences between the ONS principal projection and that of our model averaged forecasts. First, the uncertainty in the ONS rate, represented by their high and low variants, is narrower than our model averaged forecasts at all points of time. In other words, the Bayesian forecasts include a wider range of uncertainty than those produced by ONS. Second, the uncertainty in the rate of population growth of the ONS projection does not increase substantially over time, unlike those derived using probabilistic methods. The reason for this is that ONS includes the effects of a rapidly ageing population (which affects all mortality variants) in their cohort component projections. Our models, based on overall population growth rates, are unable to account for these types of effects. Third, the ONS principal population projection in 2033 of 64.1 million is about the same as our model averaged median (64.0 million). Finally, the high and low variants in the projected population totals by ONS lie within the 77th and 22nd percentiles of the posterior predictive distribution of the 2033 population forecasts. In earlier forecast years the population totals from the high projection scenario are greater than our 80th percentile. On the other hand, the projected population totals from the lower variant never fall below our 20th percentile.

Sensitivity to alternative prior

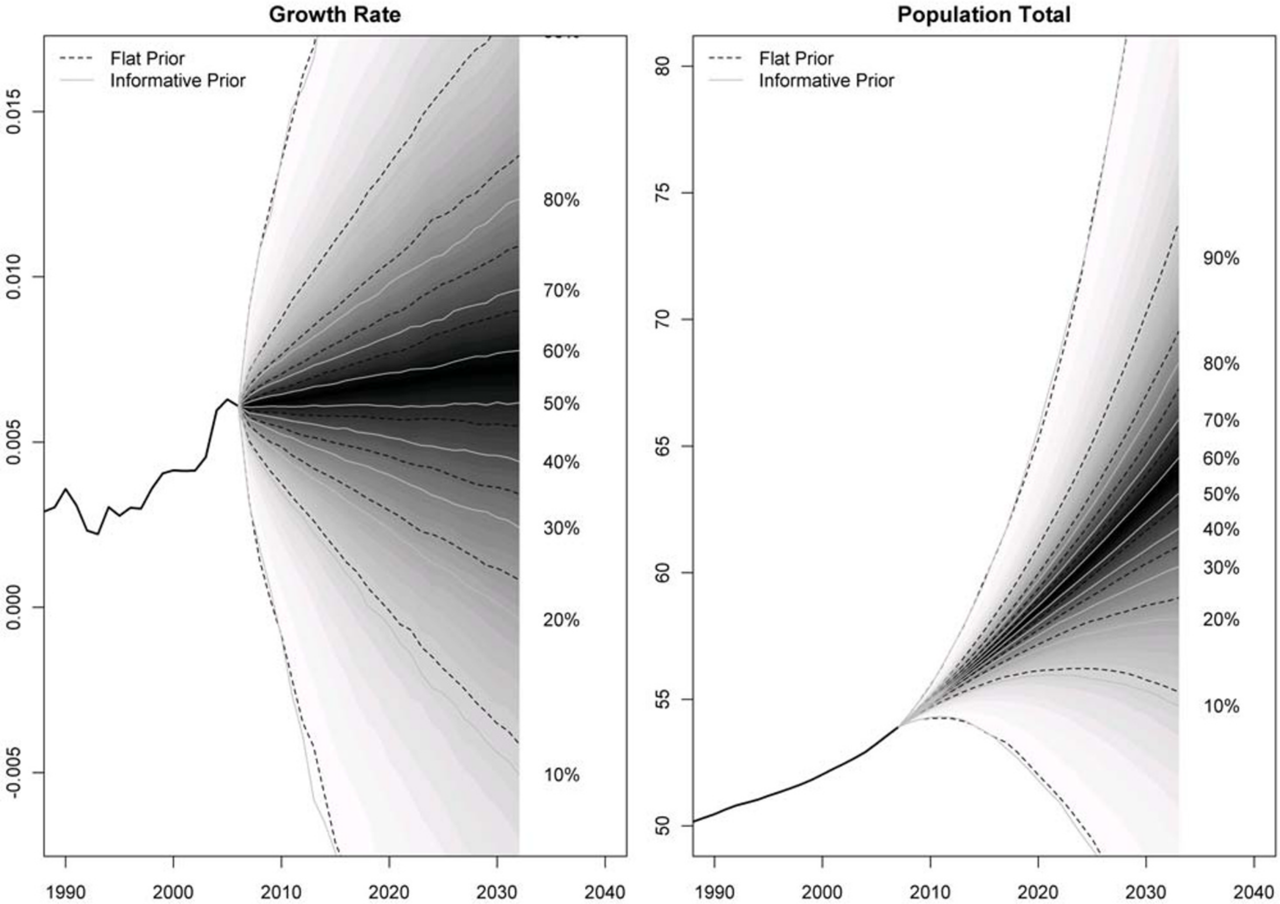
The forecasts presented in the previous section assumed flat prior distributions for all parameters. In this section we analyse the sensitivity of the posterior parameter estimates and model probabilities to the introduction of an informative prior. This is conducted by changing only the mean level of y_t (i.e., μ). This term also represents the annual mean level of increase (or decrease) in r_t and is present in all 18 models. In previous forecasts we assigned a non-informative prior distribution of $\mu \sim N(0, 100)$, where N denotes a Normal (Gaussian) distribution with mean 0 and variance 100.

To establish an informative prior, we used the ONS 2008-based principal, high and low projections to derive each variant's values for r_t and y_t from 2007 to 2032. The mean of the principal projection y_t (-0.000065) was used as the mean of new formative prior distribution. For the variance the means of the high (-0.000177) and low (-0.000035) variants of y_t were assumed to represent the 80th and 20th percentiles, respectively. These were chosen as previous authors, for example, Stoto (1983)[37] and Alho (1992)[38] have previously found high and low scenarios to represent roughly 66% confidence intervals. After a search among candidate distributions we found the $\mu \sim N$ (-0.000065, 0.0001) to approximately meet this criteria.

Given the informative prior, we calculated the corresponding posterior distributions for the parameter estimates and model probabilities. The AR models with informative priors exhibited mean values of μ similar to those with flat priors, albeit with reduced standard deviations. However, in the SV models, the mean values of μ became much closer to zero. The posterior model probabilities for two models with alternative prior assumptions for μ are shown in Table 1. Here we see that the posterior model probabilities of the informative prior remained fairly similar to the models with the flat prior assumptions.

To understand the effects of introducing informative priors on the future population growth rates and population totals, the predictive posterior distributions resulting from both model assumptions are plotted in **Figure 3**. As expected, the two plots on the right illustrate a reduced amount of uncertainty in comparison to the predictive posterior probability distributions obtained from the flat prior assumption (on the left side). For example, the 20th and 80th percentiles of p_{2033} were 58.9 million and 69.3 million, respectively, when the flat prior was used compared to 58.2 million and 68.3 million, respectively, when the informative prior was used. In addition, the median of the predictive posterior probability distribution reduced from 0.00733 for r_{2032} from the flat priors to 0.00619 from the informative priors. Consequently, the median of p_{2033} also falls from 64.0 million to 63.1 million.

Figure 3 Comparison of predictive posterior probability distributions of the population growth rates (left) and population (right) for flat and informative prior distributions



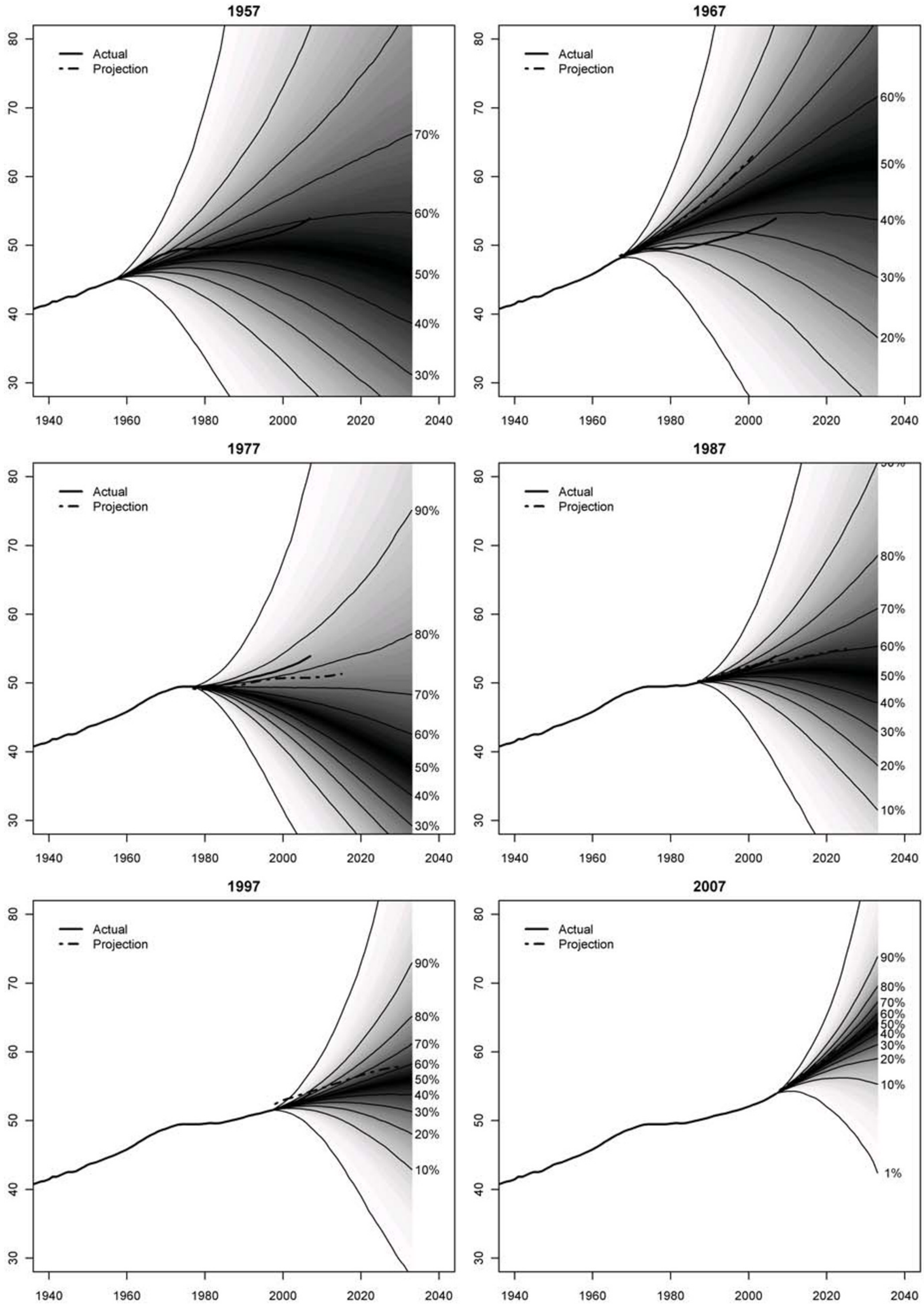
In-sample forecasts

To assess the performance of the Bayesian time series methods, in-sample forecasts (using flat priors) were conducted by using five shortened data sets with end points at 1957, 1967, 1977, 1987 and 1997 respectively. The results of these forecasts are compared against both past official population projections obtained from the Government Actuary Department (GAD) website [39] and with actual observations.

The posterior model probabilities from our forecasts with shortened series are presented in Table 1, alongside the model probabilities for those based on the full length data series ending in 2007. As expected, the forecasts based on longer time series have similar model probabilities as those based on the full data, with large support for the IN-SV model. In the shorter data series forecasts with end points in 1957 and 1967, more support is given to models that include autoregressive terms. This is most notable for the data with the end point in 1957, where the AR(3)-SV model has the highest model probability (0.39972).

The model averaged posterior predictive distributions of p_t for each of the shortened data series, along with the previously presented forecast from full data series (ending in 2007) are shown in **Figure 4**. For the shortest series, with last observation in 1957, the median of the population

Figure 4 **Sequence of joint predictive probability distribution of the population forecasts up to 2033, actual and GAD projections in the past six decades (in millions)**



forecast in the 2033 predictive distribution is 45.8 million. As we move sequentially through the results from data sets of increasing length, the median of the p_{2033} distribution increases to 61.9 million for the 1967 data set, falls to 37.6 million for the 1977 data set, and then increases to 51.0 million to 56.0 million to 64.0 million in the 1987, 1997 and 2007 data sets, respectively. An important factor explaining the large differences between the 1957, 1967 and 1977 forecasts are the ‘baby boom’ and ‘baby bust’ periods. In each of these forecasts the historical data did not provide any indications that the direction of population growth would change, thus leading to their relatively low or high projections. This is a weakness of time series forecasts, although it should be noted that experts at the time were also just as confused.

There are a number of noticeable conclusions that can be drawn when comparing the forecasted posterior distributions with the actual data and GAD projections, represented by the solid black line and dot-dashed line, respectively, in Figure 4. The median of the forecasted posterior distributions based on the 1957 data consistently underestimated the actual population (as it could not consider the future increases in fertility). This error was greatest during the early part of the forecast horizon where the actual population strays into the upper tails of our posterior distributions. However, this error improves, especially during the late 1980s when the population total moves towards the centre of our posterior distributions. In essence, the 1957 forecast averaged out the effects of the baby boom during the 1960s and the baby bust in the 1970s. In 2007 the England and Wales population was 53.9 million, which is within the 61st percentile of our p_{2033} posterior distribution. Not surprisingly, the GAD projection of 1957 suffers a very similar pattern of errors as our medians.

The median of our forecasts based on the 1967 data consistently overestimate the population, a consequence of making a forecast at a time of relatively large population growth. The error is greatest in the early part of the forecast where the growth rate of the actual population quickly decreases (see Figure 1). As with the 1957 based projection, the error improves in the latter part of the forecast horizon, with the 2007 observed population lying within the 39th percentile. The GAD projection, made in 1967, overestimates the actual population to a greater extent than our forecast, consistently following the 70th percentiles of our posterior distributions of p_t. This is presumably because they did not rely as much on historical patterns of population change as our forecasts do.

The forecasts based on the 1977 data, during the baby bust, suffer the largest errors of all the in-sample data sets. The actual population consistently remained in the upper tail, between the 80th and 90th percentiles of our posterior distributions, with the 2007 observed population lying within the 85th percentile. This large error is due to a combination of factors. First, the data series for r_t exhibits a turning point in the early 1980s, when the population began to increase once more. Second, unlike previous forecasts for shorter data series, large posterior model probabilities were estimated solely for the IN-SV model. As a result, there is a large reliance of the median forecasts on the μ parameter in this model. In addition, there is a lack of autoregressive parameters to temper the trend effect in the mean process, unlike the 1957 and 1967 based forecasts. The GAD projection in 1971 also underestimated the actual population, but with less error compared to the median of our posterior distributions.

Both the 1987 and 1997 based forecasts underestimate the actual populations, with the 2007 observed population lying within the 71st and 84th percentiles respectively. The reason for the difference is largely due to unanticipated increases in population caused by net immigration. The actual population closely follows the 70th percentile of our 1987 based forecast throughout, while the GAD projection follows our 60th percentile. The 1997 GAD projection appears to be affected by errors in the intercensal population estimates. As the forecast horizon increases, their projection becomes closer to our medians of the posterior distributions.

Conclusion

In this article we have presented a number of population forecasts for England and Wales. Utilising Bayesian methods, we have introduced uncertainty from multiple sources, including model choice and parameter estimation. We believe the resulting forecasts therefore provide a more realistic summary of future uncertainty in population forecasts in comparison with equivalent time series models fitted using classical methods.

Volatility in population growth rates was controlled for using stochastic volatility models, which tended to have the highest posterior model probabilities when fitted to historical data. The ability to control for volatility may be of importance when considered in the context of cohort component projection methods. These methods often require assumptions about future rates of population growth components. However, previous authors have noted that the success of these assumptions, when comparing their past projections with the actual population, may simply reflect the volatility or stability of the respective time series at the time the projections are made. See Shaw (2007)[40] and Keilman (2007)[41].

Bayesian methods allow the formal incorporation of expert judgement embodied in informative priors, and hence alter the forecasted population characteristics and their levels of uncertainty. The initial forecasts presented in this article were based on hardly informative flat priors and hence resulted in the large level of uncertainty in forecasted population size. This level of uncertainty was reduced through the inclusion of additional prior information. We derived our informative prior from future populations projected by ONS, which were based on expert opinion, on the future rates of the components of population change, and on cohort-component methodology. More informative priors based purely on expert opinions regarding the future of population growth rates could have been included. These could, for example, focus on the prior for the mean parameter, as well as the prior distributions for other parameters in the model, such as the degree of autocorrelation or preferences in models (for example, the inclusion of higher weights on SV models). Such prior information would result in further reductions in the estimated uncertainty due to added information in the parameter estimation and model choice procedures.

The simple time series models used to produce our population forecasts provided alternative estimates to those obtained using cohort component methods. When compared with past official population projections the medians from our simple models performed similarly well. In addition, we were able to provide multiple measures of uncertainty. Our models showed a similar degree of susceptibility to turning points, especially when low posterior probabilities were estimated for models with autoregressive terms. This feature might be tempered through the inclusion of expert opinion. For example, we might provide higher prior model weights to those that include autoregressive terms in comparison to the independent normal models.

We focused only on modelling the change in the population growth rate. This has a number of restrictions when interpreting results. For example, we are unable to provide future forecasts for the components of population change or disaggregate future population by age and sex groups.

We hope to explore these areas further in the future using Bayesian methods motivated by the augments provided throughout this article. In addition, further disaggregation of the population growth rate into components is likely to provide more accurate forecasts and further improvements in the estimated levels of uncertainty.

In conclusion, we believe the future of producing population estimates will require more emphasis on specifying uncertainty so that more informed decisions can be made by population planners and policy makers. The use of time series modelling methods allows a large library of statistical and econometric techniques to be applied to meet these demands. The use of the Bayesian approach in fitting these models also allows for further extensions over classical estimation methods, leading to more realistic forecasts and associated uncertainty measures.

Acknowledgements

This research is funded by ESRC Grant number RES-625-28-0001. The Centre for Population Change is a joint initiative between the University of Southampton and a consortium of Scottish Universities in partnership with the Office for National Statistics (ONS) and the General Register Office for Scotland (GROS). The findings, interpretations, and conclusions expressed in this article are entirely those of the author(s) and should not be attributed in any manner to ONS or GROS. The authors would like to thank Peter W.F. Smith, Jonathan J. Forster, Arkadiusz Wiśniowski, and the two anonymous reviewers for their helpful comments and suggestions.

References

1 Ahlburg, DA and Land, KC, eds. (1992) ‘Population forecasting: Guest Editors’ introduction’. *International Journal of Forecasting* 8:289-299.

2 Lee, RD and Tuljapurkar, S (1994) ‘Stochastic population projections for the United States: Beyond high, medium and low’. *Journal of the American Statistical Association* 89 (428) 1175-1189.

3 Lutz, W, ed. (1996) *The future population of the world: What can we assume today?* London: Earthscan.

4 Bongaarts, J and Bulatao, RA, eds. (2000) *Beyond six billion: Forecasting the world’s population*. Washington DC: National Academy Press.

5 Wilson, T and Rees, P (2005) ‘Recent developments in population projection methodology: A review’. *Population, Space and Place* 11:337-360 (see page 342).

6 Lutz, W and Goldstein, JR (2004) ‘Introduction: How to deal with uncertainty in population forecasting?’ *International Statistical Review* 72(1) 1-4.

7 Rowan, S and Wright, E (2010) ‘Developing stochastic population forecasts for the United Kingdom: Progress report and plans for future work’. Paper presented at the Joint Eurostat/UNECE Work Session on Demographic Projections, Lisbon. An earlier version of this paper is also available at: www.ons.gov.uk/about-statistics/methodology-and-quality/imps/updates-reports/progress-report-on-developing-stochastic-population-forecasts-for-the-uk.pdf

8 Alho, JM and Spencer, BD (2005) *Statistical demography and forecasting*. New York: Springer. (See pages 238-240).

9 Keilman, N (2001) 'Data quality and accuracy of United Nations population projections, 1950–95'. *Population Studies* 55:149-164.

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

11 Abel, GJ, Bijak, J, Forster, JJ, Raymer, J and Smith, PWF (2010) 'What do Bayesian methods offer population forecasters?' Working Paper 6/2010, ESRC Research Centre for Population Change, University of Southampton.

12 Office for National Statistics (2009) 2008-based national population projections. Available at: www.statistics.gov.uk/StatBase/Product.asp?vlnk=8519

13 Willekens, F (1990) *Demographic forecasting; State-of-the-art and research needs*. In Hazeu, CA and Frinking, GAB, eds. 'Emerging issues in demographic research', pp 9-66. Amsterdam: Elsevier.

14 De Beer, J (2000) *Dealing with uncertainty in population forecasting*. Department of Population Statistics, Central Bureau of Statistics (CBS), Voorburg.

15 O'Neill, BC, Balk, D, Brickman, M and Ezra, M (2001) 'A guide to global population projections'. *Demographic Research* 4 (8) 203–288.

16 Booth, H (2006). 'Demographic forecasting: 1980 to 2005 in review'. *International Journal of Forecasting*, 22 (3) 547–581.

17 Bijak, J (forthcoming) *Forecasting international migration in Europe: A Bayesian view*. Dordrecht: Springer.

18 Leslie, PH (1945) 'On the use of matrices in certain population mathematics'. *Biometrika* 33 (3) 183–212.

19 Rogers, A (1975) *Introduction to multiregional mathematical demography*. New York: Wiley.

20 Rees, P (2008) *What happens when international migrants settle? Projections of ethnic groups in United Kingdom regions*. In Raymer, J and Willekens, F 'International migration in Europe: Data, models and estimates', eds., pp 329-358. Chichester: Wiley.

21 Land, KC and Rogers, A, eds. (1982) *Multidimensional mathematical demography*. New York: Academic Press.

22 Schoen, R (1988) *Modeling multigroup populations*. New York: Plenum Press.

23 Schoen, R (2006) *Dynamic population models*. Dordrecht: Springer.

24 Rogers, A (1995) *Multiregional demography: Principles, methods and extensions*. Chichester: Wiley.

25 Daponte, BO, Kadane, JB and Wolfson, LJ (1997) ‘Bayesian demography: Projecting the Iraqi Kurdish population, 1977–1990’. *Journal of the American Statistical Association* 92 (440) 1256–1267.

26 Heilig, G, Buettner, T, Li, N, Gerland, P, Alkema, L, Chunn, J and Raftery, AE (2010) ‘A stochastic version of the United Nations World Population Prospects: Methodological improvements by using Bayesian fertility and mortality projections’. Paper presented at the Joint Eurostat/ UNECE Work Session on Demographic Projections, Lisbon.

27 Human Mortality Database. Available at: www.mortality.org

28 Chatfield, C (2004) *The analysis of time series: An introduction*. Boca Raton: Chapman & Hall/ CRC. Page 26.

29 Saboia, JLM (1974) ‘Modeling and Forecasting Populations by Time Series: The Swedish Case’. *Demography* 11 (3) 483-492.

30 Ahlburg, DA (1987) ‘Population Forecasts for South Pacific Nations using Autoregressive Models 1985–2000’. *Journal of Population Research* 4 (2) 157-167.

31 Pflaumer, P (1992) ‘Forecasting US population totals with the Box-Jenkins approach’. *International Journal of Forecasting* 8 (3) 329-338.

32 Bayes, T (1763) ‘n Essay towards solving a Problem in the Doctrine of Chances’ *Philosophical Transactions of the Royal Society* 53: 370–418.

33 Meyer, R and Yu, J (2000) ‘BUGS for a Bayesian analysis of stochastic volatility models’. *Econometrics Journal* 3 (2) 198-215.

34 Congdon, P (2001) *Bayesian Statistical Modelling*, Chichester: Wiley.

35 Jacquier, E, Polson, NG, Rossi, PE (2003) ‘Bayesian analysis of stochastic volatility models with fat-tails and correlated errors’. *Journal of Econometrics* 122(1) 185-212.

36 Wright, E (2010) ‘2008-based national population projections for the United Kingdom and constituent countries’. *Population Trends* 139: 91-114. ONS does not prepare forecasts for England and Wales jointly, and thus our aggregation of the two countries does not take into account internal migration between them or its uncertainty. Our aggregates for England and Wales have to be read with this in mind and may differ from the official published figures.

37 Stoto, MA (1983) ‘The accuracy of population projections’. *Journal of the American Statistical Association* 78:13-20.

38 Alho, JM (1992) ‘The magnitude of error due to different vital processes in population forecasts’. *International Journal of Forecasting* 8: 301-314.

39 See: www.gad.gov.uk

40 Shaw, C (2007) ‘Fifty years of United Kingdom national population projections: how accurate have they been?’ *Population Trends* 128: 8-23.

41 Keilman, N (2007) ‘UK national population projections in perspective: How successful compared to those in other European countries?’ *Population Trends* 129: 20-30.