ESRC Centre for Population Change Working Paper
Number 6

What Do Bayesian Methods Offer Population Forecasters?

Guy J. Abel
Jakub Bijak
Jonathan J. Forster
James Raymer
Peter W.F. Smith

June 2010

ISSN2042-4116

ABSTRACT

The Bayesian approach has a number of attractive properties for probabilistic forecasting. In this paper, we apply Bayesian time series models to obtain future population estimates with uncertainty for England and Wales. To account for heterogeneity found in the historical data, we add parameters to represent the stochastic volatility in the error terms. Uncertainty in model choice is incorporated through Bayesian model averaging techniques. The resulting predictive distributions from Bayesian forecasting models have two main advantages over those obtained using traditional stochastic models. Firstly, data and uncertainties in the parameters and model choice are explicitly included using probability distributions. As a result, more realistic probabilistic population forecasts can be obtained. Second, Bayesian models formally allow the incorporation of expert opinion, including uncertainty, into the forecast. Our results are discussed in relation to classical time series methods and existing cohort component projections. This paper demonstrates the flexibility of the Bayesian approach to simple population forecasting and provides insights into further developments of more complicated population models that include, for example, components of demographic change.

KEYWORDS

Population forecasting; Bayesian forecasting; Bayesian modelling; time series models; stochastic volatility; model averaging; England and Wales.

EDITORIAL NOTE

Guy Abel is Postdoctoral Researcher in the ESRC Centre for Population Change, University of Southampton. Jakub Bijak is a Lecturer in the Division of Social Statistics and a member of the modelling research team in the ESRC Centre for Population Change. Jonathan J. Forster is a Professor in the School of Mathematics and a member of the modelling research team in the ESRC Centre for Population Change. James Raymer is a Reader in the Division of Social Statistics and a senior member of the ESRC Centre for Population Change research team. Peter W.F. Smith is the Director of Southampton Statistical Sciences Research Institute, a Professor in the Division of Social Statistics and a member of the modelling research team in the ESRC Centre for Population Change.

Correspondence to: g.j.abel@soton.ac.uk.

ACKNOWLEDGEMENTS

The authors would like to thank Andrei Rogers for his advice concerning the direction of this paper and Arkadiusz Wisniowski for his suggestions concerning the modelling.

The ESRC Centre for Population Change Working Paper Series is edited by Teresa McGowan

ESRC Centre for Population Change

The ESRC Centre for Population Change (CPC) is a joint initiative between the Universities of Southampton, St Andrews, Edinburgh, Stirling, Strathclyde, in partnership with the Office for National Statistics (ONS) and the General Register Office Scotland (GROS). The Centre is funded by the Economic and Social Research Council (ESRC) grant number RES-625-28-0001.

Website | Email | Twitter | Facebook | Mendeley

WHAT DO BAYESIAN METHODS OFFER POPULATION FORECASTERS?

TABLE OF CONTENTS

1. INTRODUCTION **1**

2. DATA **2**

3. MODELS **4**

3.1 AUTOREGRESSION MODEL 4

3.2 STOCHASTIC VOLATILITY 5

3.3 BAYESIAN INFERENCE 5

3.4 MODEL UNCERTAINTY 6

4. FORECASTS **7**

4.1 INDIVIDUAL AR MODELS 7

4.2 INDIVIDUAL SV MODELS 10

4.3 MODEL AVERAGED FORECASTS 13

5. COMPARISONS OF RESULTS **14**

5.1 COMPARISON WITH ALTERNATIVE METHODS 14

5.2 IN SAMPLE FORECASTS 16

5.3 CONSTANT FUTURE VOLATILITY 18

6. CONCLUSION **19**

REFERENCES **21**

1 Introduction

This paper explores the use of Bayesian methods for population forecasting. The main rationale is the need for incorporating uncertainty in population forecasts, advocated by many authors since the 1980s (Alho & Spencer, 1985; Keyfitz, 1991; Lee, 1998). Statistical agencies typically provide "high" and "low" variants to communicate uncertainty around their principal population projections. Such variants have a number of drawbacks with the most prominent being a lack of specificity regarding the probability range of the high, low or even principal variants, e.g., Keilman *et al.* (2002) or Lutz & Goldstein (2004). In response, demographers and statisticians have developed methods to calculate probabilistic forecasts that describe the uncertainly of future populations by relying on time series models, expert judgements or extrapolation of past forecast errors (Keilman, 2001; Keilman *et al.*, 2002). Methods have also been developed to combine elements of each of these approaches, for example, the parameters from time series models have been constrained according to expert opinions (Lee & Tuljapurkar, 1994) or to target levels and age distributions of fertility and mortality (Lutz *et al.*, 2001).

We believe Bayesian methods offers a more natural framework than traditional frequentist methods to forecast future population with uncertainty. First, the Bayesian approach offers an explicit, coherent and transparent mechanism to include uncertainty in the data, parameters of the model and the model itself, by using probability distributions. Second, it allows the inclusion of expert judgements, including uncertainty, into the model framework. Third, the predictive distributions follow directly from the probabilistic model applied. As a result, probabilistic population forecasts, with more reliable and coherent estimates of predictive distributions, can be obtained. Together, these have the potential to improve the measurement of uncertainty in forecasts, and thus improve our potential for planning and understanding population change.

There have been several recent papers on the Bayesian estimation of demographic components for countries with inadequate data: Alkema *et al.* (2008) for fertility, and Brierley *et al.* (2008) and Raymer *et al.* (2010) for migration. For forecasting demographic components, existing examples include Tuljapurkar (1999) and Alkema *et al.* (2010) for fertility, Pedroza (2006), Girosi & King (2008) and Chunn *et al.* (2010) for mortality, and Gorbey *et al.* (1999) and Bijak (2010) for migration. A comprehensive example of using Bayesian methods for forecasting is the study of the Iraqi Kurdish population prepared by Daponte *et al.* (1997). In this paper, we focus on a developed-country situation with a relatively good availability of demographic data. We also consider a wider class of time series models and apply a Bayesian modelling averaging techniques.

To present the case for a Bayesian framework for population forecasting, we focus on a single time series of population change in England and Wales, which is described in the next section. In Section 3, we introduce the notation and describe the models used in this study. These include autoregression models for time series and stochastic volatility models to account for possible heterogeneity in historical data. We also present the Bayesian inference used for parameter calculation and model averaging, which is used to incorporate model uncertainty, and a more robust set of estimates. In Section 4, we present our fully probabilistic population forecasts from 2008 to 2033. Our results and their levels of uncertainty are discussed in relation to classical time series methods and existing cohort component estimates in Section 5. Also, comparisons are made with forecasts based on shortened data series and under the assumption that the future volatility of the population growth rate will remain at 2006 levels. Finally, we end the paper with a summary and some suggestions for extending the proposed approach.

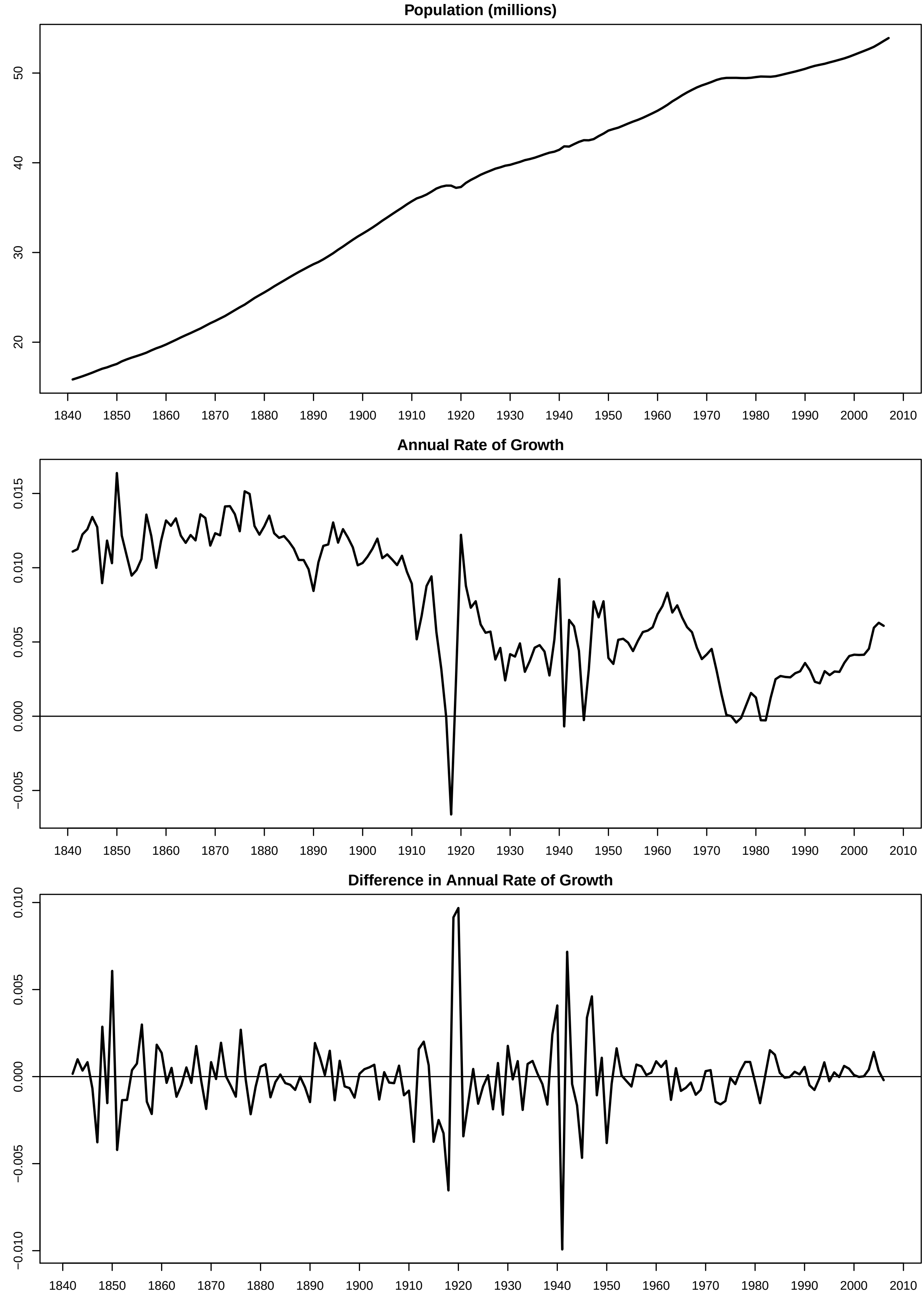
2 Data

A historical series of the England and Wales population totals are used to introduce the Bayesian approach to time series forecasting, obtained from The Human Mortality Database (`www.mortality.org`). The mid-year population totals from 1841 to 2007, including military personnel, are presented in the top panel of Figure 1. Here, we see that the population totals in England and Wales exhibited a steady increase over time, rising from 15.8 million in 1841 to 53.9 million in 2007. Brief periods of slight population decline are visible during the First World War and the 1918 influenza pandemic. Also noticeable is a period of leveling off in the population totals during the 1970s and 1980s, a result of net emigration and a slow rate of natural increase.

The features of population change are more evident when the annual rates of growth, plotted in the second panel of Figure 1, are considered. Detailed explanations for these patterns can be found in various books on British population history (Wrigley & Schofield, 1989; Coleman & Salt, 1992; Anderson, 1996; Hinde, 2003). The following provides a very brief account. The population growth rates were highest during the first third of the series. This was predominantly due to the declining mortality occurring before the decline in fertility, which remained at pre-industrial levels for much of this period. Between the two wars the rate of growth remained low in comparison with the later half of the 19th century and early 20th century. This was driven by the effects of low fertility from economic depression and a change in sociological factors. After the Second World War, population growth rates increased initially, through a short-lived fertility rise associated with demobilization, followed by a more substantial increase (baby boom) in fertility during the 1950s and early 1960s. In the late 1970s and early 1980s, the levels of population growth slowed down (as mentioned above) before rising in more recent decades though net immigration and increased fertility levels.

In this paper, we use this simple time series of population data to forecast the future population totals up to 2033. We are primarily interested in identifying the models that best fit these data in order to specify realistic probabilistic intervals in forecasted populations. As we can see from the observed data, the annual rates of growth have varied considerably over time. The models, described in the next section, take these variations into account in specifying the uncertainty, which is useful for understanding future variations in population change. As

Figure 1: England and Wales Population Data, 1841-2007.



a result, we do not have the concern of choosing a strategic starting point for the data series that might best reflect future variation. A long time series also allows us to test our models by performing in-sample forecasts (see Section 5.2).

3 Models

In this section, we specify the models and notation used to forecast future annual growth rates in England and Wales. The subsections focus on autoregression models, stochastic volatility, Bayesian inference and model uncertainty. To start, let p_t be the population size at time t for an uninterrupted series of observed time points. In population forecasting, we are interested in obtaining estimates of p_t for one or more values of $t > T$, where T is the last observed time point, and their associated measures of uncertainty.

In order to model p_t, we first derive the time series of population growth rates r_t, where

$$p_{t+1} = (1 + r_t)p_t. \tag{1}$$

However, experience suggests that if we are to use models which assume stationarity, it is more appropriate to model changes in r_t (Chatfield, 2003), denoted by y_t:

$$y_t = r_t - r_{t-1}. \tag{2}$$

For our data, time series plots of r_t and y_t are presented in the second and third panels, respectively, of Figure 1. In the next two sections, we introduce autoregression (AR) models and stochastic volatility (SV) models for y_t.

3.1 Autoregression Model

AR models have been used in the demographic context to forecast populations, see for example, Saboia (1974), Ahlburg (1987), Pflaumer (1992), Alho & Spencer (2005) and Statistics Netherlands (2005). An AR model of order p, denoted AR(p), is defined as

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + z_t, \tag{3}$$

where ϕ_j are the autoregressive coefficients representing the correlations between observations y_t and y_{t-j}, whilst j represents the time lag, and z_t are assumed to be independent observations from a probability distribution with zero mean and constant variance, σ^2. A slightly more flexible model, which we apply, also allows for a non-zero mean, μ, for y_t:

$$y_t = \mu + \sum_{j=1}^{p} \phi_j (y_{t-j} - \mu) + z_t. \tag{4}$$

This model implies a mean increase in r_t of μ each year. For a fully-specified probability model, we need to assume a distribution for z_t. Typically, a normal distribution is assumed.

3.2 Stochastic Volatility

SV models were popularised for modelling financial data where the assumption of constant variance for z_t is usually untenable. Models that account for non-constant variance have been sparsely used in the demographic context (Keilman & Pham, 2004; Bijak, 2010). Historical time series of demographic data, however, often exhibits some form of volatility due to events such as epidemics, wars or baby booms. This is certainly true for the data set out in Figure 1. SV models are time series models, similar to the AR models defined in (4), but where the variance of z_t is allowed to be time-dependent. This is achieved by replacing σ^2 with σ_t^2, and specifying a time series model for σ_t^2. In this paper, we assume an AR(1) model for $-\log \sigma_t^2$, i.e.,

$$\sigma_t^2 = e^{-h_t} \tag{5}$$

and

$$h_t \sim N(\alpha + \psi h_{t-1}, \tau^2), \tag{6}$$

where h_t represents the volatility at time t, α denotes the mean level of h_t over the entire time period, and τ is the standard deviation of h_t. Finally, ψ is the autoregressive coefficient representing the correlations between h_t and h_{t-1}.

3.3 Bayesian Inference

In Bayesian inference, uncertainty about the (multivariate) parameter θ of a statistical model is described by its posterior probability distribution given observed data $y_{\{T\}} = \{y_1, \ldots, y_T\}$. The probability density function of y_t is obtained by using Bayes Theorem:

$$f(\theta|y_{\{T\}}) = \frac{f(y_{\{T\}}|\theta)f(\theta)}{f(y_{\{T\}})}, \tag{7}$$

where $f(y_{\{T\}}|\theta)$ is the likelihood function and is defined by the model, $f(\theta)$ is the prior distribution for θ and $f(y_{\{T\}})$ is a normalising constant. The prior distribution $f(\theta)$ specifies the uncertainty about θ prior to observing any data.

Forecasting or prediction is particularly natural in a Bayesian framework. Uncertainty about the next K future values of y_t (for $t = T+1, \ldots, T+K$) is described by the joint predictive probability distribution

$$f(y_{T+1}, \ldots, y_{T+K}|y_{\{T\}}) = \int f(\theta|y_{\{T\}}) \prod_{k=1}^{K} f(y_{T+k}|y_{\{T+k-1\}}, \theta)\mathrm{d}\theta. \tag{8}$$

Note that the product term represents the joint predictive distribution in the case that parameter θ is known. The Bayesian predictive distribution simply averages (integrates) this with respect to the posterior probability distribution for θ. Hence, uncertainty about θ in light of the observed data is fully integrated.

In a Bayesian analysis we obtain forecasts and associated measures of uncertainty by calculating marginal probability distributions for quantities of interest by integrating the posterior distribution in (7) or the predictive distribution in (8). Performing these integrations analytically is typically not possible for realistically complex models such as those described above. Historically, this has prevented demographers and others from taking advantages of Bayesian methods for statistical inference. Recent developments in Bayesian computation have focussed on Markov chain Monte Carlo (MCMC) generation of samples from distributions such as (7) or (8); see Gelman *et al.* (2003) for details. Once a sample has been obtained from a joint distribution, then a sample from a distribution of any component or function of components is readily available. To generate samples from the posterior and predictive distribution in this paper, we used an MCMC sampling approach implemented using the WinBUGS software (Lunn *et al.*, 2000).

3.4 Model Uncertainty

In practical population forecasting, it is unrealistic for the analyst to be sure that any particular statistical model is the right one upon which to base their forecasts. Hence, the statistical methodology adapted should be one which allows for model uncertainty. Furthermore, we consider it essential that the measures of uncertainty associated with any forecast should incorporate both the uncertainty concerning the model and the uncertainty concerning the parameters of each model. In this paper, model uncertainty is directly integrated with parameter uncertainty into a single predictive probability distribution. An comprehensive review of Bayesian model averaging can be found in Hoeting *et al.* (1999).

Formally, let $m = 1, \ldots, M$ index the models under consideration and let θ_m represent the parameter associated with model m. Note that different models may have parameters of different dimensionality. For example, the AR(1) model with drift has a three-dimensional parameter (μ, ϕ_1, σ^2). The likelihood function for model m is $f(y_{\{T\}}|\theta_m, m)$, the prior distribution for m, is $f(\theta_m|m)$ and the posterior distribution is

$$f(\theta_m|y_{\{T\}}, m) = \frac{f(y_{\{T\}}|\theta_m, m)f(\theta_m, m)}{f(y_{\{T\}}|m)}, \tag{9}$$

where $f(y_{\{T\}}|m)$ is a normalising constant, known as the marginal likelihood for model m, and is given by

$$f(y_{\{T\}}|m) = \int f(\theta_m|m)f(y_T|\theta_m, m)\mathrm{d}\theta_m. \tag{10}$$

Prior uncertainty about models is encapsulated by a discrete probability distribution, $f(m)$, $m = 1, \ldots, M$. The prior model probabilities can be assigned the same values, $\frac{1}{M}$.

The posterior probability distribution for m given observed data $y_{\{T\}}$ is obtained by using Bayes Theorem as follows:

$$f(m|y_{\{T\}}) = \frac{f(y_{\{T\}}|m)f(m)}{f(y_{\{T\}})}. \tag{11}$$

Hence, the posterior model probability for any model m is proportional to the product of the prior model probability and the marginal likelihood. Therefore, efficient methods for computation of marginal likelihoods is essential for Bayesian inference under model uncertainty. See, for example, those described in O'Hagan & Forster (2004). In our implementation, we found that the bridge sampler (Meng & Wong, 1996) was effective for this computation.

Finally, to obtain a predictive distribution for population forecasts in the presence of model uncertainty, (8) is extended to

$$\begin{aligned} f(y_{T+1}, \ldots, y_{T+K}|y_{\{T\}}) &= \sum_{m=1}^{M} f(m|y_{\{T\}}) f(y_{T+1}, \ldots, y_{T+K}|y_{\{T\}}, m) \qquad (12) \\ &= \sum_{m=1}^{M} f(m|y_{\{T\}}) \int f(\theta_m|y_{\{T\}}, m) \prod_{k=1}^{K} f(y_{T+k}|y_{\{T+k-1\}}, \theta_m, m) \mathrm{d}\theta_m, \end{aligned}$$

which is the average of predictive distributions for individual models weighted by their posterior probabilities, $f(m|y_{\{T\}})$.

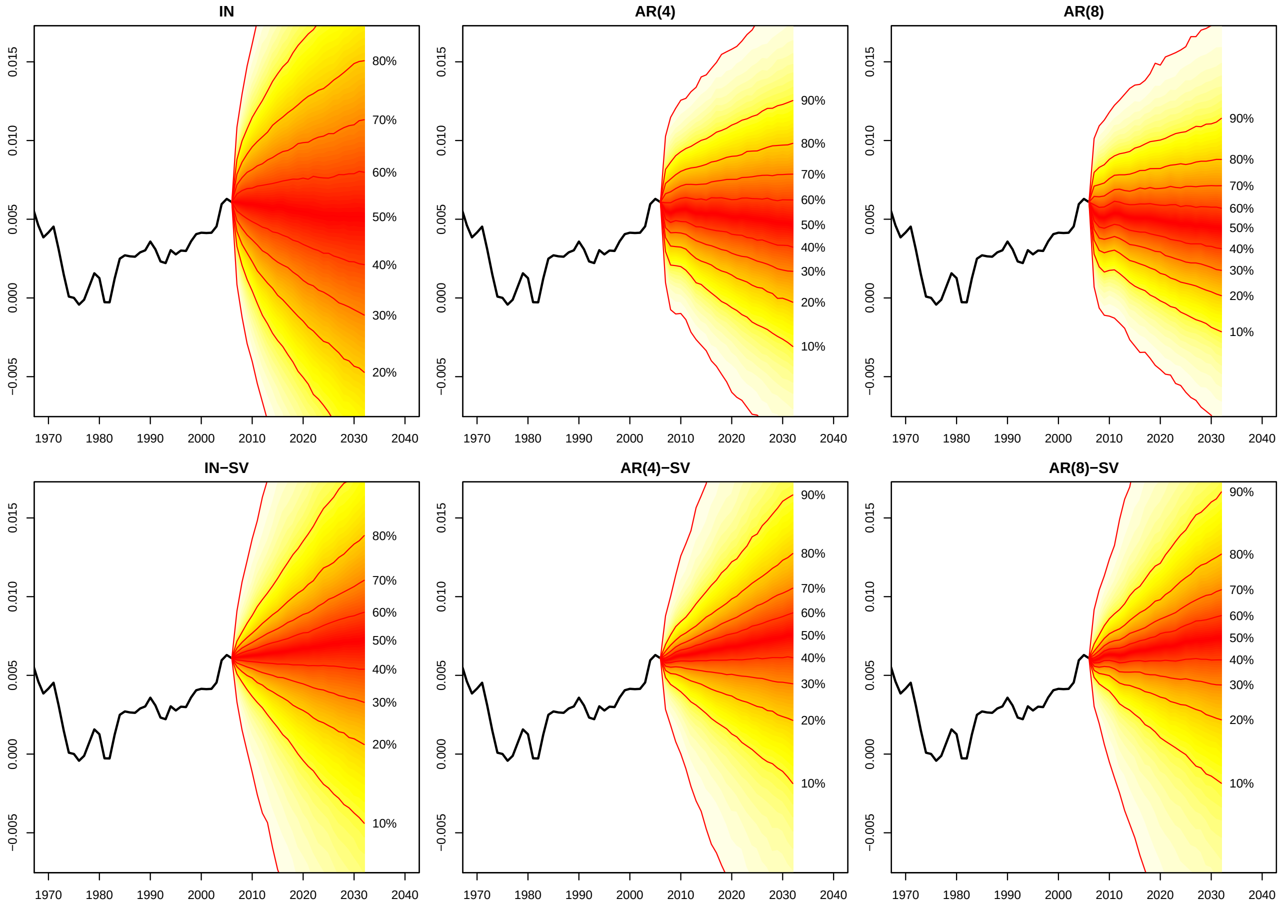
4 Forecasts

In this section, we present parameter estimates from a range of individual AR and SV models. In addition, the predictive probability distributions from a selection of these models are provided in order to gain a better understanding of the effect of expanding the dimensionality of θ_m on future population growth rates. These individual forecasts are compared in the final subsection with a single forecast that accounts for our uncertainty in model choice.

4.1 Individual AR Models

An initial set of nine models was considered for the differenced population growth rate, y_t, introduced in (1) and presented in the bottom panel of Figure 1. These consist of an independent normal (IN) model and eight autoregression models (with non-zero means), increasing in order from AR(1) to AR(8). 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 have no previous knowledge about the nature of the parameters in each model we assigned non-informative prior distributions: $\mu \sim N(0, 100^2)$, $\phi_j \sim N(0, 1), j = 1, \ldots, p$ and $\sigma \sim U(0, 100)$, where $N(\mu, \sigma^2)$ denotes a Normal (Gaussian) distribution with mean μ and variance σ^2, whereas $U(a, b)$ denotes a Uniform distribution over the interval (a, b). An MCMC sample of 10,000 observations was obtained from the posterior distribution for each model.

Figure 2: Selected Posterior Predictive Plots of Population Growth Rates from Individual Models



In a Bayesian analysis, marginal posterior distributions completely describe the uncertainty about each model parameter given the observed data. These are typically summarised using posterior means (as parameter estimates) and posterior standard deviations (as measures of uncertainty). The posterior means and standard deviations for the parameters of each of the nine models are presented in Table 1. Posterior estimates of μ tend to be centred on zero with much lower standard deviations than their prior distributions. This feature was also true for the estimates of ϕ_j. In all models, the posterior means of ϕ_j at lower values of j were below zero, indicating negative autocorrelation for their respective lags. Estimates of ϕ_j, for $j > 5$, tend to be close to zero, signifying that the association between y_t and y_{t-j} becomes weak at larger values of j.

Posterior predictive plots of the forecasted r_t from the IN, AR(4) and AR(8) models are illustrated in the top row of Figure 2. These are obtained from the forecast of y_t by rearranging (1) for each set of iterates and assuming the starting point of $r_{2006} = 0.00609$ as in the data. Each shade of the forecasted fan represents a single percentile of the estimated posterior density, where 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 the 1st and 99th percentile. Forecasts from the simple independent normal model provide a greater level of uncertainty of

Table 1: Posterior Means (Standard Deviations) of AR Model Parameters from MCMC Simulations and Model Comparisons Statistics.

Model	μ	σ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	AIC [a]	BIC[b]	$f(m\|y_{\{T\}})$[c]
IN	−0.00003 (0.00017)	0.00215 (0.00012)									−1558.75	−1556.64	0.04639
AR(1)	−0.00003 (0.00017)	0.00213 (0.00012)	−0.15360 (0.07858)								−1561.05	−1554.82	0.01879
AR(2)	−0.00004 (0.00016)	0.00208 (0.00011)	−0.19130 (0.07848)	−0.23275 (0.07776)							−1568.75	−1558.41	0.13700
AR(3)	−0.00006 (0.00016)	0.00203 (0.00011)	−0.24421 (0.07692)	−0.27889 (0.07668)	−0.22221 (0.07801)						−1575.29	−1560.84	0.55434
AR(4)	−0.00007 (0.00016)	0.00202 (0.00011)	−0.27600 (0.08036)	−0.31548 (0.07917)	−0.25796 (0.08042)	−0.13230 (0.07963)					−1576.61	−1558.05	0.15719
AR(5)	−0.00008 (0.00016)	0.00200 (0.00011)	−0.29726 (0.08085)	−0.35671 (0.08206)	−0.30694 (0.08257)	−0.17762 (0.08225)	−0.15658 (0.08042)				−1578.81	−1556.14	0.07690
AR(6)	−0.00009 (0.00016)	0.00200 (0.00011)	−0.30782 (0.08120)	−0.36911 (0.08393)	−0.32842 (0.08695)	−0.20312 (0.08828)	−0.17720 (0.08274)	−0.06892 (0.08115)			−1577.63	−1550.84	0.00818
AR(7)	−0.00009 (0.00016)	0.00200 (0.00011)	−0.30333 (0.08168)	−0.35741 (0.08478)	−0.31624 (0.08925)	−0.18207 (0.09171)	−0.15403 (0.09028)	−0.04924 (0.08455)	0.06089 (0.08002)		−1576.19	−1545.30	0.00095
AR(8)	−0.00009 (0.00016)	0.00200 (0.00011)	−0.29960 (0.08046)	−0.36279 (0.08516)	−0.32525 (0.08934)	−0.19299 (0.09333)	−0.17334 (0.09382)	−0.07220 (0.09117)	0.04309 (0.08579)	−0.06296 (0.08195)	−1574.87	−1539.86	0.00025

[a]Akaike Information Criterion.
[b]Bayesian Information Criterion.
[c]Posterior model probabilities for models with constant variance.

future values. As autoregressive parameters are added to the independent normal model, the posterior predictive distribution become comparatively tighter, illustrated by the comparisons along the top row of Figure 2. As noted previously, ϕ_j for $j > 5$ are close to zero in the higher order AR models. This results in similar posterior predictive distribution for higher order models, where the increase in the number of lagged terms no longer substantially reduces the width of the predictive distribution.

4.2 Individual SV Models

A further set of nine models were considered for the differenced population growth rate. These extend the nine models of the previous subsection with additional parameters to allow the the variance of y_t be time-dependent, as introduced in (6). Again we assigned non-informative prior distributions to the new parameters: $\alpha \sim N(0, 10^2)$, $\psi \sim U(-0.999, 0.999)$ and $\tau \sim N(0, 100^2)I(\tau > 0)$, where I indicates a truncation to the distribution. The posterior means (and standard deviations) of the parameters in the nine models are presented in Table 2 from MCMC samples of 10,000 observations for each model. As in Table 1, estimates of μ tend to be centered on zero with much lower standard deviations than their prior distributions. Estimates of autoregressive parameters tend to be close to zero for most ϕ_j with the exception of $j = 2, 3$.

The SV extension replaces the σ^2 term in the AR models with time dependent variances σ_t^2. As specified in (6) this results in three new parameters α, ψ and τ (also shown in Table 2), as well as 165 new latent parameters of h_t. Posterior means of α, the average volatility level, are similar across all models. The corresponding values of σ are slightly lower than σ in Table 1 after applying the transformation (5). Posterior means for ψ, representing the autocorrelation between a current level of volatility and that of a previous year, are above zero in all models. Values are close to 0.9 indicating a strong positive autocorrelation in the volatility levels of r_t. Estimates of τ, measuring the standard deviation of volatility, are similar across all models. The posterior distributions for the 165 latent h_t parameters are plotted (for illustrative purposes) for the SV-IN model in Figure 3. In addition, the predictive distributions of the future volatility from this model are provided. Inspection of this plot reveals a number of features. First, the estimated volatility levels decreases throughout most of the observed period. Volatility is at its lowest level in 2001, prior to a increase in subsequent years leading up to the last observation, marked by the vertical line. Second, the estimated volatility levels are highest during the 1918 influenza pandemic and war periods. During these years, the 1st quantile of the predictive distribution are higher than the 99th quantile of the final estimated volatility. Finally, the median of the predictive distributions of the future volatility gradually increases towards the value of α provided in Table 2. The width of the predictive distributions also gradually increases over time.

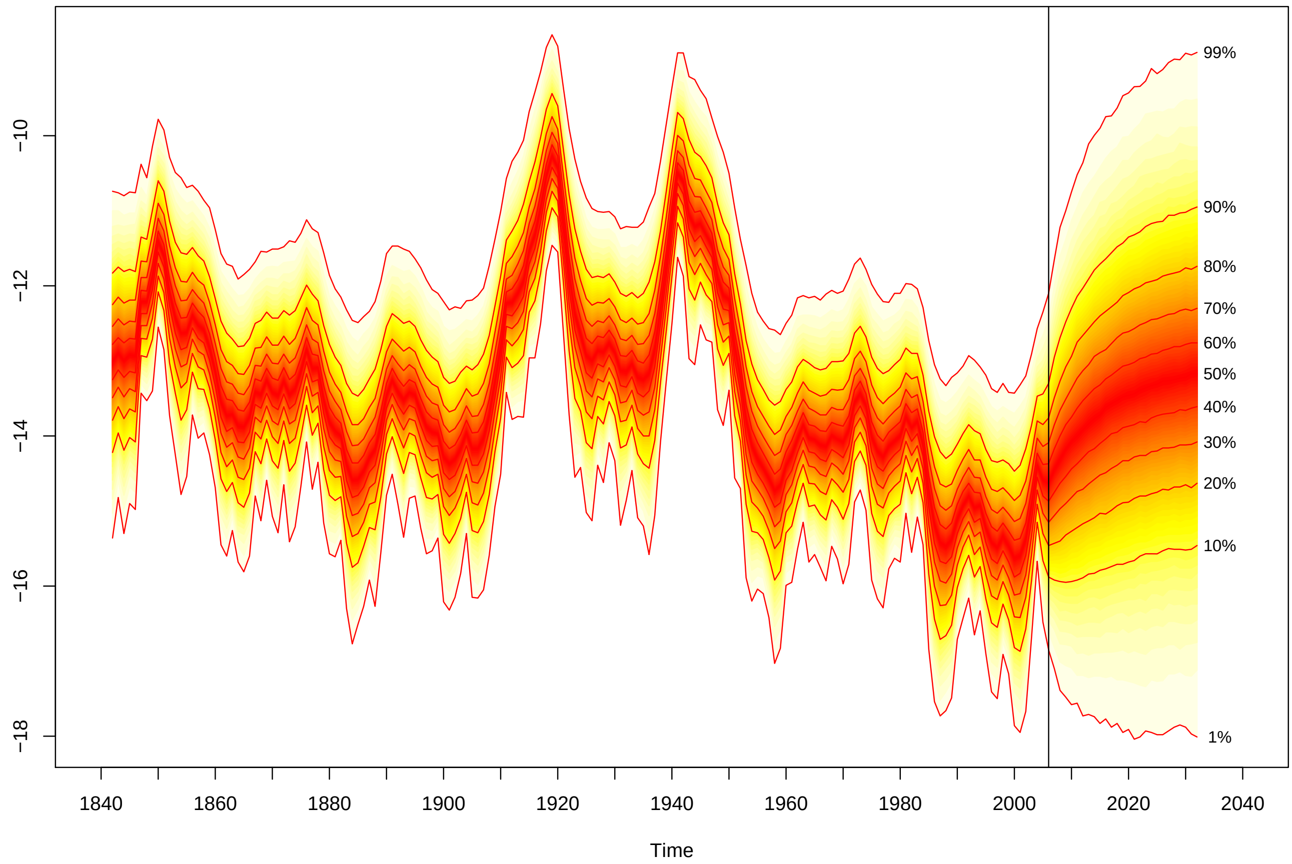
Posterior predictive plots of the forecasted r_t from the IN-SV, AR(4)-SV and AR(8)-SV models are illustrated in the bottom row of Figure 2. Comparisons between models with

Table 2: Posterior Means (Standard Deviations) of SV Model Parameters from MCMC Simulations and Posterior Model Probability.

Model	μ	α	ψ	τ	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	ϕ_7	ϕ_8	$f(m\|y_{\{T\}})^a$
IN-SV	0.00004	−12.76827	0.89797	0.69656									0.79347
	(0.00008)	(0.59693)	(0.06475)	(0.15313)									
AR(1)-SV	0.00004	−12.77901	0.89793	0.69186	−0.01478								0.08120
	(0.00008)	(0.59031)	(0.06374)	(0.15019)	(0.08703)								
AR(2)-SV	0.00005	−12.77835	0.89562	0.69926	−0.00800	−0.13723							0.01692
	(0.00008)	(0.61418)	(0.06794)	(0.16233)	(0.08802)	(0.07723)							
AR(3)-SV	0.00006	−12.75337	0.90374	0.66188	−0.06125	−0.14660	−0.13875						0.09785
	(0.00007)	(0.60338)	(0.06383)	(0.15168)	(0.09216)	(0.07777)	(0.07463)						
AR(4)-SV	0.00006	−12.73525	0.90706	0.65758	−0.05734	−0.14450	−0.14079	0.01122					0.00939
	(0.00007)	(0.61279)	(0.06168)	(0.14808)	(0.09695)	(0.08140)	(0.07591)	(0.07298)					
AR(5)-SV	0.00006	−12.73477	0.90976	0.65988	−0.05371	−0.12194	−0.12052	0.02790	0.06931				0.00108
	(0.00008)	(0.61062)	(0.06002)	(0.14437)	(0.09578)	(0.08345)	(0.07846)	(0.07513)	(0.06749)				
AR(6)-SV	0.00006	−12.72169	0.91049	0.65709	−0.05347	−0.12076	−0.11716	0.03063	0.07041	0.01294			0.00008
	(0.00008)	(0.60846)	(0.06000)	(0.14932)	(0.09583)	(0.08407)	(0.08204)	(0.07679)	(0.06808)	(0.06505)			
AR(7)-SV	0.00006	−12.70404	0.91280	0.64891	−0.05430	−0.11898	−0.11954	0.03137	0.06922	0.01473	0.00032		0.00001
	(0.00008)	(0.61567)	(0.05902)	(0.14830)	(0.09710)	(0.08519)	(0.08276)	(0.08165)	(0.07441)	(0.06661)	(0.06235)		
AR(8)-SV	0.00005	−12.74792	0.90440	0.66837	−0.06778	−0.13009	−0.12276	0.00203	0.04088	−0.01088	−0.02167	−0.05633	0.00000
	(0.00007)	(0.61084)	(0.06430)	(0.15499)	(0.09781)	(0.08559)	(0.08165)	(0.08812)	(0.07937)	(0.07112)	(0.06470)	(0.06013)	

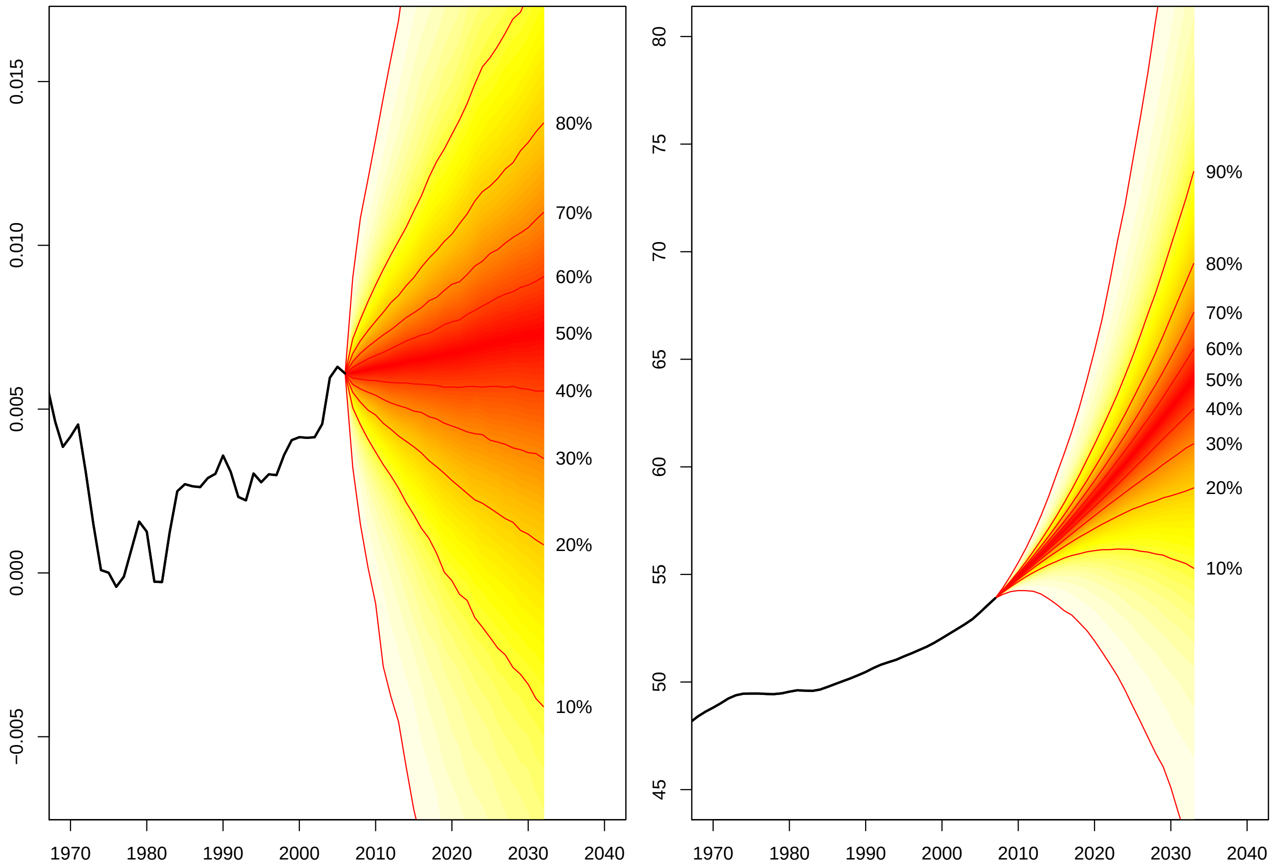
[a]Posterior model probabilities for models with stochastic variance. Note, when AR models with constant variance are included, these probabilities remain unchanged.

Figure 3: Posterior and Predictive Distributions of Volatility (h_t) from the IN-SV Model.



SV terms reveal that uncertainty in forecasted r_t is slightly reduced through the addition of autoregressive parameters, as was the case with the AR models with a constant variance parameter. However, this reduction in uncertainly is small due to the values of ϕ_j, for $j > 3$, in the SV models being close to zero. Comparison of the forecasted population growth rates between the selected individual models with constant variance and the SV models (between the top and bottom row in Figure 2) demonstrates a different shape in the forecast fans, caused by a combination of lower ϕ values and additional terms for a non-constant variance in the SV models. The width of the decile contour lines of the predictive distributions in the SV models increase at a steady rate. The equivialnt contor lines in the constant variance model tend to spread quickly (depending on the order of the AR model) and then continue to widen at a steady, but slower, rate. Consequently, for the simplest models (IN and IN-SV) the inclusion of the SV terms reduces the width of the predictive distribution, as illustrated in 2033 where the difference between the 80th and 20th percentile is 0.01449 compared to 0.01942 for the IN-SV model. As the model order for the mean process increases, this relationship is reversed. For example the difference between the 80th and 20th percentile for the AR(4)-SV model is 0.01010 compared to 0.01120 for the AR(4) model.

Figure 4: Joint Predictive Probability Distribution of the Model Averaged Growth Rates (left) and Resulting Population Forecast in Millions (right).

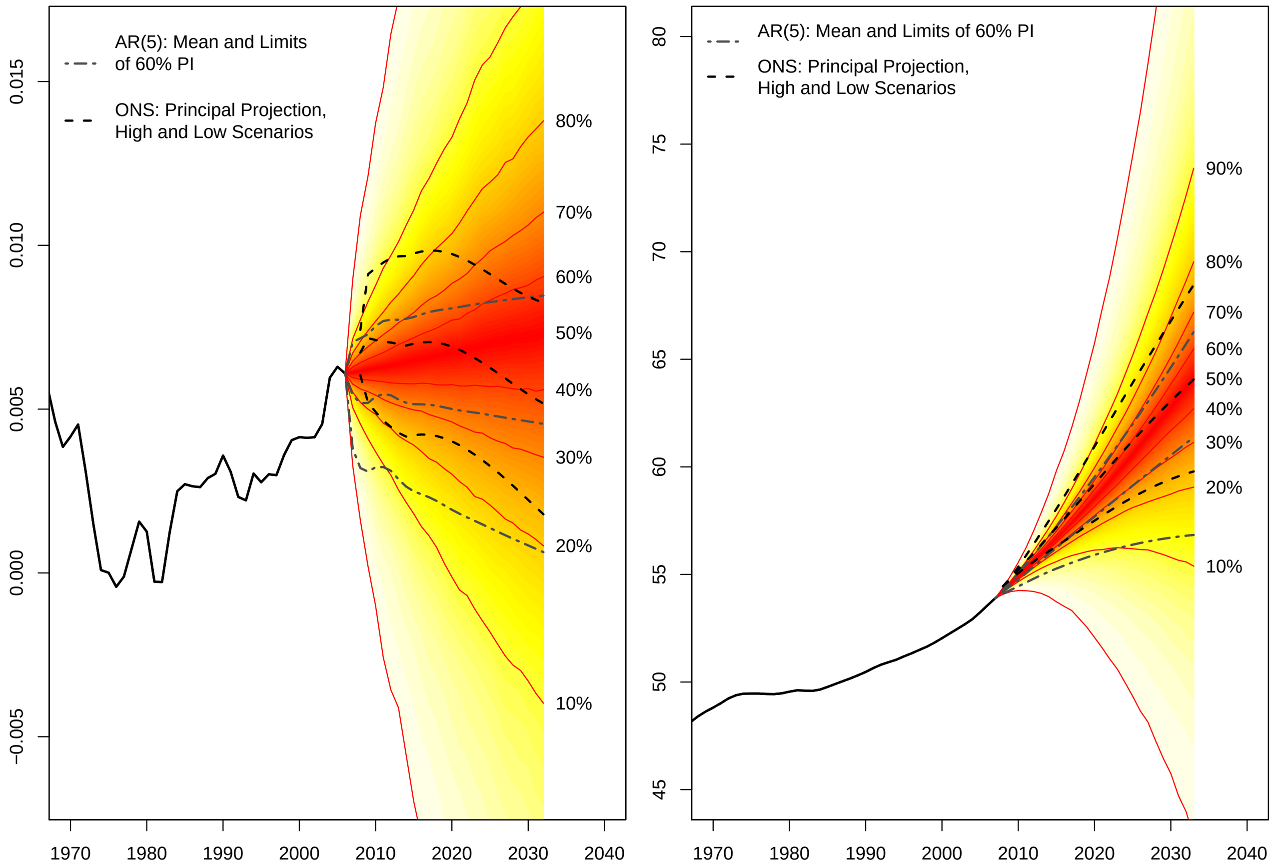


4.3 Model Averaged Forecasts

We calculated posterior probabilities, $f(m|y_{\{T\}})$ for all models (with and without SV terms) as described in Section 3.4. These are only displayed in the last column of Table 2. Note, all posterior probabilities for the models with out SV term (in Table 1) were very close to zero. This result give strong support for the SV-IN model (posterior probability of 0.79347). The next most likely model is the SV-AR(3) model, followed by the SV-AR(1). All SV models with higher order AR terms, in addition to the models with constant variance terms, appear very unlikely with model probabilities below 0.01.

The predictive probability distribution of r_t averaged over all models, given the model probabilities, are presented in the left hand panel of Figure 4. Because a sample from the posterior of probability distribution of each individual model is generated in the analysis, calculation of the averaged predictive probability distribution is straightforward. Unsurprisingly, this plot bares a large resemblance to the SV-IN model forecasts in Figure 2, from which a large posterior model probability was estimated. On the right hand panel of Figure 4 we also present the resulting population forecast from the predictive probability distribution of r_t. Our results provide a median predictive population of 64.0 million in 2033. Numerous measures of uncertainty are

Figure 5: Comparison of Alternative Future Growth Rates (left) and Population in Millions (right).



also available, for example in 2033 the 20th percentile is 59.0 million and the 80th percentile is 69.4 million.

5 Comparisons of Results

In this section the presented results and their levels of uncertainty are discussed in relation to alternative methods and model specifications in three stages. First, the results in Figure 4 are compared with forecasts from classical time series methods and existing cohort component estimates. Second, accuracy is assessed by re-estimating parameters and forecasts using shortened data series and comparing these against past observed data. Third, comparisons of our results are made with forecasts conditional on the volatility fixed at the last observed level, under the scenario that past observed variation in historical population growth rates (such as during the influenza or war periods) will not occur again.

5.1 Comparison with Alternative Methods

For comparative purposes, traditional frequentist time series models corresponding to the nine AR models (with the constant variance assumption) were estimated using the `arima` function in R 2.10.1 (R Development Core Team, n.d.). As noted at the beginning of Section 3.1 such models have previously been used to forecast future populations, as opposed to SV models. Estimates of ϕ_j in all models were within 0.1 of the mean values presented in Table 1. Estimates from the `arima` function of μ and σ were also very similar (to a higher degree of accuracy) to those estimated using the Bayesian methodology. The close correspondences between parameter estimates are due to the reliance on data, rather than the (uninformative) priors, in the calculation of posterior distributions.

Model summary statistics from the models fitted in R are also provided in Table 1. For comparison with a Bayesian model averaging approach, the last column of Table 1 shows the re-estimated posterior model probabilities, when only the models with constant variance are considered. The Akaike Information Criterion (AIC) of Akaike (1973) is commonly used for model selection for time series methods (Chatfield, 2003, p256). This criterion favoured the AR(5) model, as opposed to the model probabilities which provided this model with a probability of 0.07690 (and zero when SV models are considered). Hence, if we were to use the AIC as an alternative method for model selection in a frequentist setting, only a single model, which we estimated to have a low probability, would be selected. The Bayesian Information Criterion (BIC) of Schwarz (1978), which penalises the inclusion of extra parameters more severely, is also presented in Table 1. The BIC closely resembles the posterior model probabilities (as might be expected given the established link (Raftery, 1995)) and suggests AR(3) as a suitable model. Note, neither the AIC or BIC can be easily obtained for equivalent SV models as they require a direct calculation of the likelihood.

In Figure 5 we compare the results of the choice of a single model, based on the AIC, against our model averaged forecast over all 18 models. In left hand panel the mean forecast of r_t from the AR(5) model is displayed using the dot-dashed line. This was calculated using the `predict.arima` function in R. In addition, the 60% prediction intervals from the AR(5) forecast are plotted. Comparisons of the two measures of central tendency provide differing estimates of the future growth rate of population. The forecasted mean growth rate from the AR(5) model is 0.00455 in 2032. In contrast, the median of the predictive probability for r_{2032} was 0.00733. This difference is also reflected in the forecasts of the total population plotted on the right hand panel of Figure 5. The mean forecasted population for the AR(5) model is 61.4 million in 2033, 3.6 million lower than the median of the posterior predictive probability for p_{2033}. This discrepancy is predominantly caused by allowing the model selection process to consider models that account for stochastic variance terms. As noted above, when only AR models are considered the model selection using the AIC still favours a model with a low estimated model probability. However, the median of the predictive probability for r_t in 2032 averaged over only the AR model was 0.00493, much closer to that of the AR(5) model.

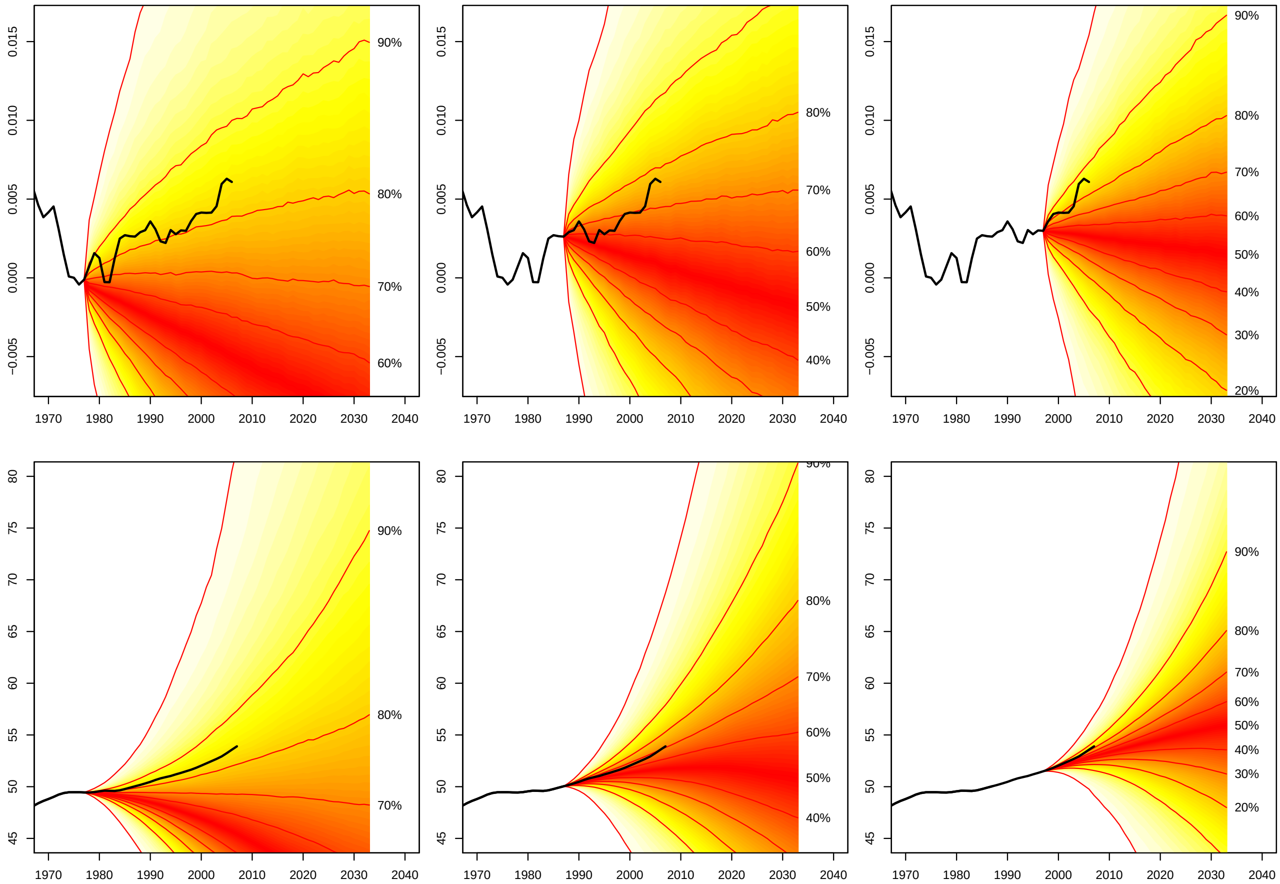
Uncertainty from the selection of a single model can be compared to that of the model averaged uncertainty. We use the 60% prediction intervals, obtained from the fit in R and plotted in Figure 5, and compare these to the equivalent 20th and 80th percentiles of the posterior predicative distributions. The model selection method provides a smaller amount of uncertainty for future forecasts than the model averaged forecast. The 60% prediction interval for r_{2032} from the AR(5) model is (0.00063, 0.00846), smaller in range to that of the 20th and 80th percentiles (0.00083, 0.01380) in the same year. The same story is reflected in the range of forecasted population, where the model selection method provides a prediction interval of (56.8 million, 66.3 million) compared to (58.9 million, 69.3 million) from the model averaged method. This discrepancy is derived predominantly from the averaging over models in the Bayesian framework. It is interesting to note that when only AR models are averaged, the range between the 20th and 80th percentiles (not plotted) of r_{2032} (-0.00055, 0.01057) remains large in comparison with the AR(5) model.

In the United Kingdom, the Office for National Statistics (ONS) prepare a set of projected total populations estimated using a cohort component methodology under a range of deterministic scenarios. We focus on comparing our results with the three variants (principal, high and low) published in the latest set of projections for England and Wales, see Wright (2010). All three variants are based on sets of demographic trend-based assumptions for future fertility, mortality and net migration. The principal variant relies on assumptions considered to best reflect demographic patterns at the time they were adopted. The high (or low) population variant assumes a combination of high (or low) fertility, life expectancy and net migration. They are intended to provide users with a better understanding of future uncertainty in population change. All three variants of population totals are displayed on the right hand panel in Figure 5. On the left hand panel are the derived values of r_t. The central, dashed lines represents the principal projections, whilst the upper and lower dashed line represent the high and low population variants respectively. The panels in Figure 5 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 far smaller than that of our model averaged forecasts at all points of time. 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. Third, the ONS principal population projection in 2033 of 63.7 million is slightly lower than our model averaged median (64.0 million), despite a reduction in the rate away from the median of the model averaged forecast towards the end of the horizon. Finally, the high and low variants in the projected population totals by the ONS lie within the 77th and 24th percentiles of the posterior predictive distribution of the 2033 population forecasts.

5.2 In Sample Forecasts

To asses the accuracy of the Bayesian time series methods in-sample forecasts are conducted using three shortened data sets with end points at 1977, 1987 and 1997 respectively. The posterior model probabilities of these shortened series were very similar those presented in Table

Figure 6: Growth Rates (top) and Population (bottom) Forecasts Based on Data Shortened at 1977 (Left), 1987 (Middle) and 1997 (Right).

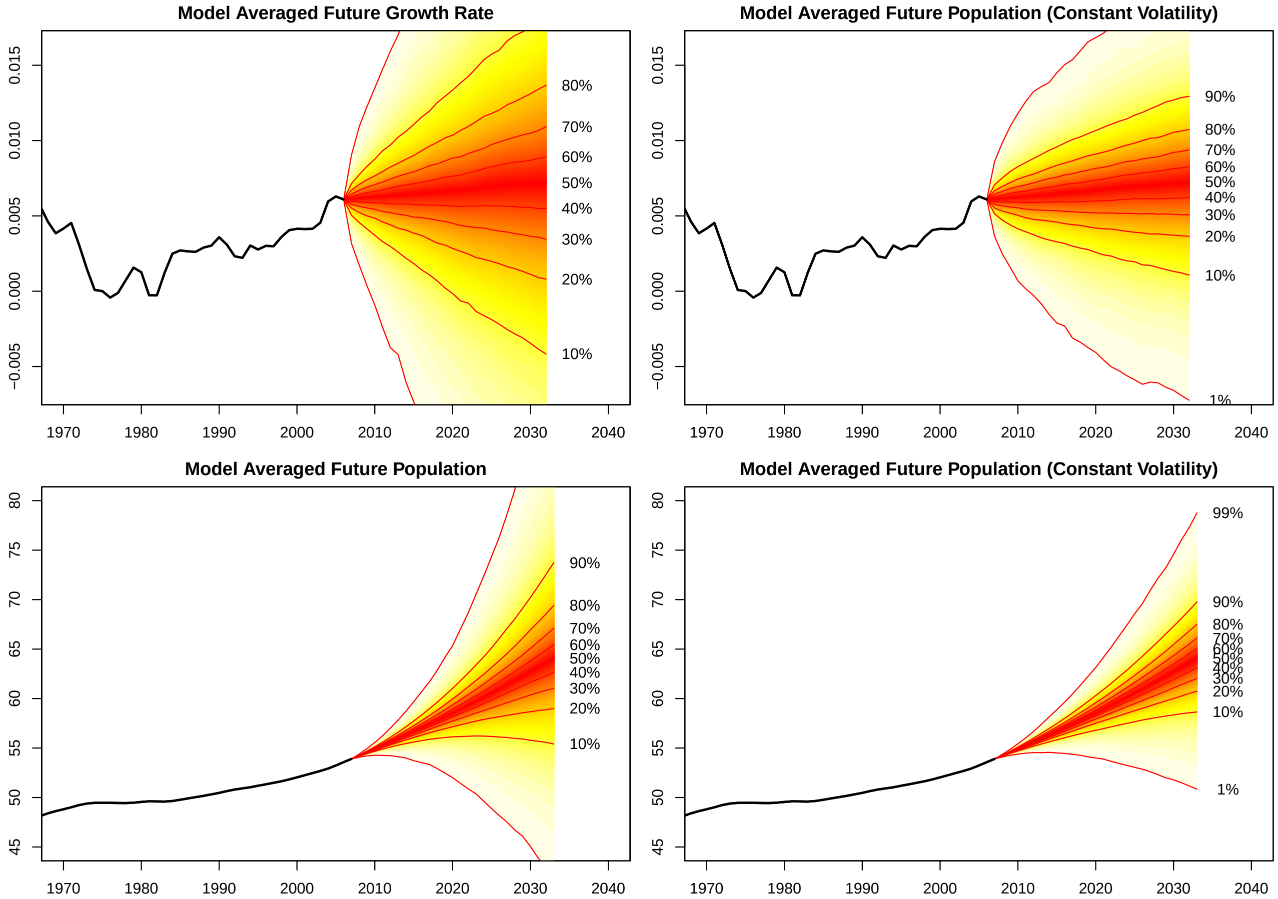


2, with the IN-SV model having a model probability of 0.74872, 0.79542 and 0.67155 for data ending in 1977, 1987 and 1997 respectively. Model averaged posterior predictive distributions of r_t (top) and p_t (bottom) are shown in Figure 6.

The plots show altering forecasted predictive distributions for each series. For the shortest series, with last observation in 1977, the median of the population forecast in the 2033 predictive distribution is 37.6 million. For the data set with the last observation in 1987 the median population in 2033 is 51.0 million and for the longest data set with the last observation in 1997 the median population in 56.0 million. There are a number of noticeable conclusions that can be drawn from Figure 6.

Inspection of the median from the in-sample forecasts illustrate the susceptibility of time series models to turning points in the data series, as occurred during the mid-1970's and early 1980's. This weakness is accentuated by the results of the model averaging process each of the shortened data series, where the large estimated probabilities on the IN-SV model provides a reliance on the median forecasts of y_t on the μ parameter. As noted previously, this parameter acts a trend term for the entire r_t series. Hence, as the end point alters, so does the posterior mean estimates of μ. The observed data fall below the 90th percentile in each of the in-sample forecasts of r_t. Forecasts from a single time series model, such as the AR(5) model discussed

Figure 7: Comparisons of Joint Predictive Probability Distribution of the Model Averaged Growth Rates (top) and Resulting Population Forecast in Millions (bottom) When Forecasted Volatility is Fixed at the 2006 Estimate (right) or Model Based (left).



in the previous sub-section, have narrower intervals and hence would provide less support for the resulting observations.

5.3 Constant Future Volatility

In this section, we compare the model averaged forecast that allows future uncertainty in volatility with one that keeps volatility constant at the last observed level. This is motivated by the reduced volatility exhibited in more recent years (see Figure 3), where predictive distributions of the volatility, observed in historical periods (e.g., during epidemics or times of war) might be deemed unrealistic for plausible future forecasts. Using a modification the WinBUGS code, the volatility estimated in 2006 are fixed for all SV models. The posterior predictive distributions of the model averaged forecasts were then calculated using the probabilities provided in Table 2. The resulting forecasts of r_t and p_t are shown on the right hand panels of Figure 7, where the results in Figure 4 are reproduced on the left hand panel.

Unsurprisingly, comparisons between the left and right hand panels results in narrower predictive distributions when the volatility is fixed at 2006 levels. The difference between the

20th percentile and 80th percentile of p_{2033} reduces from 10.5 million to 6.8 million. The median, as expected, remains the same, at 64.0 million.

6 Conclusion

In this paper, we have demonstrated the use of Bayesian time series methods for the forecasting of the future population of England and Wales by using a historical series of population growth rates. The forecasts have explicitly allowed for uncertainties in the data, parameters of the model and the model itself by using probability distributions, which are fully represented in the final probabilistic population forecast.

Volatility in the observed y_t series was handled using SV models. Alternative approaches to deal with non-constant variances are available, such as autoregressive conditional heteroscedastic models (ARCH) or generalised ARCH models (GARCH); see for example Chatfield (2003). This family of models imply a deterministic structure for σ_t in addition to the mean structure for y_t. Further extensions to the modelling of the growth rate can also be explored by decomposing r_t into demographic components of population change. Separate series of births, deaths and migration can be modelled as a multivariate time series process using Bayesian Vector Autoregressive (VAR) models. This decomposition may be further continued by modelling subnational populations (and the flows between them) or by incorporating age structures. These extensions, which we are currently investigating, are likely to reduce the uncertainty of population forecasts in comparison to those presented in this paper.

As Booth (2006) notes, the incorporation of informed judgements have formed the basis of many of the assumptions in traditional population projections. She also notes that methods tend to be unsystematic or inadequately documented, even in developed countries. The Bayesian approach allows data and uncertainty in parameters and model choice to be fully quantified using probability distributions. In our implementation, prior informed opinion (which was deliberately kept uninformative) had minimal influence on the final forecasts. Further work, in collecting expert opinion and translating these to priors for Bayesian time series models, may lead to alternative forecasts and reduced levels of uncertainty.

Time series models were used in this paper to forecast future population growth. The medians of our predictive distributions for future populations are slightly higher, but not drastically different to, the principal projection estimated by the ONS using a more complex cohort component methodology. Such methodologies require a large amount of data on current age and sex structures, along with numerous assumptions on rates of demographic components. However, unlike the more complex, yet deterministic, cohort component method, the forecasting methods used in this paper are able to quantify our uncertainty through a posterior predictive distribution.

Our model averaged posterior predictive distribution tended to be wider than those provided by predictive intervals from traditional frequentist time series methods. This is not unexpected as intervals for a single model selected on the basis of a model fit statistics (such as the AIC or BIC) will tend to be too narrow since they do not take into account model uncertainty (Chatfield, 2003, p86). Thus, the use of model averaging allows a more realistic picture of the uncertainty of future population to be obtained.

REFERENCES

Ahlburg, D. (1987) Population Forecasts for South Pacific Nations using Autoregressive Models 1985 - 2000. Journal of Population Research, 4(2), 157-167.

Akaike, H. (1973) Information Theory and an Extension of the Maximum Likelihood Principle. Pages 267-281 of: 2nd International Symposium on Information Theory, Tsahkadsor, Armenian SSR.

Alho, J.M. and Spencer, B.D. (1985) Uncertain Population Forecasting. Journal of the American Statistical Association, 80(390), 306-314.

Alho, J.M. and Spencer, B.D. (2005) Statistical Demography and Forecasting. Springer.

Alkema, L. Raftery, A.E. Gerl, P. Clark, S.J. and Pelletier, F. (2008) Estimating the Total Fertility Rate from Multiple Imperfect Data Sources and Assessing its Uncertainty. Tech. rept. 89. Center for Statistics and the Social Sciences.

Alkema, L. Raftery, A.E. Gerland, P. Clark, S.J. Pelletier, F. and Buettner, T. (2010) Probabilistic Projections of the Total Fertility Rate for All Countries. Tech. rept. 97. Center for Statistics and the Social Sciences, University of Washington.

Anderson, M. (ed). (1996) British Population History: From the Black Death to the Present Day (New Studies in Economic and Social History). Cambridge University Press.

Bijak, J. (2010) Forecasting International Migration in Europe: The Bayesian Approach. Springer.

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

Brierley, M.J. Forster, J.J. McDonald, J.W. and Smith, P.W.F. (2008) Bayesian Estimation of Migration Flows. Chap. 7, pages 149-174 of: Raymer, J., & Willekens, F. (eds), International Migration in Europe. Chichester, United Kingdom: Wiley.

Chatfield, C. (2003) The Analysis of Time Series: An Introduction. 6th edition. Chapman and Hall/CRC.

Chunn, J.L. Raftery, A.E. and Gerland, P. (2010) Bayesian Probabilistic Projections of Mortality. In: Annual Meeting of the Population Association of America. Population Association of America, University of Washington.

Coleman, D. and Salt, J. (1992) The British Population: Patterns, Trends, and Processes. Oxford University Press, USA.

Daponte, B.O. Kadane, J.B. and Wolfson, L.J. (1997) Bayesian Demography: Projecting the Iraqi Kurdish Population, 1977-1990. Journal of the American Statistical Association, 92(440), 1256-1267.

Gelman, A. Carlin, J.B. Stern, H.S. and Rubin, D.B. (2003) Bayesian Data Analysis. 2nd edition. Chapman & Hall/CRC.

Girosi, F, and King, G. (2008) Demographic Forecasting. Princeton University Press.

Gorbey, S. James, D. and Poot, J. (1999) Population Forecasting with Endogenous Migration: An Application to Trans-Tasman Migration. International Regional Science Review, 22(1), 69-101.

Hinde, A. (2003) England's Population: A History since the Domesday Survey. Bloomsbury.

Hoeting, J.A. Madigan, D. Raftery, A.E. and Volinsky, C.T. (1999) Bayesian Model Averaging: A Tutorial. Statistical Science, 14(4), 382-401.

Keilman, N. (2001) Demography. Uncertain population forecasts. Nature, 412(6846), 490-491.

Keilman, N, and Pham, D.Q. (2004) Time Series Based Errors and Empirical Errors in Fertility Forecasts in the Nordic Countries. International Statistical Review, 72(1), 5-18.

Keilman, N. Pham, D.Q. and Hetland, A. (2002) Why population forecasts should be probabilistic - illustrated by the case of Norway. Demographic Research, 6, 409-454.

Kefitz, N. (1991) Uncertainty in National Population Forecasting: Issues, Backgrounds, Analyses, Recommendations. Population Studies, 45(3), 545-546.

Lee, R.D. (1998) Probabilistic Approaches to Population Forecasting. Population and Development Review, 24, 156-190.

Lee, R.D. and Tuljapurkar, S. (1994) Stochastic Population Forecasts for the United States: Beyond High, Medium, and Low. Journal of the American Statistical Association, 89(428), 1175-1189.

Lunn, D.J. Thomas, A. Best, N. and Spiegelhalter, D. (2000) WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing, 10(4), 325-337.

Lutz, W. and Goldstein, J.R. (2004) Introduction: How to Deal with Uncertainty in Population Forecasting? International Statistical Review, 72(1), 1-4.

Lutz, W. Sanderson, W. and Scherbov, S. (2001) The end of world population growth. Nature, 412(6846), 543-545.

Meng, X.L. and Wong, W.H. (1996) Simulating Ratios of Normalizing Constants via a Simple Identity: A Theoretical Exploration. Statistica Sinica, 6, 831-860.

O'Hagan, A. and Forster, J. (2004) The Advanced Theory of Statistics, Vol. 2B: Bayesian Inference. 2nd edition. Wiley.

Pedroza, C. (2006) A Bayesian forecasting model: predicting U.S. male mortality. Biostatistics, 7(4), 530-550.

Paumer, P. (1992) Forecasting US population totals with the Box-Jenkins approach. International Journal of Forecasting, 8(3), 329-338.

R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raftery, A.E. (1995) Bayesian Model Selection in Social Research. Sociological Methodology, 25, 111-163.

Raymer, J. Forster, J.J. Smith, P.W.F. Bijak, J. Wisniowski, A. and Abel, G.J. (2010) The IMEM model for estimating international migration ows in the European Union. In: Joint UNECE/Eurostat Work Session on Migration Statistics.

Saboia, J.L.M. (1974) Modeling and Forecasting Populations by Time Series: The Swedish Case. Demography, 11(3), 483-492.

Schwarz, G. (1978) Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461-464.

Statistics Netherlands (2005) Changing Population of Europe: Uncertain Future. Final report of the project "Uncertain Population of Europe" (UPE). Tech. rept. Division of Social and spatial statistics, Department of Demography, Statistics Netherlands.

Tuljapurkar, S. (1999) Validation, probability-weighted priors, and information in stochastic forecasts. International Journal of Forecasting, 15(3), 259-271.

Wright, E. (2010) 2008-based national population projections for the United Kingdom and constituent countries. Population Trends, March, 91-114.

Wrigley, E.A. and Schofield, R.S. (1989) The Population History of England 1541-1871. Cambridge University Press.

ESRC Centre for Population Change
Building 58, Room 2001
Faculty of Social and Human Sciences
University of Southampton
SO17 1BJ

T: +44 (0)2380 592579
E: cpc@soton.ac.uk
www.cpc.ac.uk

To subscribe to the CPC newsletter and keep up-to-date with research activity, news and events, please register online: www.cpc.ac.uk/newsletter

You can also follow CPC on Twitter, Facebook and Mendeley for our latest research and updates:

www.facebook.com/CPCpopulation

www.twitter.com/CPCpopulation

www.mendeley.com/groups/3241781/centre-for-population-change

The ESRC Centre for Population Change (CPC) is a joint initiative between the University of Southampton and a consortium of Scottish universities including St Andrews, Edinburgh, Stirling and Strathclyde, in partnership with the Office for National Statistics and National Records of Scotland.