Sequential Monte Carlo Methods for United Kingdom Population Dynamics

Zikai Liu

March 4, 2024

Contents

1 Abstract

2 Introduction

The study of population dynamics, particularly in the context of human populations, is a critical area of research with significant implications for policy and planning. Traditionally, deterministic models such as logistic growth, and time series of population counts, have been used to infer the demographic processes driving observed population fluctuations [3]. However, this approach presents certain challenges. Changes in observed population size reflect a combination of demographic variability and measurement error, making it difficult to accurately discern underlying demographic trends [4]. In this context, state space models (SSMs) have become increasingly popular for fitting population dynamics. These models offer a flexible framework for modeling time series data, accommodating (in this case) the intrinsic variability in population processes and the observation error inherent in population counts. The state space approach separates the true population state, which is often unobservable, (see Table 1 for references) from the observation process, allowing for a more nuanced understanding of population dynamics. In the case of human population dynamics, this account for the complexity heightened by factors such as international migration, age structure, and spatial variation.

To compile these models effectively, we employ a Bayesian inference framework. This involves specifying priors for demographic parameters of interest, such as birth and death rates, as well as the initial state if it is not assumed known [10]. Given a set of observations, denoted as $y_{1:T}$, we then derive a posterior distribution for these parameters using Bayes' theorem:

$$p(\theta, s_{1:T}|y_{1:T}) \propto p(y_{1:T}|s_{1:T}, \theta)p(s_{1:T}|\theta)p(\theta) \tag{1}$$

where θ represents the parameters, $p(\theta)$ is the prior distribution, $p(s_{1:T}|\theta)$ is the prior distribution of the states, and $p(y_{1:T}|s_{1:T}, \theta)$ is the likelihood of the data given the states and the parameters. However, in many state space models, the analytic calculation of the likelihood $p(y_{1:T}|s_{1:T}, \theta)$ is prohibitively difficult. This intractability arises due to the hidden states in the model and is a challenge for both Bayesian and likelihood-based methods.

Approximate methods such as Kalman Filters and Monte Carlo algorithms are often used to overcome this issue. Kalman Filters are particularly

Term	Definition
State	s_t represents the unobserved state (hidden states) of the system at time t, which is the quantity of interest
Observation	y_t represents the observed measurements at time t
Process Equation	$p(s_t|s_{t-1})$ represents the probabilistic transition dynamics of the s_t in the system, capturing the temporal dependencies from one state to the next.
Observation Equation	$p(y_t|s_t)$ models the probabilistic relationship between the hidden states and the observed measurements, accounting for the uncertainty in observations due to **measurement errors**
Smoothing	Using all T the observations, $y_{1:T}$, to estimate the $\hat{s}_{t|1:T}$
Filtering	Using observations up to and including time t, $y_{1:t}$, to estimate $\hat{s}_{t|1:t}$
Forecasting	Using observations up to s time steps before t, denoted as $y_{1:t-s}$, to predict the state at t

Table 1: Definitions of terms used in state-space models

useful for linear models with Gaussian errors, providing an efficient recursive solution to the filtering problem [2]. Sequential Monte Carlo (SMC) methods are simulation techniques that enable the estimation of posterior distributions in complex probabilistic models by using a series of weighted random samples across time [5]. Particle Filtering, a type of SMC algorithms, provide a flexible solution to handle non-linear models with non-Gaussian errors. These methods are essential tools in the estimation of measurement error, the elucidation of hidden demographic processes, and ultimately, the accurate estimation of population size.

For the research presented in this dissertation, we have used extensive data from the UK National Statistics, covering the period from 1839 to 2021. This data comprises detailed records of births, deaths, population estimates, and migration trends[1] in England and Wales. By employing state space models in conjunction with Kalman Filter and Particle Filtering techniques, our objective is to construct a robust model that effectively captures the nuances of population dynamics in the UK. This model aims to disentangle the complexities of demographic factors such as birth rates, death rates, and mi-

---
[1]The migration data used in this study only includes the years 2000 to 2021.

gration trends, to provide a clearer understanding of population fluctuations over the past 150 years.

In our future work, we aim to extend our current model to separately analyze the population dynamics in England, Scotland, Wales, and Northern Ireland. We will leverage the datasets provided by the National Records of Scotland and the Northern Ireland Statistics and Research Agency (NISRA). These datasets include time series data on births, deaths, migration, and population estimates dating back to the mid-19th century.

In addition to the Sequential Monte Carlo and Kalman Filter methods used in this study, we also plan to explore the use of Markov Chain Monte Carlo (MCMC) methods for model fitting in our future work. This approach will provide us with another perspective and potentially enhance our ability to capture the unique demographic trends and influences in each of these regions.

3 Methods

3.1 Data Sources

3.1.1 Population Data

The data source compromises National and subnational population estimates for the UK and its constituent countries, the number of births and deaths, and the number of international and internal migrants. The data is collected and published by the Office for National Statistics (ONS) and the National Records of Scotland (NRS) and the Northern Ireland Statistics and Research Agency (NISRA).It includes all usual residents regardless of nationality, incorporating international migrants, wirh specific historical adjustments. [7]

3.1.2 Migration Data

The net migration, calculated as the difference between the inflow and outflow of international migrants, is added to the process equations of the state-space model. The migration data, which only dates back to 2004 for England and Wales, is collected from various sources and only includes long-term migration due to the lack of a population registration system in the UK.[6]

3.2 Inference Methods for State-Space Models

In the context of State Space Models (SSMs), the unobserved states are denoted as s_t, which follow a Markov Process initiated from s_0 and adhere to a process equation $p(s_t|s_{t-1})$ for $t \geq 1$. Similarly, the observations y_t is conditionally independent given s_t with observation equation as $p(y_t|s_t)$ for $t \geq 1$. gether, they describe how the system evolves and how observations relate to the underlying state.

To facilitate inference, we define $s_{0:t} \equiv \{s_0, \ldots, s_t\}$ and $y_{1:t} \equiv \{y_1, \ldots, y_t\}$ respectively as the state and observation at time t. Our goal is to recursively estimate the posterior distribution $p(s_{0:t}|y_{1:t})$. This is also known as the smoothing distribution, defined as $s_t|y_{1:T}$, the conditional probability of the state given all observations up to time T. We may also be interested in the filtering distribution defined as $P(s_t|y_{1:t})$ conditional only observation up to time t. Applying Bayes' theorem, the posterior distribution at time t is given by:

$$p(s_{0:t} \mid y_{1:t}) = \frac{p(y_{1:t} \mid s_{0:t})p(s_{0:t})}{\int p(y_{1:t} \mid s_{0:t})p(s_{0:t}) \, ds_{0:t}} \tag{2}$$

And we can update the posterior recursively using the following steps:

$$p(s_{0:t+1}|y_{1:t+1}) = \frac{p(s_{0:t+1}, y_{1:t+1})}{p(y_{1:t+1})} \tag{3}$$

$$= \frac{p(s_{t+1}, y_{t+1}|s_{0:t}, y_{0:t})p(s_{0:t}, y_{1:t})}{p(y_{t+1}|y_{1:t})p(y_{1:t})} \tag{4}$$

$$= \frac{p(y_{t+1}|s_{t+1}, s_{0:t}, y_{0:t})p(s_{t+1}|s_{0:t}, y_{0:t})p(s_{0:t}, y_{1:t})}{p(y_{t+1}|y_{1:t})p(y_{1:t})} \tag{5}$$

$$= \frac{p(y_{t+1}|s_{t+1})p(s_{t+1}|s_t)p(s_{0:t}|y_{1:t})}{p(y_{t+1}|y_{1:t})} \tag{6}$$

$$= \frac{p(y_{t+1}|s_{t+1})p(s_{t+1}|s_t)}{p(y_{t+1}|y_{1:t})}p(s_{0:t}|y_{1:t}) \tag{7}$$

We assume conditional independence in (4.5), such that y_{t+1} depends solely on the current state s_{t+1} and the Markov chain property in (4.6), where the transition to the next state s_{t+1} relies only on the current state s_t. Hence the posterior distribution $p(s_{0:t+1}|y_{1:t+1})$ can be expressed as the observation likelihood $p(y_{t+1}|s_{t+1})$ and the state transition probability $p(s_{t+1}|s_t)$ multiplied by the posterior distribution $p(s_{0:t}|y_{1:t})$ at time t.

Evaluating the posterior distribution $p(s_{0:t}|y_{1:t})$ in (4.6) is challenging as it often involves a high-dimensional distribution known proportionally to time t [5]. To address the problem, we can either assume a specific form of the state-space model, such as a linear Gaussian model (Model 1), or use Monte Carlo (MC) Integration methods (Model 2 & 3) to approximate the posterior distribution.

3.3 Model 1 - Linear Gaussian with Kalman Filter

In this model, we employ a Kalman filter to interactively estimate the population starting from the initial state S_0 with a 'diffuse' initial state covariance matrix Σ_0 of 1000, a common approach when no prior information regarding state variance is known and which has been shown to have little influence on estimated parameters [1]. The evolution of the population state is governed by the variables T_t which represent the net effects of $\phi + \beta$ (death and birth rates) derived from previous year data The process noise H_t is assumed to be an i.i.d. normal variable with mean zero and variance estimated from the yearly differences in population.[2], capturing the inherent variability in population dynamics not explained by the birth and death rates alone:

$$S_{t+1} = T_t \cdot S_t + H_t \tag{8}$$

Observations y_t are modeled as direct measurements of the state variable with added noise, to account for the uncertainty and potential inaccuracies in population data. The observation noise G_t, assumed normal similar to H_t is set based on a 3% relative standard error of the population's prior year value:

$$Y_t = S_t + G_t \tag{9}$$

We can compile the Kalman Filter as a recursive algorithm with analytical solutions. Given initialization values S_0 and Σ_0, at time t: For each step t:

1. The prior distribution at time $t-1$ is summarized by a Gaussian distribution $s_{t-1} \sim N(\hat{s}_{t-1}, \Sigma_{t-1})$

2. Prior to the observation y_t, the state s_t is estimated from the process equation and the prior covariance $R_t = \Sigma_{t-1} + G_t$ [3]. So we have $s_t \sim N(G\hat{s}_{t-1}, R_t)$

---

[2]In theory H_t should follow a normal distribution with mean zero and covariance matrix Σ_{t*t} but is simplified to a single value in one dimensional.

[3]The coefficient of S_t in equation (4.8) equals 1

3. We then calculate the Innovation (Prediction Error) as $e_t = y_t - T_t \hat{s}_{t-1}$

4. The Kalman Gain $K_t = R_t T_t^T (H_t + T_t R_t T_t^T)^{-1}$

5. The Posterior for $P(s_t|y_t)$ has mean $\hat{S}_t = S_{t-1} + K_t e_t$ and Variance: $\Sigma_t = R_t - R_t T_t^T (H_t + T_t R_t T_t^T)^{-1} T_t R_t$ [9].

We iteratively apply the Kalman filter to update our estimates of the population size. The filtered estimates $P(s_t|y_{1:t})$ provide a refined view of the population size, incorporating both the model's predictions and the observed data. We also compute standard errors from the diagonal elements of the post-update covariance matrix Σ_t to construct 95% confidence intervals around the estimates.

3.3.1 Optimization for Estimating Standard Deviation

A common challenge in applying Kalman filters lies in tuning the process error H_t and observation error G_t [12]. While the Kalman filter assumes zero-mean white noise processes (4.7, 4.8), real-world population dynamics rarely exhibit the constant birth and death rates implied by the process model. Process noise, H_t, addresses this uncertainty, allowing the Kalman filter to prioritize recent measurements. A high process error (relative to population) enables the model to closely track observations, deviating from the linear trajectory suggested by the process equation. Similarly, observation noise, G_t, accounts for errors in population data collection, such as non-response and double-counting in censuses [8]. The observation error G_t is initialized based on the 3% relative standard error associated with 2021 census estimates [8]. To further refine the process error H_t, we employ Maximum Likelihood Estimation (MLE). This optimization aims to minimize the negative log-likelihood of the Kalman filter's output:[4]

$$\underset{H_t}{\text{minimize}} \ - \mathcal{L}(H_t, \bar{G}_t, Y) \tag{10}$$

We employ the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm, a quasi-Newton method, to solve this optimization problem. BFGS builds upon the principles of Newton's method which uses the following update step:

$$x_{k+1} = x_k - [H_f(x_k)]^{-1} \nabla f(x_k) \tag{11}$$

---

[4]$\bar{G}_t$ represents the default value of G_t

where:

- x_k is the current estimate of the solution,

- $\nabla f(x_k)$ is the gradient of the function f at x_k,

- $H_f(x_k)$ is the Hessian matrix of the function f at x_k.

BFGS avoids the computationally expensive task of directly calculating and inverting the Hessian matrix $[H_f(x_k)]$ at each iteration. Instead, it iteratively constructs an approximation of the inverse Hessian $[B_f(x_k)]^{-1}$ based on gradient information, guiding the search for optimal H_t [11]. We optimize the process errors H_t, with initial values set to the logarithm of their defaults to ensure positivity [2]. The inverse Hessian $[B_f(x_k)]^{-1}$ is initialized as the identity matrix I. The search direction d_k is calculated as:

$$d_k = -B_k^{-1}\nabla f(x_k) \tag{12}$$

The algorithm then finds a step size α_k that minimizes $f(x_k + \alpha_k d_k)$ as defined (4.9) above, and updates the positions $x_{k+1} = x_k + \alpha_k d_k$. We calculate change in parameter values s_k and gradient y_k as:

$$s_k = x_{k+1} - x_k = \alpha_k d_k, \ y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \tag{13}$$

Finally, we update the inverse Hessian B_{k+1}^{-1} using the BFGS formula:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{14}$$

Final estimates for H_t is obtained by exponentiating the optimized results.

3.4 Model 2 - Sequential Monte Carlo with Bootstrap Filter

To address potential nonlinearities and non-Gaussian in the population dynamics, we employ a Bayesian BRS state-space model in conjunction with Sequential Monte1 Carlo1 (SMC) methods, specifically a bootstrap filter. The model is formulated as follows with the process equation:[10]:

$$S_t \sim Poisson(\lambda = S_{t-1} + b_t - d_t + I_t) \tag{15}$$

where S_t is the surviving population at time t, b_t and d_t are the number of births and deaths at each year respectively. The death process d_t is modeled as a binomial process with probability ϕ depending on the population size S_{t-1} from the previous year. After the death process, the surviving population is updated as $S_{t-1} - d_t$. The birth process b_t is then modeled as a Poisson process with rate $\beta \times (S_{t-1} - d_t)$, based on the updated surviving population. The net immigration process, I_t, is incorporated as an exogenous input, directly augmenting the population size. Y_t, the observed population at time t, is given by:

$$Y_t \sim Lognormal(log(S_t), log(G_t)) \tag{16}$$

where G_t is the standard deviation of log-scale observation error.

Two models are compiled in this estimation process. Initially, we employ the precise death rate ϕ and birth rate β, calculated as proportions of the annual total population corresponding to the number of deaths and births, respectively. Subsequently, we estimate ϕ and β using a second model with a vague uniform distribution prior between 0 and 0.1. The initial state of the population (S_1) is modeled as a Poisson distribution with rate parameter equal to the initial population, and the observation for the initial state is modeled as a log normal distribution with mean equal to the log of S_t and a fixed observation error as 3% of the YOY population variation [6]. The model is then compiled with a bootstrap filter to estimate state sequences, and equally weighted posterior samples of state variables are used to generate time series estimates of the population states.

Steps for Each Time Step t each with M the number of particles [5]:

1. Propagation (Sample new particles from the state transition distribution): $\tilde{\mathbf{s}}_t^{(i)} \sim q(\mathbf{s}_t | \mathbf{s}_{t-1}^{(i)})$, $i = 1, ..., M$ where $q(*)$ is the proposal distribution, which here is the process equation (4.14) defined above.

2. Weighting (Calculate importance weights for each particle): $w_t^{(i)} \propto p(\mathbf{y}_t | \tilde{\mathbf{s}}_t^{(i)})$, $i = 1, ..., M$ The weights reflect how well each propagated particle explains the current observation y_t. Normalized the weights.

3. Resampling particles (with replacement proportional to their previous time step weights $w_{t-1}^{(i)}$): $\mathbf{s}_t^{(i)} \sim \tilde{\mathbf{s}}_t^{(i)}$, $i = 1, ..., M$ promotional to $w_t^{(i)}$

4. The new set of particles $\{\mathbf{s}_t^{(i)}\}_{i=1}^M$ approximates the posterior $p(\mathbf{s}_t | \mathbf{y}_{1:t})$.

The particle filtering enables us to estimate the posterior distribution of the population with a non-linear process equation in (4.9) which we cannot solve the likelihood analytically. However, in the result section, we may see that such model may suffer from 'particle degeneracy' issues [5] where only a few particles have significant weights, leading to a poor approximation of the posterior distribution.

3.5 Model 3 - Particle Filter with Dynamic Birth and Death Rates

3.5.1 Birth Growth Survival Model (BRS)

Population dynamics are shaped by the interplay of birth, aging, death, and immigration processes [7]. The Birth, Growth, Survival (BRS) model is a state-space model that represents the dynamics of a population with two states, typically representing immature and mature individuals. The model is defined by a sequence of sub-processes: survival (S), growth (R), and birth (B). Each of these sub-processes is defined as a stochastic process. The survival and birth processes are modeled as binomial processes, with the probability of survival and birth being ϕ and β respectively. The growth process is implicitly defined by the survival and birth processes. The state of the population in each year is represented by S_t, which is modeled as a Poisson process with the rate parameter being the sum of the surviving population, births, and deaths from the previous year. The observed population Y_t is modeled as a log-normal distribution with the mean being the log of the surviving population and the standard deviation being a function of the previous year's population. This formulation allows us to capture the inherent variability in population dynamics and provides a full joint distribution of the states in S_t, conditional on S_{t-1}.

3.5.2 Hard-coded Dynamics Particle Filter

Initialization: Birth and death rate particles are initialized from a normal distribution with mean μ_β and μ_ϕ and standard deviation σ_β and σ_ϕ respectively, truncated to the interval $[0, 1]$ to ensure valid probability values.

$$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2) \tag{17}$$

$$\phi_i \sim \mathcal{N}(\mu_\phi, \sigma_\phi^2) \tag{18}$$

Given the previous year's population, P_{t-1}, and a death rate particle, ϕ_i, deaths for year t are simulated using a binomial distribution:

$$D_{t,i} \sim \text{Binom}(P_{t-1}, \phi_i)$$

The surviving population after deaths, $S_{t,i} = P_{t-1} - D_{t,i}$, is used along with a birth rate particle, β_i, to simulate births using a Poisson distribution:

$$B_{t,i} \sim \text{Poisson}(\beta_i \cdot S_{t,i})$$

The total estimated population for year t, $\hat{P}_{t,i}$, is then the sum of the surviving population, births, and immigration count I_t, with an added normal variation for observation error:

$$\hat{P}_{t,i} = S_{t,i} + B_{t,i} + I_t + \epsilon_{t,i}$$

Where $\epsilon_{t,i} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and σ_ϵ is proportional to the previous population size, reflecting observation error.

The weights for each particle are updated based on the likelihood of the observed population given the estimated population, using a Poisson likelihood:

$$w_{t,i} = w_{t-1,i} \cdot \text{Poisson}(P_t | \hat{P}_{t,i})$$

Feature	Model 2: Hard-coded Bayesian Particle Filter	Model 3: Nimble SMC with Bootstrap Filter
Initialization	Birth and death rate particles initialized with $\mathcal{N}(\mu, \sigma^2)$ and truncated to $[0, 1]$	Specific ϕ and β initially used, then estimated with uniform priors; S_1 modeled with Poisson distribution
Process Model	Binomial death process and Poisson birth process, incorporating immigration and normal variation for observation error	Sequential steps with propagation, weighting, and resampling to approximate the posterior $p(S_t\|Y_{1:t})$
Observation Model	Weights updated using Poisson likelihood of observed vs. estimated population	Log-normal observation model with log-scale error, $Y_t \sim$ Lognormal$(\log(S_t), \log(G_t))$
Proposal Distribution	Implicit within the binomial and Poisson processes	Defined by the process equation, allowing for adaptation to non-linearities
Resampling Strategy	Weight-based, at each time step to address particle degeneracy	Bootstrap filter employs resampling less frequently, potentially reducing particle impoverishment
Posterior Estimation	Direct estimation of population size and parameter values	Employs SMC to estimate state sequences and generate time series estimates of the population states

4 Results

5 Analysis of Population Dynamics

5.1 England & Wales

5.2 Scotland

5.3 Northern Ireland

6 Discussion

6.1 Implications of Findings

Implications of the findings

6.2 Limitations and Future Work

Limitations of the current study and directions for future research.

6.3 Conclusion

Key findings and their significance.

A Supplementary Code & Visualizations

References

[1] Tom Aspinall. Fkf.sp: Fast kalman filtering through sequential processing.

[2] Marie Auger-Méthé, Ken Newman, Diana Cole, Fanny Empacher, Rowenna Gryba, Aaron A. King, Vianey Leos-Barajas, Joanna Mills Flemming, Anders Nielsen, Giovanni Petris, and Len Thomas. A guide to state–space modeling of ecological time series. *Ecological Monographs*, 91(4):e01470, 2021.
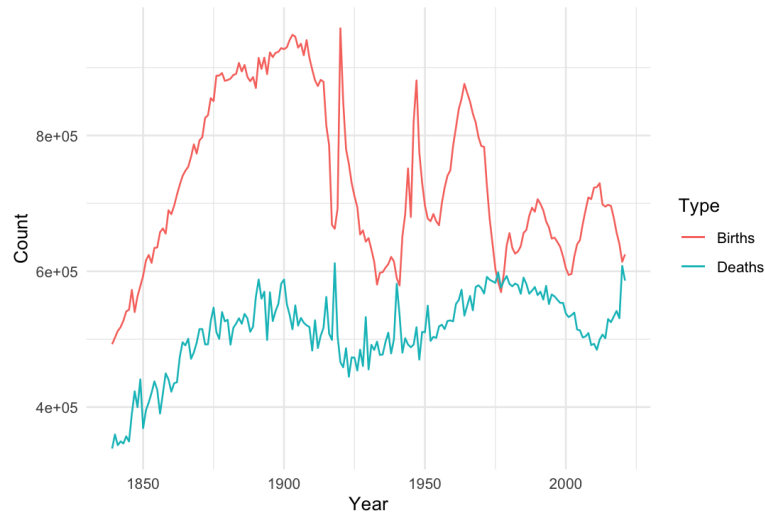
Figure 1: Birth and Death- England & Wales - 1838-2021

[3] N. Cappuccino and P. W. Price. *Population Dynamics: New Approaches and Synthesis*. Academic Press, San Diego, CA, 1995.

[4] J. S. Clark and O. N. Bjørnstad. Population time series: Process variability, observation errors, missing values, lags, and hidden states. *Ecology*, 85:3140–3150, 2004.

[5] A. Doucet, N. de Freitas, and N. Gordon. *An Introduction to Sequential Monte Carlo Methods*, page 3–14. Springer, New York, NY, 2001.

[6] National Statistics Centre for Demography. Estimating international migration for population estimates - an information paper. Technical report, 2005.

[7] Office for National Statistics. Population estimates for the uk, england, wales, scotland and northern ireland: mid-2021, 2022.

[8] Office for National Statistics (ONS). Coverage estimation for census 2021 in england and wales. *ONS website*, November 2022.

[9] Richard Meinhold and Nozer Singpurwalla. Understanding the kalman filter. *American Statistician*, 37:123–127, 05 1983.
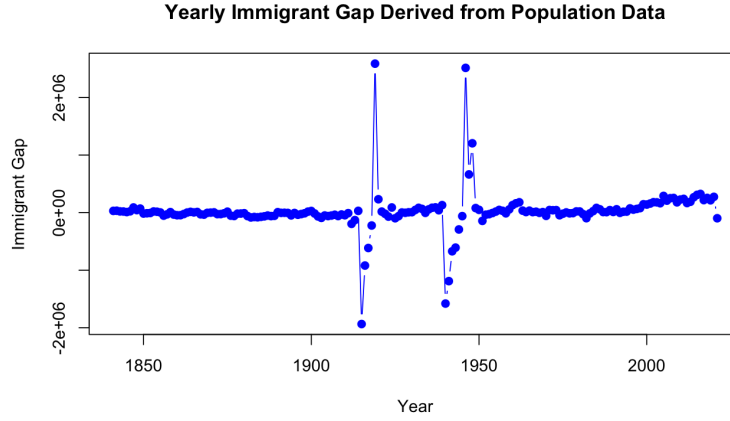
Yearly Immigrant Gap Derived from Population Data

Figure 2: Immigration Gap England - 1838-2021

[10] K. B. Newman, S. T. Buckland, B. J. T. Morgan, R. King, D. L. Borchers, and D. J. Cole. *Modelling Population Dynamics: Model Formulation, Fitting and Assessment Using State-Space Methods*. Methods in Statistical Ecology. Springer, New York, NY, 1 edition, 2014.

[11] Jorge Nocedal and Stephen J. Wright. *Quasi-Newton Methods*, pages 135–163. Springer New York, New York, NY, 2006.

[12] Bernt M. Åkesson, John Bagterp Jørgensen, Niels Kjølstad Poulsen, and Sten Bay Jørgensen. A generalized autocovariance least-squares method for kalman filter tuning. *Journal of Process Control*, 18(7):769–779, 2008.
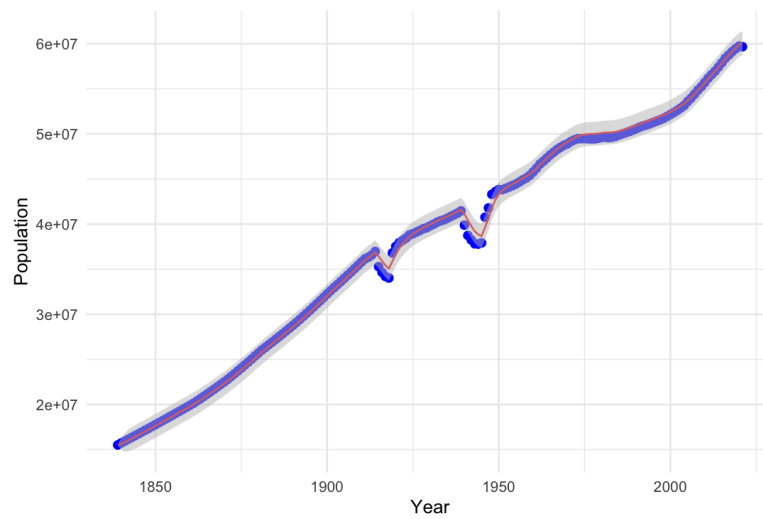
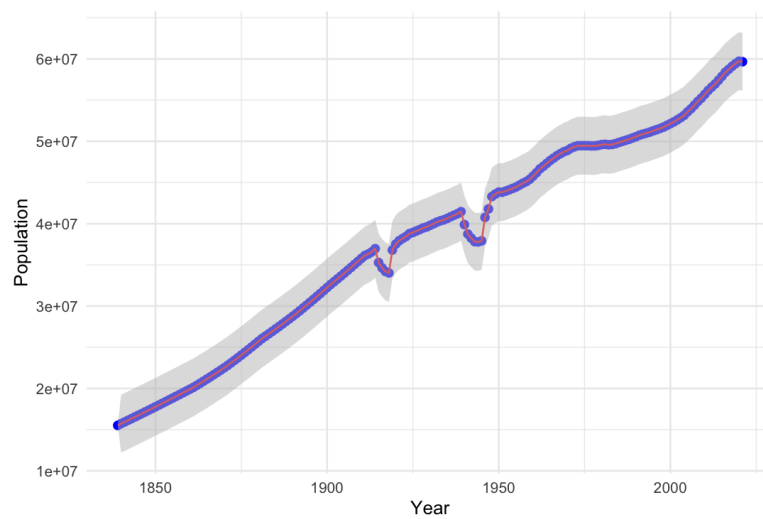Figure 3: Model 1 - Kalman Filter Default Error- England & Wales

Figure 4: Model 1 - Kalman Filter Optim Error- England & Wales

```r
stateSpaceModelCode <- nimbleCode({
  # Priors
  phi ~ dunif(0, 0.1)   # phi (death rate)
  beta ~ dunif(0, 0.1)  # beta (birth rate)

  # Initial state
  S[1] ~ dpois(lambda = N_initial) # Initial population
  Y[1] ~ dlnorm(meanlog = log(S[1]), sdlog = log(sdo)) # Initial state observation

  # Process model
  for (t in 2:T) {
    # Binomial death process
    d[t] ~ dbin(size = S[t-1], prob = phi)

    # Surviving population after deaths but before births
    S_post_death[t] <- S[t-1] - d[t]

    # Poisson birth process based on the updated surviving population
    b[t] ~ dpois(lambda = beta * S_post_death[t])

    # Update total surviving population including new births
    S[t] ~ dpois(lambda = S_post_death[t] + b[t])

    # Observation process
    Y[t] ~ dlnorm(meanlog = log(S[t]), sdlog = log(sdo))
  }
})
```
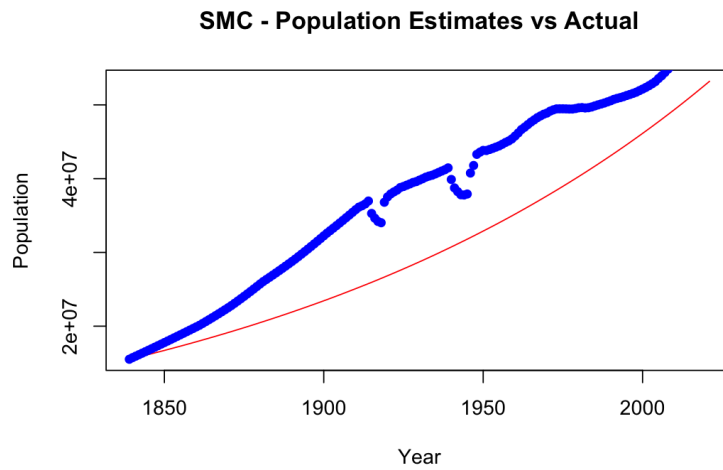
Figure 5: Model 3 - R Code- England & Wales

Figure 6: Model 3 - Bootstrap Filter with BRS Model- England & Wales