# Sequential Monte Carlo Methods for United Kingdom Population Dynamics

Zikai Liu

March 8, 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 [2]. 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 [3]. 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) 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.

In this study, we apply three distinct state space models, delineating our analysis into two distinct inferential paradigms: frequentist statistical inference and Bayesian inference. The frequentist approach is exemplified by our first model, which employs Kalman Filters. Kalman Filter is a special case of linear SSMs with Gaussian errors and initial states, enabling an efficient recursive solution to the filtering problem [8]. Unlike Bayesian methods that integrate prior knowledge (of parameters and initial states) into analysis, the Kalman Filter adheres to the principles of Maximum Likelihood Estimation (MLE), focusing on estimate the hidden states of the model given a sequence of noisy observations. Although the two-stage predictive and updates steps can be linked to the posterior update of Bayesian inference, the Kalman Filter contrast the posterior density approach embodied by subsequent models. Our second and third models adopt a Bayesian approach, employing Sequential Monte Carlo (SMC) methods – namely, Particle Filtering – to approximate posterior distributions in non-linear models with non-Gaussian error structures [4]. These models allow for vague (or informative) prior distributions

| Term | Definition |
| --- | --- |
| Filtering | $p(x_t|y_{1:t})$ use observations to time $t$, $y_{1:t}$, to estimate $x_t$ |
| Forecasting | $p(x_t|y_{1:t-s})$ use observations $s$ time steps before $t$ to predict $x_t$ |
| Observation | The observed measurements $y_t$ at time $t$ |
| Observation Equation | $g(y_t|x_t)$, the probabilistic relationship between the hidden states and the observed measurements |
| Observation Error | The gap between the observed data and the true state, often modeled as noise in the observation equation. |
| Online Estimation | Continually estimate $x_t$ based on information available at time $t$ |
| Process Equation | $f(x_t|x_{t-1})$ represents the probabilistic transition dynamics of the $x_t$ in the system, capturing the temporal dependencies from one state to the next. |
| Process Error | The discrepancy between the predicted state and the actual hidden state, often modeled as noise in the process equation. |
| Smoothing | $p(x_t|y_{1:T})$ use all $T$ the observations to estimate the $x_t$ |
| State | $x_t$ represents the unobserved state (hidden states) of the system at time $t$, which is the quantity of interest |
| State Space Model | A statistical model to capture the system dynamics through a set of **process equations** for the state evolution over time and **observation equations** linking the hidden states to the observed data. |

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

on parameters like birth and death rates. The second model simulates population dynamics using a Bootstrap Filter through the `nimbleSMC` packages in R, capturing birth, death, and immigration processes. Our third model refines this approach, incorporating dynamic birth and death rates within a hard-coded particle filter algorithm. This model accounts for the stochasticity of demographic events, offering a nuanced understanding of population changes over time. These methods are crucial for accurately estimating measurement error, uncovering hidden demographic processes, and ultimately, determining population size with precision.

For the research presented in this dissertation, we have used extensive

data from the UK National Statistics, covering the period from 1839 to 2023. 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 migration trends, to provide a clearer understanding of population fluctuations over the past 150 years.

# 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 are 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). They include all usual residents regardless of nationality, incorporating international migrants, with specific historical adjustments [6].

### 3.1.2 Migration Data

The net migration was calculated as the difference between the inflow and outflow of international migrants. The migration data published by ONS, which only dates back to 1964 for England and Wales, was collected from various sources and only includes long-term migration due to the lack of a population registration system in the UK [5].

---

[1]The migration data used in this study only includes the years 1964 to 2023.

## 3.2 Inference Methods for State-Space Models

In the context of State Space Models (SSMs), the unobserved states are denoted as $x_t$, which follow a Markov Process [2] initiated from $x_0$ and adhere to a process equation $p(x_t|x_{t-1})$ for $t \geq 1$ [8]. It can be written as:

$$x_t = f(x_{t-1}, \eta_t) \tag{1}$$

Where $f$ is a function that describes the state transition, and $\eta_t$ is the process noise.

Similarly, the observations $y_t$ are conditionally independent given $x_t$ with observation equation as $p(y_t|x_t)$ for $t \geq 1$, which can be written as:

$$y_t = g(x_t, \epsilon_t) \tag{2}$$

Where $g$ is a function that describes how the state is observed, and $\epsilon_t$ is the observation noise. Together, they describe how the system evolves and how observations relate to the underlying state.

As mentioned earlier, we can fit the SSMs with both frequentist and Bayesian inference frameworks. For the frequentist approach, such as the Kalman Filter, we estimate the hidden states using the conditional distribution of the states given the observations and the fixed (or estimated) parameter values. The joint likelihood $L_j$ for the states $x_{1:T}$ given a length $T$ observations $y_{1:T}$ is defined as:

$$L_j(x_{1:T}|y_{1:T}) = \prod_{t=1}^{T} g(x_t, \epsilon_t) f(x_{t-1}, \eta_t), \tag{3}$$

For our online estimation setting, to estimate the state with a known parameter values $\hat{\theta}$, this (the probability distribution of states) can be represented as a filtering problem:

$$\hat{p}(x_{1:t}|y_{1:t}, \hat{\theta}) = \frac{L_j(x_{1:t}|y_{1:t}, \hat{\theta})}{\int L_j(x_{1:t}|y_{1:t}, \hat{\theta}) \, dx_{1:t}} \tag{4}$$

Where the marginal likelihood, i.e. the integral in the denominator ensures that the probabilities sum to one.

---

[2]A Markov process is a stochastic process where the future state depends solely on the present state, not the entire past history of the process.

For the 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 [11]. To facilitate inference, we define $x_{0:t} \equiv \{x_0, \ldots, x_t\}$ and $y_{1:t} \equiv \{y_1, \ldots, y_t\}$ respectively as a set of states and observations at time $t$. Our goal is to recursively estimate the posterior distribution $p(x_{0:t}|y_{1:t})$. This is also known as the smoothing distribution, defined as $x_t|y_{1:T}$, the conditional probability of the state given all observations up to time $T$. We may also be interested in the filtering distribution defined as $p(x_t|y_{1:t})$ conditional only observation up to time $t$. Applying Bayes' theorem, the posterior distribution at time $t$ is given by:

$$p(\theta, x_{0:t}|y_{1:t}) = \frac{p(y_{1:t}|x_{0:t}, \theta)p(x_{0:t}|\theta)p(\theta)}{\int p(y_{1:t}|x_{0:t}, \theta)p(x_{0:t}|\theta)p(\theta) \, d\theta \, dx_{0:t}} \tag{5}$$

Where $\theta$ represents the parameters, $p(\theta)$ is the prior distribution, $p(x_{1:t}|\theta)$ is the **prior distribution** of the states, and $p(y_{1:t}|x_{1:t}, \theta)$ is the likelihood of the data (**joint likelihood**) given the states and the parameters, and the denominator integral is the **Marginal Likelihood** serves as a normalization factor, ensuring the posterior distribution sums up to 1 [10].

To simplify the estimation, we'll assume the parameters $\theta$ are known and fixed, hence we can update the posterior recursively using the following steps:

$$
\begin{aligned}
p(x_{0:t+1}|y_{1:t+1}) &= \frac{p(x_{0:t+1}, y_{1:t+1})}{p(y_{1:t+1})} \\
&= \frac{p(x_{t+1}, y_{t+1}|x_{0:t}, y_{0:t})p(x_{0:t}, y_{1:t})}{p(y_{t+1}|y_{1:t})p(y_{1:t})} \\
&= \frac{p(y_{t+1}|x_{t+1}, x_{0:t}, y_{0:t})p(x_{t+1}|x_{0:t}, y_{0:t})p(x_{0:t}, y_{1:t})}{p(y_{t+1}|y_{1:t})p(y_{1:t})} \\
&= \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)p(x_{0:t}|y_{1:t})}{p(y_{t+1}|y_{1:t})} \\
&= \frac{p(y_{t+1}|x_{t+1})p(x_{t+1}|x_t)}{p(y_{t+1}|y_{1:t})}p(x_{0:t}|y_{1:t})
\end{aligned}
$$

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

Evaluating the marginal likelihood in (4) & (5) is challenging as it often involves a high-dimensional integration known proportionally to time $t$ [4]. For example, for a SSM with a 1-dimensional state at each time step. If we want to evaluate the posterior distribution at time $t = 100$, we would need to compute a 100-dimensional integral. This is because the denominator in Bayes' theorem, as shown in (5), involves an integral over all possible values of $x_{0:t}$, which is a $t$-dimensional integral. The denominator of this equation involves an integral over all possible values of $x_{0:t}$, which is a $t$-dimensional integral. As $t$ increases, the dimensionality of this integral also increases, making it more difficult to compute. To address the problem, we can either assume a specific form of the state-space model, such as a linear Gaussian model (Model 3.3) by maximizing the likelihood, or use Monte Carlo (MC) Integration methods (Model 3.4 and 3.5) to approximate the posterior distribution.

## 3.3 Model 1 - Linear Gaussian with Kalman Filter

In this model, we assume that the population dynamics can be described by a linear Gaussian SSMs, suitable to be fit with a Kalman Filter using R packages `Fast Kalman filter`. The state-space model is defined by the following process and observation equations:

$$x_t = c_t + T_t \cdot x_{t-1} + \eta_t, \tag{6}$$

where $\eta_t \sim \mathcal{N}(0, H_t)$. Deriving from $x_t = c_t + T_t \cdot x_{t-1} + Q_t \cdot \eta_t$ with $d_t$ being a vector of annual net migration, $\eta_t$ was assumed to be mean zero and variance $H_t$, and $Q_t$ equals identity.

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 $\epsilon_t$ was assumed normal with mean zero and variance $G_t$:

$$y_t = x_t + \epsilon_t, \tag{7}$$

where $\epsilon_t \sim \mathcal{N}(0, G_t)$. Deriving from $y_t = d_t + Z_t \cdot x_t + K_t \cdot \eta_t$ with $d_t$ being a vector of annual net migration, $Z_t$ equal 1 and $\eta_t$ was assumed to be mean zero and variance $G_t$, and $K_t$ equals identity.

- $x_t$ is the state of the system at time $t$, with the starting value $x_0$ being the initial population count (in 1839).

8

- $c_t$ is the number of net immigration count each year [3]

- $T_t$ is the state transition matrix, describing how the state evolves from time $t - 1$ to $t$, utilizing net effects of $\phi + \beta$ (mean death and birth rates) derived from the data. In the univariate case here, $T_t$ is a scalar, representing the growth rate of the population.

- $y_t$ is the observed population at time $t$.

- $H_t$ and $G_t$ are the process and observation noise variances, respectively. Both are assumed to be independent and identically distributed (i.i.d.) normal variables with a mean of zero [8]. The variance of $H_t$ was estimated from the year-over-year population differences to capture stochasticity not explained by process equation (6). The variance of $G_t$ was set based on 3% of the population count base on the 2021 ONS census report suggestion [7]. The implications of these assumptions will be discussed in 5.1.1.

The Kalman filter output the analytical solutions recursively in two steps: prediction and update [8]. Given the initialization values $x_0$ and the covariance matrix $P_0$, which, in this univariate case, simplifies to the variance of the estimation error. For each step at time $t - 1$:

1. Given state mean $x_{t-1}$ and variance $P_{t-1}$

2. **Prediction**

   (a) we predict the state as $\hat{x}_t = T_t x_{t-1} + c_t$

   (b) with the variance as $\hat{P}_t = T_t P_{t-1} T_t^T + H_t$ where $\hat{P}_t$ is the predicted variance at time $t$.

3. **Update**

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

   (b) The Kalman Gain $K_t = \hat{P}_t + G_t$

   (c) We then update our estimation of $x_t$ with a normal distribution of mean $x_t = \hat{x}_t + \hat{P}_t K_t^{-1} e_t$ and variance: $P_t = (I - \hat{P}_t K_t^{-1}) \hat{P}_t$ [8].

---

[3]Available data from 1964 to 2023 and set to zero otherwise.

We iteratively apply the Kalman Filter to update our population estimates. The filtering distribution $p(x_t|y_{1:t})$ provide a refined view of the population size, by adjusting model's predictions with observed data. We also compute standard errors from the diagonal elements of the post-update variance $P_t$ to construct 95% confidence intervals around the estimates.

### 3.3.1 Optimization for Estimating Standard Deviation

In the above example, we assume that the parameters (i.e. $T_t$) are known. In situations where the parameters vector $\theta$ - which may include $H_t$, $G_t$, among others - are unknown, a two-step approach is often employed. The first step involves the use of Maximum Likelihood Estimation (MLE) to estimate the values of $\theta$ that maximize the likelihood of the observed data [13]. This is achieved by integrating out the hidden states $x_{1:t}$ from the joint likelihood $L_j$ defined in Equation (3) [1]. Once the parameters are estimated, the second step involves using the estimated $\hat{\theta}$ to perform state estimation. The marginal likelihood for $\theta$ is given by:

$$
\begin{aligned}
L_m(\theta|y_{1:t}) &= \int L_j(\theta, x_{1:t}|y_{1:t}) \, dx_{1:t} \\
&= \int [g(y_1|x_1, \epsilon_1; \theta) \ldots g(y_t|x_t, \epsilon_t; \theta) \\
&\quad \times f(x_1|x_0, \eta_1; \theta) \ldots f(x_t|x_{t-1}, \eta_t; \theta)] \, dx_{1:t}
\end{aligned}
$$

The log-likelihood function $\ell(\theta)$ for a vector of unknown parameters $\theta$ is then computed as[4]:

$$
\ell(\theta) = \sum_{t=1}^{T} \log L_m(\theta|y_{1:t-1}), \tag{8}
$$

While the Kalman Filter assumes zero mean white processes (observation) noise, real-world population dynamics rarely exhibit the constant birth and death rates and immigration flows implied by the process model. Process variance, $H_t$, addresses this uncertainty, allowing the Kalman Filter to prioritize recent measurements. A high $H_t$ (relative to population) enables the model to closely track observations, deviating from the linear trajectory suggested by the process equation. Similarly, observation variance, $G_t$, accounts for errors in population data collection, such as non-response and

---

[4]Derivation of maximum likelihood estimation equation can be found in Appendix A

double-counting in censuses [7]. For the parameter optimization, we fix the observation error $G_t$ to be 3% of the population count. To further refine the process error $H_t$, we construct a function that takes $G_t$ as its sole arguments. This function is based on the negative log-likelihood function detailed in 3.2. This optimization aims to minimize the negative log-likelihood of the Kalman Filter's output:[5]

$$\underset{H_t}{\text{minimize}} \ -\mathcal{L}(H_t, \bar{G}_t, Y) \tag{9}$$

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:

$$x_{k+1} = x_k - [H_f(x_k)]^{-1}\nabla f(x_k) \tag{10}$$

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 $H_t$ [12]. We optimize the process errors $H_t$, with initial values set to the logarithm of their defaults to ensure positivity [1]. The inverse Hessian $[B_f(x_k)]^{-1}$ was initialized as the identity matrix $I$. The search direction $d_k$ was calculated as:

$$d_k = -B_k^{-1}\nabla f(x_k) \tag{11}$$

The algorithm then finds a step size $\alpha_k$ 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 $x_k$ and gradient $y_k$ as:

$$x_k = x_{k+1} - x_k = \alpha_k d_k, \ \ y_k = \nabla f(x_{k+1}) - \nabla f(x_k) \tag{12}$$

---

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

11

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

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T x_k} - \frac{B_k x_k x_k^T B_k}{x_k^T B_k x_k} \tag{13}$$

Final estimates for $H_t$ was 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 [4]. The model was formulated as follows with the process equation [11]:

$$X_t \sim Poisson(\lambda = X_{t-1} + b_t - d_t + I_t) \tag{14}$$

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

$$Y_t \sim Lognormal(log(X_t), log(G_t)) \tag{15}$$

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

Two models are fitted in this estimation process. Initially, we employ the precise death rate $\phi$ and birth rate $\beta$, calculated as proportions of the annual total population corresponding to the number of deaths and births, respectively. Subsequently, we estimate $\phi$ and $\beta$ using a second model with a vague uniform distribution prior between 0 and 0.1. The initial state of the population $(X_1)$ was modeled as a Poisson distribution with rate parameter equal to the initial population, and the observation for the initial state was modeled as a log normal distribution with mean equal to the log of $X_t$ and

a fixed observation error as 3% of the YOY population variation [5]. The model was 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 [4]:

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

2. Weighting (Calculate importance weights for each particle): $w_t^{(i)} \propto p(\mathbf{y}_t|\tilde{\mathbf{x}}_t^{(i)})$, $i = 1,...,M$ 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 $w_{t-1}^{(i)}$): $\mathbf{x}_t^{(i)} \sim \tilde{\mathbf{x}}_t^{(i)}$, $i = 1,...,M$ promotional to $w_t^{(i)}$

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

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 [4] 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 [6]. 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 $\phi$ and $\beta$ respectively. The growth

13

process is implicitly defined by the survival and birth processes. The state of the population in each year is represented by $X_t$, 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 $Y_t$ was 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 $X_t$, conditional on $X_{t-1}$.

### 3.5.2 Hard-coded Dynamics Particle Filter

**Initialization:** Birth and death rate particles are initialized from a normal distribution with mean $\mu_\beta$ and $\mu_\phi$ and standard deviation $\sigma_\beta$ and $\sigma_\phi$ respectively, truncated to the interval $[0, 1]$ to ensure valid probability values.

$$\beta_i \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2) \tag{16}$$

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

Given the previous year's population, $P_{t-1}$, and a death rate particle, $\phi_i$, 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, $X_{t,i} = P_{t-1} - D_{t,i}$, was used along with a birth rate particle, $\beta_i$, to simulate births using a Poisson distribution:

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

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

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

Where $\epsilon_{t,i} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and $\sigma_\epsilon$ 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 $\phi$ and $\beta$ initially used, then estimated with uniform priors; $X_1$ 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 $p(X_t|Y_{1:t})$ |
| Observation Model | Weights updated using Poisson likelihood of observed vs. estimated population | Log-normal observation model with log-scale error, $Y_t \sim$ Lognormal$(\log(X_t), \log(G_t))$ |
| Proposal Distribution | Implicit within the binomial and Poisson processes | Defined by the process equation, allowing for adaptation to non-linearity |
| 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

### 5.1.1 Kalman Filter

Starting with initial parameters $H_t = 1.61 \times 10^{11}$ and $G_t = 1.13 \times 10^6$, we achieved a close fit to the observations, as shown in Figure 4, with a log-likelihood of 2623.38. Optimization resulted in a slightly higher $H_t$ estimate of $1.77 \times 10^{11}$ and a new model likelihood of 2622.96. Regardless of the initial uncertainty in $P_0$, the filter converges to an optimal solution, as evidenced by the log-likelihood values for $P_0$ values of 1000 and 1000000.

As the process and observation error variances approach zero, the Kalman filter remains valid [8], but increasingly approximates a deterministic model with a near-fixed mean, as shown in Figure 3. When the process error was squared ($H_t = 4 \times 10^5$), the confidence interval significantly narrowed, resulting in more linear filtering and reduced responsiveness to new observations.

As mentioned in 3.3, the assumption of constant observation error variance is questionable given the tripling of the population over 150 years. The variability in process noise is also likely changes over time due to factors such as advancements in census methods, suggesting the need for a time-dependent $H_t$ (and $G_t$) model.

Lastly, the fitting results suggest an overestimation of the population from 2000, possibly due to the combined effects of immigration and constant birth and death rate assumptions, despite a likely decline in birth rates over time [6].

## 5.2 Scotland

## 5.3 Northern Ireland

# 6 Discussion

## 6.1 Implications of Findings

Implications of the findings

16

## 6.2 Limitations and Future Work

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.

## 6.3 Conclusion

Key findings and their significance.

# A    Proof of MLE for Kalman Filter

For the time series, a conditional probability density function is used to write the joint density function as

$$L(y_1, ..., y_T; \theta) = \prod_{t=1}^{T} p(y_t|y_{1:t-1}) \tag{18}$$

Recall from Section 3.3 that, after the prediction step, we have the conditional distribution of $y_t$ being normal with a mean $\hat{x}_t + e_t$ and variance $P_t$

$$e_t = y_t - \hat{x}_t - \epsilon_t$$

where the innovation covariance matrix $P_t$ is given by:

$$P_t = \hat{P}_t + G_t \tag{19}$$

Assuming Gaussian noise, the PDF of the innovation at time t is:

$$p(e_t|y_1, ..., y_{t-1}, \theta) = \frac{1}{\sqrt{(2\pi)^t \det(P_t)}}$$
$$\cdot \exp\left(-\frac{1}{2}e_t^T P_t^{-1} e_t\right)$$

The joint likelihood function for all observations is the product of individual:

$$L(y_1, ..., y_T; \theta) = \prod_{t=1}^{T} p(e_t|y_1, ..., y_{t-1}, \theta) \tag{20}$$

We often work with the log-likelihood function for numerical stability:

$$\log L(y_1, ..., y_T; \theta) = \sum_{t=1}^{T} \left[ -\frac{1}{2}\log(\det(P_t)) \right.$$
$$\left. -\frac{1}{2}e_t^T P_t^{-1} e_t - \frac{N}{2}\log(2\pi) \right]$$

We then able to use an optimization algorithm (Newton-Raphson or BFGS) to find the values of $T_t, H_t, G_t$ that maximize the log-likelihood function.

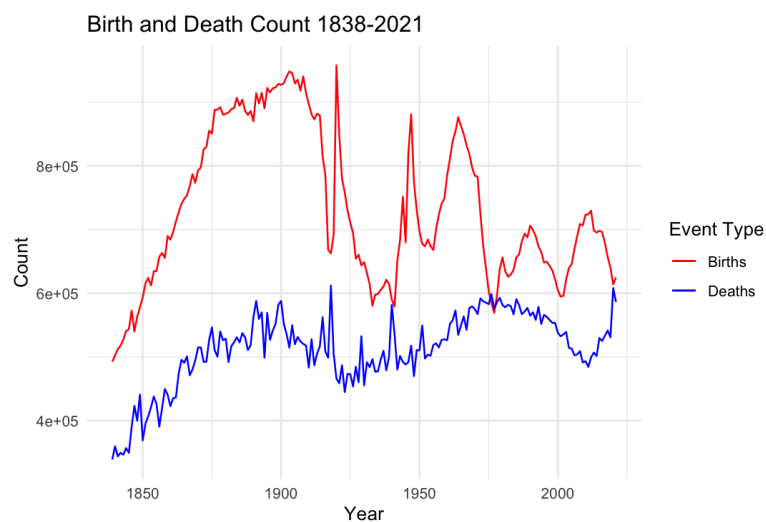# B    Supplementary Code & Visualizations

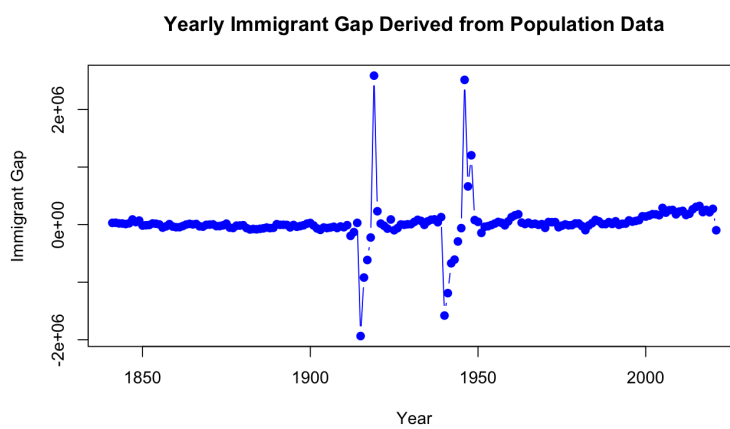Figure 1: Birth and Death - England & Wales - 1838-2021



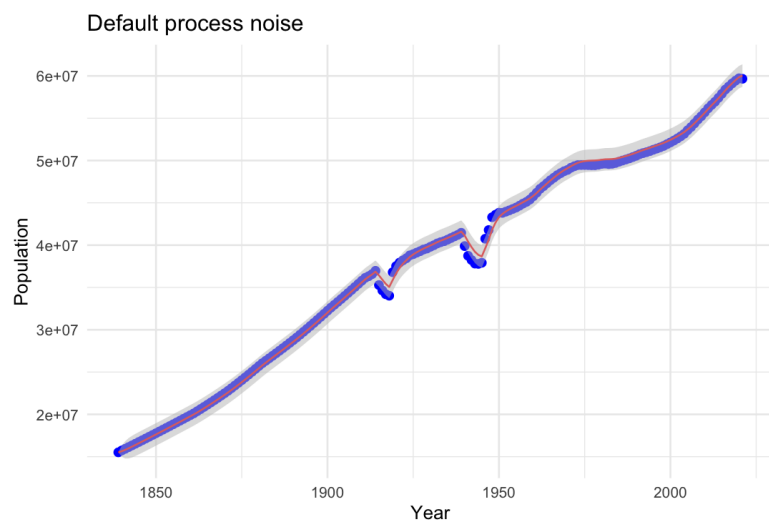Figure 2: Immigration Gap - England & Wales - 1838-2021

Figure 3: Model 1 - Kalman Filter Default Error- England & Wales
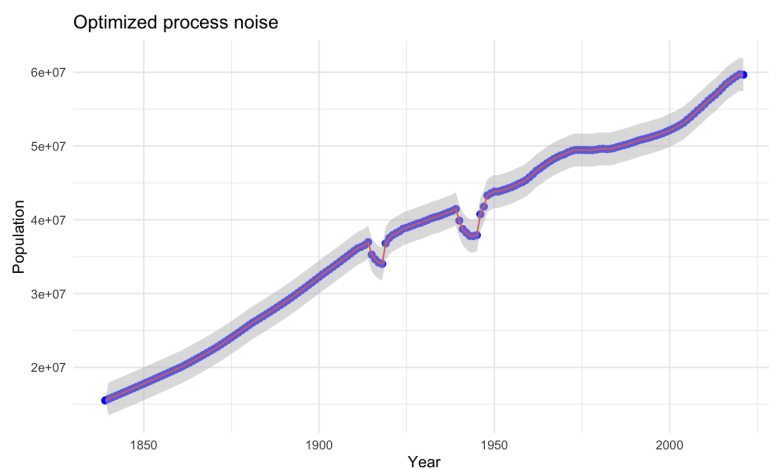


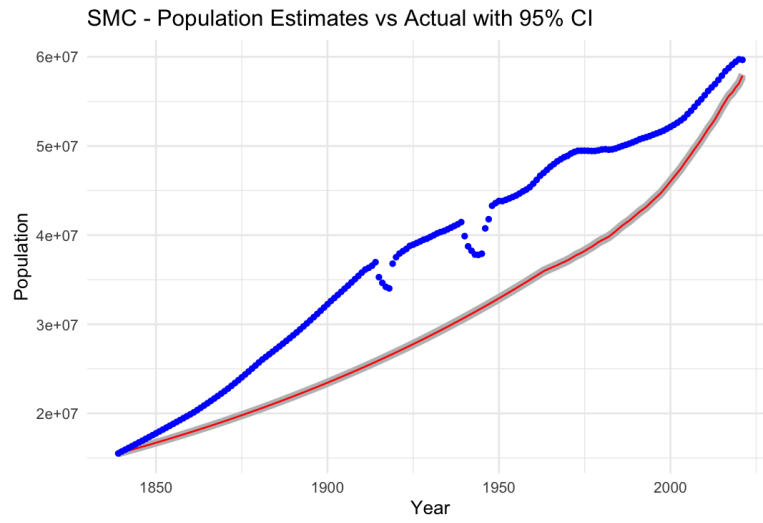Figure 4: Model 1 - Kalman Filter Optim Error- England & Wales

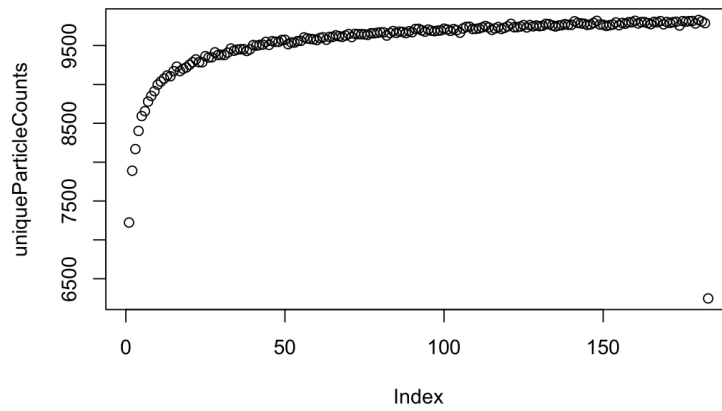Figure 5: Model 2 - SMC Bootstrap Filter - England & Wales



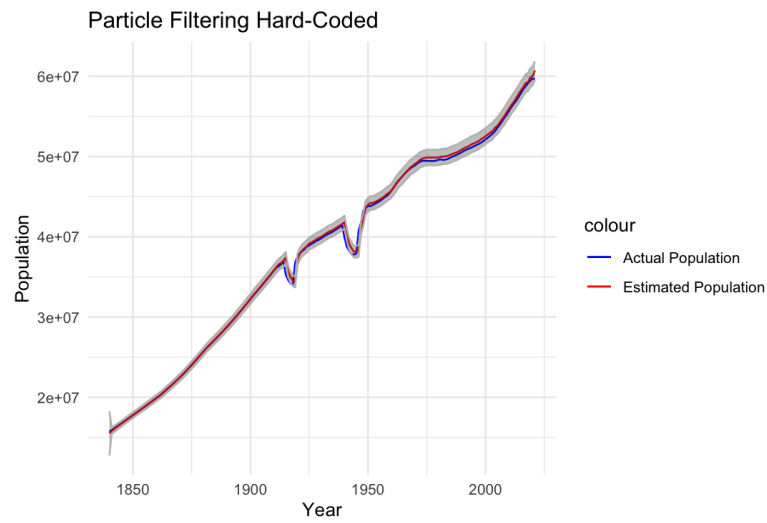Figure 6: Model 2 - Unique Particles Numbers - England & Wales

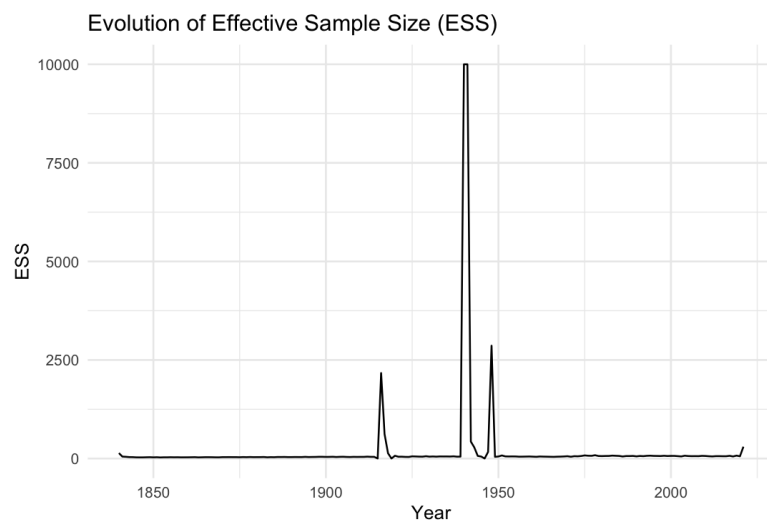Figure 7: Model 3 - SMC Hard-Coded - England & Wales



Figure 8: Model 3 - SMC Hard-Coded ESS- England & Wales

# References

[1] 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.

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

[3] 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.

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

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

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

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

[8] Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, 1990.

[9] Richard Meinhold and Nozer Singpurwalla. Understanding the kalman filter. *American Statistician*, 37:123–127, 05 1983.

[10] Nicholas Michaud, Perry de Valpine, Daniel Turek, Christopher J. Paciorek, and Dao Nguyen. Sequential monte carlo methods in the nimble and nimblesmc r packages. *Journal of Statistical Software*, 100(3):1–39, 2021.

[11] 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.

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

[13] 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.