A Particle Markov Chain Monte Carlo Approach for the Estimation of CBD-Type Models

by

Xueyi Xu

B.Sc., Simon Fraser University, 2018

Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Fall 2021

Declaration of Committee

Name:	**Xueyi Xu**
Degree:	**Master of Science**
Thesis title:	**A Particle Markov Chain Monte Carlo Approach for the Estimation of CBD-Type Models**
Committee:	**Chair:** Joan Hu Professor, Statistics and Actuarial Science
	Barbara Sanders Co-Supervisor Associate Professor, Statistics and Actuarial Science
	Jean-François Bégin Co-Supervisor Assistant Professor, Statistics and Actuarial Science
	Himchan Jeong Examiner Assistant Professor, Statistics and Actuarial Science

Abstract

The current literature on mortality has mainly focused on model specification, giving less regard to parameter estimation. Indeed, over the last three decades, multiple mortality models have been introduced, most being extensions of the well-known Lee-Carter model or the Cairns-Black-Dowd (CBD) model. However, the estimation of these models has been somewhat overlooked; most papers focus on frequentist methods, such as the (two-stage) maximum likelihood estimation method that estimates the mortality parameters first and then the parameters of the mortality improvement dynamics second. In this report, we present a new Bayesian-based estimation procedure for CBD-type models that relies on the particle Markov chain Monte Carlo (pMCMC) method of Andrieu et al. (2010). This methodology captures the dynamic nature of the mortality improvement factors (and their underlying parameters) consistently, unlike most two-stage estimation methods used in the literature.

Keywords: particle Markov chain Monte Carlo; CBD-X models; state-space model

Acknowledgements

Firstly, I would like to express my deepest gratitude to my supervisors, Prof. Barbara Sanders and Dr. Jean-François Bégin, without whom this project would not have been successful. Their invaluable knowledge and continuous guidance are incredibly useful and are greatly appreciated.

I would like to extend my thanks to Dr. X. Joan Hu for chairing my defence, and Dr. Himchan Jeong for investing his precious time in serving the examining committee, making insightful suggestions and helping refine the project. I also would like to thank all staff and faculty members of the Department of Statistics and Actuarial Science for their dedication, especially Dr. Cary Tsai, Dr. Yi Lu, Dr. Rachel Altman, Dr. Tim Swartz, and Dr. Haolun Shi, whose courses greatly broadened my knowledge in actuarial science and statistics. Furthermore, I would like to take this opportunity to thank the Graduate Student Society and the Department of Statistics and Actuarial Science at Simon Fraser University for offering financial supports throughout my graduate study.

Finally, I am always grateful to my academic fellows and loving friends for sharing their experiences and being with me during my studies at Simon Fraser University. Special thanks to my cat, Dummy, for companying with me when I stay up for studying, coding, and writing up the report, although he always falls asleep after few minutes.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Longevity risk is one of the most significant risks that insurance companies, pension plan sponsors, and government are exposed to. Longevity risk arises when people live longer than expected, which means pensions or annuities need to be paid for a much longer time than expected, resulting in higher pension plan and insurance liabilities.

To capture and understand longevity risk, a wide variety of stochastic mortality models has been developed during the last three decades. Most of these are extensions of the well-known Lee-Carter (LC) model or of the Cairns-Black-Dowd (CBD) model. Lee and Carter (1992) introduce the first stochastic mortality model, and since then, multiple LC-type mortality models were proposed; for instance, Renshaw and Haberman (2006) extend the LC model to incorporate cohort effect. Currie et al. (2006) introduce a simpler age-period-cohort model, which can be seen as a special case of the Renshaw and Haberman (2006) model.

Cairns et al. (2006) contribute the main mortality modelling alternative to the LC model. Their framework is also extended; for instance, Cairns et al. (2009) propose new CBD-type models that include cohort-effect factors and other features. Plat (2009) combines the CBD model with cohort effect with some features of the LC model. Inspired by Plat (2009), Dowd et al. (2020) extend the CBD-type models of Cairns et al. (2009) by including a static age function α_x; these are called CBD-X models.

All the mortality models above are constructed from a mixture of independent deterministic and stochastic components which capture age, period and, in some cases, cohort effects. The number and form of these effects usually distinguish one model from another. We focus on the CBD-X models in this report and ignore the cohort-effect factors for ease of presentation.

Alongside the development of stochastic mortality models, various estimation methods for these models have also been proposed. Early work by Lee and Carter (1992) and Koissi et al. (2006) estimates stochastic mortality models based on a singular value decomposition (SVD) approach—a method that decomposes a matrix into three components by factorization.

Specifically, the authors above set α_x as the arithmetic average of the logarithm of the central death rate (i.e., $\log(m_{x,t})$) over time; then, they apply SVD on $(\log(m_{x,t}) - \alpha_x)$ to extract the other parameters. For mortality forecasting purposes, period-effect factor time-series dynamics are assumed, and the parameters of this time-series model are estimated from the parameters obtained in the previous step.

Nowadays, most mortality studies rely on maximum likelihood estimation method (e.g., Brouhns et al., 2002; Cairns et al., 2009; Plat, 2009, and Dowd et al., 2020). This is usually done in two stages: first, by postulating a Poisson distribution for the number of deaths to facilitate maximum likelihood estimation; then, same as in the SVD method, some time-series dynamics for the period-effect factors are assumed. However, these two-stage estimation methods do not capture the dynamic nature of the period-effect factors and their underlying parameters consistently.

In contrast, a one-step approach incorporates the dynamic nature of the period-effect factors within the model estimation and thus can solve the inconsistency problem. The one-step approach can be achieved by at least two different methods: Markov chain Monte Carlo (MCMC) or state-space representation. Czado et al. (2005) use MCMC methods to estimate parameters within the LC model and test it on French male data. Cairns et al. (2011) propose a two-population Bayesian stochastic mortality model, which is designed for modelling sub-population (e.g., UK male insured lives) within a dominant population (e.g., the populations of England and Wales males) based on the MCMC approach. Cairns et al. (2019) apply the MCMC method to estimate parameters within a new multi-population Bayesian CBD model, which is used to capture the socio-economic difference in the mortality of Danish males.

Pedroza (2006) proposes an alternative solution; that is, recasting the LC model in a state-space formulation. State-space models (SSM) describe the probabilistic dependence between the observed measurement (i.e., the observed death rates) from a system and the latent variables (i.e., period-effect factors) that drive the dynamics of that system. By recasting the mortality model in a state-space representation, we can incorporate the dynamics of period-effect factors within the model estimation. Reichmuth and Sarferaz (2008) further extend such frameworks to the Renshaw and Haberman (2006) model. All the authors mentioned above use the Kalman filter method to estimate the distributions of the latent variables, but this filter requires the SSM to be linear and Gaussian. Fung et al. (2017) introduce particle filters for sampling non-linear and non-Gaussian SSM and apply this method to estimate different LC-type mortality models. The distributions of the latent variables are estimated by particle filters, while the unknown parameters within the SSM can be estimated either by MLE or MCMC methods.

We follow the study of Fung et al. (2017), and we sample the distribution of the latent variables using particle filters and estimating the other unknown parameters in the SSM using a sample-based approach—MCMC methods. This estimation procedure is also known

as the particle Markov chain Monte Carlo (pMCMC) method. As stated in Andrieu et al. (2010), the pMCMC method is more computationally efficient than classic MCMC methods. Therefore, in this report, we present a new Bayesian-based estimation procedure for CBD-X models that relies on the pMCMC method of Andrieu et al. (2010).

The contributions of this report lie in explaining the technicalities of pMCMC sampling for single-population CBD-X mortality models under the Bayesian paradigm and applying these models to Canadian mortality data. Although Bayesian estimation has been widely used in recent years (see, e.g., Cairns et al., 2019, apply MCMC methods to estimate the CBD-X model for the Danish population, and Fung et al., 2017, use the pMCMC method for the LC models), we are the first to apply the pMCMC method to estimate the CBD-X models. The CBD-X models in this report are estimated with Canadian male data to assess the quality of mortality estimations and predictions. Statistical diagnostic tests and model selection are also conducted. We further compare the model forecasting performance with the forecasts obtained with the maximum likelihood method.

This report consists of seven chapters. In Chapter 2, we review some important mortality models and the standard approach for specifying the structure of the period effects in the actuarial literature. In Chapter 3, we introduce a state-space representation of the preferred mortality model and explain filtering. The state-space formulation of the CBD-X models and the sequential Monte Carlo algorithm are also given in Chapter 3. Chapter 4 focuses on Bayesian inference and the methodology applied for model estimation; that is, the Metropolis-Hastings algorithm, the Gibbs sampler, and the pMCMC algorithm. In Chapter 5, we estimate the CBD-X models introduced in Chapter 2 using Canadian male mortality data. Chapter 6 compares model predictions estimated under both frequentist and Bayesian approaches. Chapter 7 concludes.

Chapter 2

One-Population Mortality Models

Sound mortality models are essential for pension plans and life insurers to correctly price annuities, value pension liabilities, and hedge against losses due to mortality and longevity risk. Several deterministic mortality laws were established during the 19^{th} and 20^{th} centuries, (e.g., Gompertz (1825) law for the force of mortality). Deterministic models are adequate for forecasting expected present values but do not capture the uncertainty about future mortality developments. Thus, a stochastic mortality model is necessary and, since the early 1990s, multiple stochastic mortality models were developed to provide accurate predictions of mortality improvements.

The force of mortality $\mu_{x,t}$ is commonly used to estimate and forecast mortality, where $x \geq 0$ represents ages and $t \geq 0$ represents calendar years. For fixed age x and time t, $\mu_{x,t}$ expresses the instantaneous probability of immediate death. It cannot be directly observed, unfortunately, as mortality data are not recorded in continuous time.

Typically, the central death rate $m_{x,t}$ is used as a proxy for $\mu_{x,t}$. We define the central death rate $m_{x,t}$ for person aged x during year t for $x \in [x_{\min}, x_{\max}]$ and $t \in [t_{\min}, t_{\max}]$ by

$$m_{x,t} = \frac{D_{x,t}}{E_{x,t}},$$

where $x_{\min}$ is the youngest age fitted within a mortality model and $x_{\max}$ is the oldest age. Likewise, $t_{\min}$ is the first calendar year fitted within a mortality model and $t_{\max}$ is the last calendar year. For convenience, we set $t_{\min} = 1$ and $t_{\max} = T$. Variable $D_{x,t}$ represents the death count during calendar year t at age x, and $E_{x,t}$ is the exposure of the population at age x during year t. That is to say, $m_{x,t}$ is the proportion of the number of deaths over the total exposure at age x during year t.

To obtain the force of mortality from discrete observations, a constant force of mortality assumption is commonly imposed. Under this assumption, $\mu_{x,t} = m_{x,t}$. Also, $\mu_{x,t} = -\log(1 - q_{x,t})$, where $q_{x,t}$ is the mortality rate for an individual aged x during year t, so we have the

following relationship:

$$q_{x,t} = 1 - \exp(-\mu_{x,t}) = 1 - \exp(-m_{x,t}).$$

Age effects represent mortality variation by age, regardless of birth cohort. Period effects capture mortality changes over time that equally affect all ages during a particular calendar year. Cohort effects represent mortality variations resulting from different generations represented by the year of birth. Throughout this report, we use $c = t - x$ to represent the year of birth or cohort year. According to Hunt and Blake (2021), most of the existing stochastic mortality models in the literature can be written as the following age-period-cohort effect model,

$$\eta_{x,t} = \alpha_x + \sum_{i=1}^{N} \beta_x^{(i)} \kappa_t^{(i)} \gamma_c^{(i)}, \tag{2.1}$$

where

- $\eta_{x,t}$ can be the death rate on logarithmic scale (i.e., $\eta_{x,t} = \log(m_{x,t})$) or the mortality rate in logit scale (i.e., $\eta_{x,t} = \log\left(\frac{q_{x,t}}{1-q_{x,t}}\right) = \text{logit}(q_{x,t})$) for people aged x in year t;
- α_x is a static age function, which is the basic age effect and can be treated as the mortality table without mortality improvement;
- N is the number of period and/or cohort terms within the model;
- $\beta_x^{(i)}$ is the i^{th} age-effect factor, which variations are associated with physiological and social ageing processes;
- $\kappa_t^{(i)}$ is the i^{th} period-effect factor that affects mortality trend for people of all ages that are alive in period t; for example, this can be related to environmental conditions, medical developments, and living conditions;
- $\gamma_c^{(i)}$ represents the i^{th} cohort effect; that is, the impact of the past conditions on current mortality rate. For instance, generation-specific habits, wars, and catastrophes, for individuals born in the same year.

The two major mortality model families used in actuarial science are nested within this general model. They are the LC-type models and the CBD-type models.

2.1 Generalized Lee and Carter-Type Models

Lee and Carter (1992) present an original model that fits and predicts mortality rates for the United States. It is a two-factor model that has one period-effect factor and two age-effect factors. Due to its simplicity and relatively good performance, this type of model has been

widely used for various countries' demographic and actuarial applications (e.g., Lundström and Qvist, 2004; Haqqi Anna Zili et al., 2018; Kamaruddin and Noriszura, 2018).

2.1.1 The Original LC Model

Lee and Carter (1992) propose the following two-factor model for death rates:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} \ ,$$

where $\eta_{x,t} = \log(m_{x,t})$ and parameters α_x describe the basic pattern of $\log(m_{x,t})$ averaged over time for each age. Parameters $\kappa_t^{(1)}$ express the overall evolution of mortality over time; their related impacts on the basic mortality patterns are weighted for the the different ages through $\beta_x^{(1)}$. This model can be written as a nested case of Equation (2.1) by setting $N = 1$ and $\gamma_c^{(1)} = 1$ for all c.

2.1.2 The Renshaw and Haberman Model

Renshaw and Haberman (2006) conclude that the original LC model does not fit mortality data from England and Wales well; however, they find that adding a cohort-effect term can increase the goodness-of-fit significantly. The resulting model is as follows:

$$\eta_{x,t} = \alpha_x + \beta_x^{(1)}\kappa_t^{(1)} + \beta_x^{(2)}\gamma_c^{(2)} \ ,$$

where $\eta_{x,t} = \log(m_{x,t})$, parameters α_x, $\beta_x^{(1)}$, and $\kappa_t^{(1)}$ have the same meaning as those in the original LC model, and parameters $\beta_x^{(2)}$ adjust the cohort effect across age groups. This model can be written as a nested case of Equation (2.1) by setting $N = 2$, $\kappa_t^{(2)} = 1$ for all t, and $\gamma_c^{(1)} = 1$ for all c.

2.2 Generalized CBD-Type Models

Cairns et al. (2006) develop another stochastic mortality model originally used for forecasting United Kingdom mortality data. This model plays an important role in predicting mortality at older ages (i.e., people older than 60). This model relies on two period-effect factors. The first factor affects mortality equally for all ages, whereas the second factor has a higher impact on older ages. This type of model has been accepted widely and applied in pension funds, life insurance, and pricing longevity bonds (e.g., Cairns et al., 2009; Cairns, 2011; Chan et al., 2014; Cairns et al., 2019).

2.2.1 The Original CBD Model

Cairns et al. (2006) propose the original CBD model as follows:

$$\eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x}),$$

where $\overline{x} = \frac{\sum_{x=x_{\min}}^{x=x_{\max}} x}{x_{\max}-x_{\min}+1}$ is the average of fitted ages. This model can be written as a nested version of Equation (2.1) by setting $N = 2$, $\alpha_x = 0$ for all x, $\beta_x^{(1)} = 1$ for all x, $\beta_x^{(2)} = (x - \overline{x})$, and $\gamma_c^{(1)} = \gamma_c^{(2)} = 1$ for all c. The first period-effect factor $\kappa_t^{(1)}$ represents the overall mortality improvement over time, and the second $\kappa_t^{(2)}$ captures the different mortality improvement by ages; specifically, people aged below the mean age $\overline{x}$ have larger mortality improvements than people aged above the mean age $\overline{x}$. They use two period-effect factors because the quantification of mortality improvement through only one period-effect factor is not adequate. The same authors point out that having only one period-effect factor $\kappa_t^{(1)}$ in the LC model implies perfect correlation among the changes of central death rates for all ages (Cairns et al., 2009). The original CBD paper sets $\eta_{x,t} = \text{logit}(q_{x,t})$ and uses the England and Wales data. For those data, the $\text{logit}(q_{x,t})$ increases approximately linearly in age x, but this is not necessarily true for all data sets. Thus, it is not necessary to set $\eta_{x,t} = \text{logit}(q_{x,t})$ for all CBD-type mortality models.

2.2.2 The CBD Model with a Cohort Effect

Since a cohort effect might exist in some populations and can become an important factor in mortality modelling, Cairns et al. (2009) extend the original CBD model to include a cohort effect, as follows:

$$\eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x}) + \gamma_c^{(3)},$$

where $\eta_{x,t} = \text{logit}(q_{x,t})$. This model is not as widely used as the three-factor CBD model discussed below because the three-factor CBD model performs better, as shown in Cairns et al. (2009). The CBD model with a cohort effect is nested in the general framework of Equation (2.1) by setting $N = 3$, $\alpha_x = 0$ for all x, $\beta_x^{(1)} = \beta_x^{(3)} = 1$ for all x, $\beta_x^{(2)} = (x - \overline{x})$, $\kappa_t^{(3)} = 1$ for all t, and $\gamma_c^{(1)} = \gamma_c^{(2)} = 1$ for all c.

2.2.3 A Three-Factor CBD Model

Cairns et al. (2009) extend the CBD model with a cohort effect above by adding a quadratic age effect, which is inspired by the curvature that appears in the $\text{logit}(q_{x,t})$ plot when fitting US mortality data. Figure 2.1, which is taken from the working paper version of Cairns et al. (2009), shows $\text{logit}(q_{x,t})$ for the year 2003 and ages $x \in [60, 89]$ when fitting US mortality data. Cairns et al. (2009) point out that the curvature is not all that prominent, but it impacts the results when the different models are compared. This model is called M7 in Cairns et al. (2009), but we call it the three-factor CBD model because it has three period-effect factors. The formula of this generalization is as follows:

$$\eta_{x,t} = \kappa_t^{(1)} + \kappa_t^{(2)}(x - \overline{x}) + \kappa_t^{(3)}((x - \overline{x})^2 - \sigma_x^2) + \gamma_c^{(4)},$$

where $\eta_{x,t} = \text{logit}(q_{x,t})$ and $\sigma_x^2 = \frac{\sum_{x=x_{\min}}^{x=x_{\max}} (x-\overline{x})^2}{x_{\max}-x_{\min}+1}$. This model can be written as a nested case of Equation (2.1) by setting $N = 4$, $\alpha_x = 0$ for all x, $\beta_x^{(1)} = \beta_x^{(4)} = 1$ for all x, $\beta_x^{(2)} = (x-\overline{x})$, $\beta_x^{(3)} = ((x-\overline{x})^2 - \sigma_x^2)$, $\kappa_t^{(4)} = 1$ for all t, and $\gamma_c^{(1)} = \gamma_c^{(2)} = \gamma_c^{(3)} = 1$ for all c. Parameters $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ have the same meaning as their counterparts in the original CBD model, and $\kappa_t^{(3)}$ captures non-linearity with respect to age in year t.

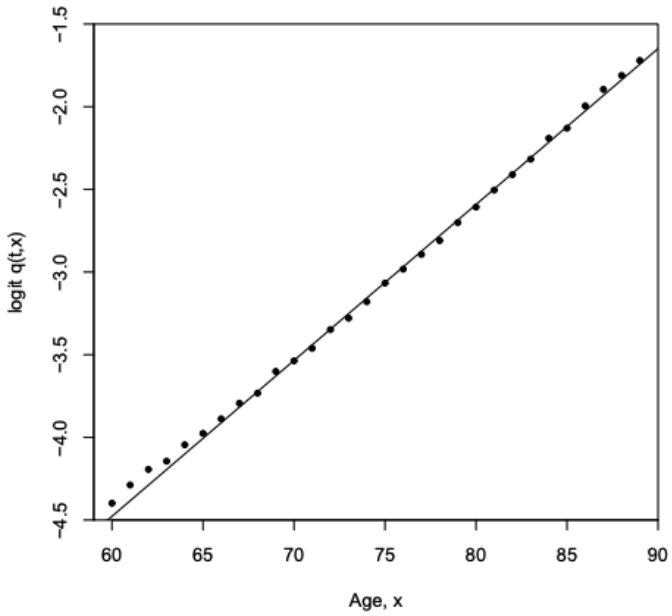

Figure 2.1: US mortality rates for the year 2003. The dots are $\text{logit}(q_{x,t})$ calculated from the observed data and the solid line is the model-based estimated mortality rate in logit scale, $\text{logit}(\hat{q}_{x,t})$.

2.3 The CBD-X Models

Several extensions of the CBD-type model are introduced by Dowd et al. (2020), which originate from Plat (2009). These models combine elements of LC-type and CBD-type models. The resulting models can be used for full age ranges while capturing the cohort effect.

2.3.1 The Plat Model

Plat (2009) assumes a static age function α_x, three period-effect terms, and cohort effect. The Plat model is thus given by:

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\overline{x}) + \kappa_t^{(3)}(x-\overline{x})^+ + \gamma_c^{(4)},$$

where $\eta_{x,t} = \log(m_{x,t})$ and $(x-\overline{x})^+ = \max(0, x-\overline{x})$. This model can be written as a nested case of the Equation (2.1) by setting $N = 4$, $\beta_x^{(1)} = \beta_x^{(4)} = \kappa_t^{(4)} = \gamma_c^{(1)} = \gamma_c^{(2)} = \gamma_c^{(3)} = 1$ for all x, t, or c, $\beta_x^{(2)} = (x-\overline{x})$, and $\beta_x^{(3)} = (x-\overline{x})^+$.

Plat (2009) suggests removing the third period-effect term if the user is only interested in the older ages. The two-factor reduced Plat model then becomes:

$$\eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\overline{x}) + \gamma_c^{(3)}.$$

This model is virtually the same as the CBD-X(2) model that we will discuss in the following section.

2.3.2 The CBD-X Models

Inspired by Plat (2009), Dowd et al. (2020) investigate a new type of mortality model which combines the static age function α_x with the CBD-type models. Models of this type are called CBD-X[1] models. The CBD-X models are used to model mortality of groups of adults over a wider range of ages than usually advisable for the CBD-type models. The CBD-X models create a hybrid between the LC and the CBD models, just like the Plat model, as the LC model can be written in the CBD-X form. The CBD-X(1) model is the CBD-X model with one period effect, and the same naming logic applies to the other CBD-X models in this report. The CBDX model in Hunt and Blake (2020) is the CBD-X(2) model without cohort effect. Also, the CBD-X model in Cairns et al. (2019) can be treated as a multi-population version of the CBD-X(2) model without cohort effect.

The CBD-X models are given by the following:

$$\text{CBD-X(1)} \ : \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \gamma_c^{(2)}, \tag{2.2}$$

$$\text{CBD-X(2)} \ : \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\overline{x}) + \gamma_c^{(3)}, \tag{2.3}$$

$$\text{CBD-X(3)} \ : \ \eta_{x,t} = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x-\overline{x}) + \kappa_t^{(3)}((x-\overline{x})^2 - \sigma_x^2) + \gamma_c^{(4)}, \tag{2.4}$$

where $\eta_{x,t} = \log(m_{x,t})$. The CBD-X(1) model can be written as a nested case of Equation (2.1) by setting $N = 2$, and $\beta_x^{(1)} = \beta_x^{(2)} = \kappa_t^{(2)} = \gamma_c^{(1)} = 1$ for all x, t, or c. The CBD-X(1) model without cohort effect is essentially the special case of the LC model with $\beta_x^{(1)} = 1$. The CBD-X(2) model is an extension of the CBD model with a cohort effect, and it can be written as a nested case of Equation (2.1) by setting $N = 3$, $\beta_x^{(1)} = \beta_x^{(3)} = \kappa_t^{(3)} = \gamma_c^{(1)} = \gamma_c^{(2)} = 1$ for all x, t, or c, and $\beta_x^{(2)} = (x-\overline{x})$. The CBD-X(3) extends the three-factor CBD model, and it can be written as a nested case of Equation (2.1) by setting $N = 4$, $\beta_x^{(1)} = \beta_x^{(4)} = \kappa_t^{(4)} = \gamma_c^{(1)} = \gamma_c^{(2)} = \gamma_c^{(3)} = 1$ for all x, t, or c, $\beta_x^{(2)} = (x-\overline{x})$, and

[1]Please note that in Dowd et al. (2020), these models are called CBDX models, but we prefer to call them as CBD-X models.

$\beta_x^{(3)} = (x - \overline{x})^2 - \sigma_x^2$. Moreover, α_x, $\beta_x^{(i)}$, $\kappa_t^{(i)}$, and $\gamma_c^{(i)}$ have the same meaning as their counterparts in the CBD-type models.

We will focus on the CBD-X models in this report because they combine the best features of the LC-type and CBD-type models. Furthermore, we will ignore the cohort effect to ease the model estimation.

2.4 Structure of the Period Effects

The dynamics of mortality are driven by the period effects $\kappa_t^{(i)}$ for $i = 1, ..., N$. Therefore, we usually need to specify a structure for them by applying time-series techniques. The standard approach in the actuarial literature (e.g., Lee and Carter, 1992; Cairns et al., 2006, 2011; Georgios et al., 2017), is to assume that period effects follow a (multivariate) random walk with drift.

A random walk is a stochastic process that forces dependence from one time step to the next. The dependence provides some consistency from one step to the next. A random walk with drift is a special case such that, at each point in time, the series takes a random step away from its last position with a trend equal to the so-called drift term. A random walk with drift can be written as

$$\boldsymbol{\kappa}_t = \boldsymbol{\mu} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{Z}_t, \quad \boldsymbol{Z}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2.5}$$

where $\boldsymbol{\kappa}_t = \begin{bmatrix} \kappa_t^{(1)} & \dots & \kappa_t^{(N)} \end{bmatrix}^\top$, $\boldsymbol{\mu}$ is an N-dimensional vector of drift parameters, and $\boldsymbol{\Sigma}$ is the $N \times N$ covariance matrix of the multivariate white noise $\boldsymbol{Z}_t$. Moreover, $\mathbf{0}$ is an N-dimensional vector of zeros.

For the CBD-X(3) model, we can specify Equation (2.5) as follows:

$$\boldsymbol{\kappa}_t = \begin{bmatrix} \kappa_t^{(1)} & \kappa_t^{(2)} & \kappa_t^{(3)} \end{bmatrix}^\top, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}^\top, \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{12} & v_{22} & v_{23} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}.$$

The CBD-X(2) and CBD-X(1) models can be treated as special cases of the CBD-X(3) model by setting $\kappa_t^{(i)} = 0$ for $i = 3$ and $i = 2, 3$, respectively.

Chapter 3

Particle Filters for State-Space Models

3.1 State-Space Model

The estimation of the mortality models mainly adopts two-stage frequentist methods, which use a maximum likelihood-based approach first, and then estimate a model for the period effects for forecasting purposes. However, by using this two-stage method, the dynamics of the period effects are not directly incorporated in the first step. Fung et al. (2017) argue that recasting different classes of mortality models in a state-space formulation allows for state-space-based inference under either the frequentist or Bayesian paradigm. A key advantage of this approach is that we can reduce the two-stage estimation into a single step by including the dynamics of the period effects within the model estimation.

A state-space model (SSM) is a model that describes the probabilistic dependence between the observed measurement from a system and the latent variables that drive the dynamics of that system. Latent variables cannot be directly observed; we need to infer them through other directly observed variables. A general SSM consists of two equations: the transition equation and the measurement equation. The transition equation explains the relationship between latent variables at the current time and those at the previous time. In the SSM, we assume that latent variables satisfy the Markov property, so the current state of the latent variables only depend on previous states. Mathematically, we have that

$$\boldsymbol{\kappa}_t = f\left(\boldsymbol{\kappa}_{t-1}, \boldsymbol{Z}_t\right), \quad \text{for } t \in \{1, \ldots, T\}, \tag{3.1}$$

where $\boldsymbol{\kappa}_t \in \mathbb{R}^N$ is a vector of latent variables at time t, and the random vector $\boldsymbol{Z}_t \in \mathbb{R}^N$ contains the error terms used within the transition equation. The function $f : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ explains how the latent variables change from one step to the next. Because the latent variables are unobserved, we use a measurement equation to link them to the observations.

The measurement equation can be written as:

$$\boldsymbol{y}_t = g\left(\boldsymbol{\kappa}_t, \boldsymbol{\varepsilon}_t\right), \quad \text{for } t \in \{1, \ldots, T\}, \tag{3.2}$$

where $\boldsymbol{y}_t \in \mathbb{R}^n$ is the vector of observed data at time t, $\boldsymbol{\kappa}_t$ are the latent variables, and $\boldsymbol{\varepsilon}_t \in \mathbb{R}^n$ is the vector of error terms used in the measurement equation. The function $g : \mathbb{R}^N \times \mathbb{R}^n \to \mathbb{R}^n$ handles the relationship between the current state of the latent variables and the observations. If g and f are both linear functions, then the SSM is linear; otherwise, it is non-linear. If the error terms $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{Z}_t$ are modelled with Gaussian distributions, then we have a Gaussian SSM; otherwise, we have a non-Gaussian SSM.

The transition and measurement equations usually involve some unknown parameters $\boldsymbol{\Theta}$. In the model, both the latent variables $\boldsymbol{\kappa}_t$ and these unknown parameters $\boldsymbol{\Theta}$ need to be estimated. The difference between them is that latent variables vary from time to time while parameters are fixed. They should, therefore, be estimated differently.

3.1.1 State-Space Formulation of the CBD-X Models

The mortality models introduced in Chapter 2 are cast into a state-space representation. The latent variables are the period effects. The observed data are the logged estimated central death rates, $\hat{m}_{x,t}$.

Recall that the CBD-X(3) model can be written as in Equation (2.4):

$$\log(m_{x,t}) = \alpha_x + \kappa_t^{(1)} + \kappa_t^{(2)}(x - \bar{x}) + \kappa_t^{(3)}\left((x - \bar{x})^2 - \sigma_x^2\right).$$

Let the logged observed central death rate, $\log(\hat{m}_{x,t})$, be the true logged central death rate plus some noise:

$$\log(\hat{m}_{x,t}) = \log(m_{x,t}) + \epsilon_{x,t}.$$

Let $\boldsymbol{y}_t = [\log(\hat{m}_{x_{\min},t}) \quad \ldots \quad \log(\hat{m}_{x_{\max},t})]^\top$ be the set of $\log(\hat{m}_{x,t})$ for all the age x in year t. The latent variables are the period effects $\boldsymbol{\kappa}_t = \begin{bmatrix}\kappa_t^{(1)} & \kappa_t^{(2)} & \kappa_t^{(3)}\end{bmatrix}^\top$. The error terms within the measurement equation are $\boldsymbol{\varepsilon}_t = [\varepsilon_{x_{\min},t} \quad \ldots \quad \varepsilon_{x_{\max},t}]^\top$. Thus, by using matrix notation, we obtain

$$\boldsymbol{y}_t = \boldsymbol{\alpha}_x + \boldsymbol{\beta}_x \boldsymbol{\kappa}_t + \boldsymbol{\varepsilon}_t,$$

where

$$\boldsymbol{\alpha}_x = [\alpha_{x_{\min}} \quad \ldots \quad \alpha_{x_{\max}}]^\top,$$
$$\boldsymbol{\beta}_x = \begin{bmatrix} 1 & (x_{\min} - \bar{x}) & (x_{\min} - \bar{x})^2 - \sigma_x^2 \\ \vdots & \vdots & \vdots \\ 1 & (x_{\max} - \bar{x}) & (x_{\max} - \bar{x})^2 - \sigma_x^2 \end{bmatrix}.$$

The transition equation is

$$\boldsymbol{\kappa}_t = \boldsymbol{\mu} + \boldsymbol{\kappa}_{t-1} + \boldsymbol{Z}_t, \quad \boldsymbol{Z}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\kappa}_t = \begin{bmatrix} \kappa_t^{(1)} & \kappa_t^{(2)} & \kappa_t^{(3)} \end{bmatrix}^\top, \quad \boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \mu_3 \end{bmatrix}^\top, \quad \boldsymbol{\Sigma} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{12} & v_{22} & v_{23} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}.$$

A key component of the state-space representation is the distribution of the measurement errors. We assume that $\log(\hat{m}_{x,t}) \sim \mathcal{N}(\log(m_{x,t}), \frac{1}{\hat{D}_{x,t}})$ yields $\epsilon_{x,t} \sim \mathcal{N}(0, \frac{1}{\hat{D}_{x,t}})$, where $\hat{D}_{x,t}$ is the observed number of deaths at age x in year t. See Appendix A for the justification of this assumption.

By writing the CBD-X models in the state-space form, the dynamics of the observations $\boldsymbol{y}_t$ and the dynamics of the period effect $\boldsymbol{\kappa}_t$ are combined into one system. Furthermore, since logged observed central death rates follow normal distribution, the measurement density for single population mortality models listed in Chapter 2 can be written as:

$$\begin{aligned} p(\boldsymbol{y}_{1:T} \,|\, \boldsymbol{\kappa}_{0:T}, \boldsymbol{\Theta}) &\propto \prod_{x,t} \sqrt{\frac{\hat{D}_{x,t}}{2\pi}} \exp\left(-\frac{\hat{D}_{x,t}}{2} \left(\log(\hat{m}_{x,t}) - \log(m_{x,t})\right)^2 \right) \\ &\propto \exp\left(-\sum_{x,t} \frac{\hat{D}_{x,t}}{2} \left(\log(\hat{m}_{x,t}) - \log(m_{x,t})\right)^2 \right). \end{aligned} \tag{3.3}$$

3.2 Filtering

The uncertainty about the latent variables is expressed via a joint conditional probability distribution $p\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$, where $\boldsymbol{\kappa}_{0:t} = \{\boldsymbol{\kappa}_0, ..., \boldsymbol{\kappa}_t\}$, $\boldsymbol{y}_{1:t} = \{\boldsymbol{y}_1, ..., \boldsymbol{y}_t\}$, and $t \in \{1, ..., T\}$. We can find the joint conditional probability distribution $p\left(\boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}\right)$ by applying filtering. The unknown parameter $\boldsymbol{\Theta}$ is assumed to be known in this section as running filtering algorithms require fixed and known values for $\boldsymbol{\Theta}$. The estimation of $\boldsymbol{\Theta}$ will be discussed in the next chapter.

Filtering allows us to estimate the distribution of the current latent state of an SSM given the observations up to and including the current time, $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$. We find the joint conditional probability distribution by calculating the one-step ahead predictive distribution $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)$ and the filtering distribution $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$ recursively. The recursive approach begins by setting an initial distribution for the latent factors, $p\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right)$. Then, each iteration can be divided into two steps. The first step is called the prediction step: at each time t, we obtain the one-step ahead predictive distribution of the latent variable $\boldsymbol{\kappa}_t$ based on the information from the last period's filtering distribution $p\left(\boldsymbol{\kappa}_{t-1} \,|\, \boldsymbol{y}_{1,...t-1}, \boldsymbol{\Theta}\right)$

and the transition density:

$$\underbrace{p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1},\, \boldsymbol{\Theta}\right)}_{\text{Predictive Distribution}} = \int \underbrace{p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{\kappa}_{t-1},\, \boldsymbol{\Theta}\right)}_{\text{Transition Density}} \underbrace{p\left(\boldsymbol{\kappa}_{t-1} \,|\, \boldsymbol{y}_{1:t-1},\, \boldsymbol{\Theta}\right)}_{\text{Filtering Distribution}} d\boldsymbol{\kappa}_{t-1},$$

where the transition density $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{\kappa}_{t-1}, \boldsymbol{\Theta}\right)$ is implied by the transition equation as $\boldsymbol{\kappa}_t = f\left(\boldsymbol{\kappa}_{t-1}, \boldsymbol{Z}_t\right)$.

Second, in the update step, we apply Bayes' rule to update the filtering distribution by combining the predicted $\boldsymbol{\kappa}_t$ with the additional time-t observations $\boldsymbol{y}_t$:

$$\begin{aligned} p\left(\boldsymbol{\kappa}_t | \, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) &= \frac{p\left(\boldsymbol{y}_t, \boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)} \\ &= \frac{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t,\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)}{\int p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right) d\boldsymbol{\kappa}_t}. \end{aligned}$$

Given $\boldsymbol{\kappa}_t$, the observations $\boldsymbol{y}_t$ are statistically independent with the previous observations $\boldsymbol{y}_{1:t-1}$, thus the filtering distribution becomes:

$$p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) = \frac{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)}{\int p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right) d\boldsymbol{\kappa}_t}.$$

If the dynamics of the latent variables are linear and Gaussian, we can derive analytical solutions for this filtering distribution by using the Kalman filter. However, as most of the SSMs are non-linear and non-Gaussian, it is not possible to derive analytical solutions, and we therefore need to approximate the joint conditional probability distribution numerically. The most efficient and popular method to do so is the sequential Monte Carlo algorithm (SMC; also called particle filter or PF), which uses simulation to approximate the target distribution.

Although the dynamics of the period effects in our models are linear Gaussian, we still use the sequential Monte Carlo algorithm to allow for possible non-linear period-effect dynamics in the future.

3.2.1 Sequential Monte Carlo Algorithm

The SMC is less restrictive than the Kalman filter in the SSM context as it does not require assumptions about the linearity or normality of the state distributions. The aim of applying the SMC methods for the state-space model is to approximate the joint conditional distribution $p\left(\boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta}\right)$ sequentially. More precisely, SMC aims to approximate $p\left(\boldsymbol{\kappa}_{0:1} \,|\, \boldsymbol{y}_1, \boldsymbol{\Theta}\right)$ and $p\left(\boldsymbol{y}_1 \,|\, \boldsymbol{\Theta}\right)$ first, then $p\left(\boldsymbol{\kappa}_{0:2} \,|\, \boldsymbol{y}_{1:2}, \boldsymbol{\Theta}\right)$ and $p\left(\boldsymbol{y}_{1:2} \,|\, \boldsymbol{\Theta}\right)$, and so on. Instead of deriving an analytical equation as the Kalman filter does, the PF approximates the joint conditional distribution by using a sequence of weighted random samples of $p\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$ called particles.

If we sample J independent random variables, $\boldsymbol{\kappa}_{0:t} \sim p\left(\boldsymbol{\kappa}_{0:t} \,|\boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$, then standard Monte Carlo methods approximate $p\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$ by generating an empirical distribution made up of J samples of $\boldsymbol{\kappa}_{0:t}^{j}$:

$$\hat{p}\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) = \frac{1}{J} \sum_{j=1}^{J} \delta_{\boldsymbol{\kappa}_{0:t}^{j}}(\boldsymbol{\kappa}_{0:t}),$$

where $\delta_x(x)$ is a Dirac mass function centred at x. However, if $p\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$ is a complex high-dimensional probability distribution, then we cannot sample $\boldsymbol{\kappa}_{0:t}$ directly from it. The SMC algorithm addresses this problem by sampling $\boldsymbol{\kappa}_{0:t}$ from the proposal distribution $q\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)$. The proposal distribution can be any distribution from which it is easy to generate a sample. The approximation of the conditional probability becomes:

$$\hat{p}\left(\boldsymbol{\kappa}_{0:t} \,|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) \propto \sum_{j=1}^{J} W_t^j \delta_{\boldsymbol{\kappa}_{0:t}^{j}}(\boldsymbol{\kappa}_{0:t}),$$

where W_t^j is the normalized importance weight associated with particle $\boldsymbol{\kappa}_{0:t}^{j}$ to correct for the fact that $\boldsymbol{\kappa}_{0:t}$ are not sampled from the right distribution. We define the importance weight w_t^j as

$$w_t^j = \frac{p\left(\boldsymbol{\kappa}_{0:t}^{j} \,\middle|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_{0:t}^{j} \,\middle|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)},$$

and the normalized importance weight is given by:

$$W_t^j = \frac{w_t^j}{\sum_{k=1}^{J} w_t^k}. \tag{3.4}$$

To avoid having to recompute the entire expression for the importance weights at each iteration, and to increase computational efficiency, instead of sampling all the particles $\boldsymbol{\kappa}_{0:t}^{j}$ from a joint proposal distribution at once, we sample particles from a sequence of conditional distributions. We rewrite the proposal distribution in recursive form:

$$q\left(\boldsymbol{\kappa}_{0:t}^{j} \,\middle|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) \equiv q\left(\boldsymbol{\kappa}_{t}^{j} \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^{j}, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) q\left(\boldsymbol{\kappa}_{0:t-1}^{j} \,\middle|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right).$$

To obtain particles $\boldsymbol{\kappa}_{0:t}^{j}$ at time t, we sample $\boldsymbol{\kappa}_1$ from a proposal distribution of $\boldsymbol{\kappa}_1$, $q\left(\boldsymbol{\kappa}_1 \,|\, \boldsymbol{\kappa}_0, \boldsymbol{y}_1, \boldsymbol{\Theta}\right)$ at time 1, then append to these the new sample $\boldsymbol{\kappa}_2$ from a proposal distribution of $\boldsymbol{\kappa}_2$, $q\left(\boldsymbol{\kappa}_2 \,|\, \boldsymbol{\kappa}_{0:1}, \boldsymbol{y}_{1:2}, \boldsymbol{\Theta}\right)$, and repeat this procedure until time t. The joint conditional distribution can be written as

$$p\left(\boldsymbol{\kappa}_{0:t}^{j} \,\middle|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) = \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_{0:t}^{j}, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\kappa}_{0:t}^{j} \,\middle|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}$$

$$= \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_{0:t}^j,\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_{0:t-1}^j \,\middle|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}.$$

In the SSM, latent variables satisfy the Markov property and, given $\boldsymbol{\kappa}_t$, the observations $\boldsymbol{y}_t$ are independent of $\boldsymbol{y}_{1:t-1}$, leading to

$$p\left(\boldsymbol{\kappa}_{0:t}^j \,\middle|\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right) = \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_{0:t-1}^j \,\middle|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}.$$

Then, the importance weight becomes:

$$\begin{aligned} w_t^j &= \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_{0:t-1}^j \,\middle|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}; \boldsymbol{\Theta}\right)\, q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)\, q\left(\boldsymbol{\kappa}_{0:t-1}^j \,\middle|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)} \\ &\propto \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)}\, w_{t-1}^j. \end{aligned}$$

We thus define

$$\tilde{w}_t^j \equiv \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)},$$

which is known as the incremental importance weight. The denominator of the incremental importance weight is typically reduced to $q(\boldsymbol{\kappa}_t^j \,|\, \boldsymbol{\kappa}_{t-1}^j,\, \boldsymbol{y}_t, \boldsymbol{\Theta})$ for computational convenience, so

$$\tilde{w}_t^j = \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j,\, \boldsymbol{y}_t, \boldsymbol{\Theta}\right)}.$$

As a result, at each iteration, only the incremental importance weight needs to be computed, and a new sequence of particles is obtained by keeping the trajectories of the particles sampled up to time $t-1$.

The SMC produces its approximation by an iterative process. The algorithm presented below is based on Andrieu et al. (2010). At first, we need to generate initial values for our particles $\boldsymbol{\kappa}_0$. Then, for each time t, we start by sampling, J random samples of $\boldsymbol{\kappa}_t$, denoted by $\boldsymbol{\kappa}_t^j$, from a proposal distribution $q\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_t, \boldsymbol{\kappa}_{t-1}, \boldsymbol{\Theta}\right)$ to approximate $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_t, \boldsymbol{\kappa}_{t-1}, \boldsymbol{\Theta}\right)$. Then, the particles are weighted and the corresponding normalized importance weights are calculated; the latter is needed as we sample particles from the proposal distribution and not the target distribution.

The importance weight of one particle might grow exponentially over time and, as the number of iterations increases, all the probability mass will eventually be allocated to that particle; that is, one particle could end up with normalized importance weight close

to one, while the other particles' normalized importance weights would be close to zero. This is known as weight degeneracy. Adding a resampling step can prevent the weight degeneracy problem. Thus, the last step of the SMC algorithm is resampling. Resampling means a new sequence of particles is replicated from the existing particles based on their normalized importance weights. The simplest resampling method (and the one used in this study) is called multinomial resampling. Specifically, in this study, we draw J random variables with replacement from a multinomial distribution with probabilities $\boldsymbol{W}_t$, where $\boldsymbol{W}_t = [W_t^1 \quad ... \quad W_t^J]$, and W_t^J is obtained by Equation (3.4). As a consequence of this resampling, the particles with small weights will be eliminated while those with large weights will be duplicated. The resampled particle's weights are set equal to $w_t^j = \frac{1}{N}$ for $t \in \{1, \ldots, T\}$, which forces the weights not to permanently degenerate. If resampling is done at the end of every step, then the importance weight becomes

$$\begin{aligned} w_t^j &\propto \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^j, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{0:t-1}^j, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)} \\ &\propto \tilde{w}_t^j. \end{aligned}$$

Andrieu et al. (2010) introduce an ancestor variable, A_t^j, to keep track of the particles. Instead of resampling $\boldsymbol{\kappa}_t$, the resampling step is done by resampling A_t^j from a multinomial distribution. Hence, in the sampling step, we sample $\boldsymbol{\kappa}_t^j \sim q\left(\boldsymbol{\kappa}_t \,\middle|\, \boldsymbol{y}_t; \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)$ and set $\boldsymbol{\kappa}_{0:t}^j = \left[\boldsymbol{\kappa}_{0:t-1}^{A_{t-1}^j}, \boldsymbol{\kappa}_t^j\right]$. The ancestor particles will be useful in the next chapter. Then, we repeat this process for each time t until time T. The pseudocode of the SMC algorithm is summarized in Algorithm 1.

3.2.2 Bootstrap Filter

If we use the transition equation as the proposal distribution, then the filter is called the bootstrap filter. In the bootstrap filter, because $q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{y}_t; \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right) = p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)$, the importance weights are simply given by $w_t^j = p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)$, and the bootstrap filter does not use the information in the current observation $\boldsymbol{y}_t$ to generate a new particle, because the importance density is unconditional to $\boldsymbol{y}_t$. We apply the bootstrap filter in our model estimation.

3.2.3 Likelihood Evaluation

In addition to a sequence of particles, we can also obtain an evaluation of the likelihood function by:

$$\mathcal{L}\left(\boldsymbol{y}_{1:T} \,|\, \boldsymbol{\Theta}\right) = p\left(\boldsymbol{y}_1 \,|\, \boldsymbol{\Theta}\right) \prod_{t=2}^{T} p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right), \tag{3.5}$$

Algorithm 1 Sequential Monte Carlo

1: sample $\boldsymbol{\kappa}_0^j \sim q\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right)$ and set $A_0^j = j$
2: **for** $t = 1, ..., T$ **do**
3: sample $\boldsymbol{\kappa}_t^j \sim q\left(\boldsymbol{\kappa}_t \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)$ and set $\boldsymbol{\kappa}_{0:t}^j = \left[\boldsymbol{\kappa}_{0:t-1}^{A_{t-1}^j}, \boldsymbol{\kappa}_t^j\right]$
4: compute the weight:

$$
\begin{aligned}
w_t^j &= \frac{p\left(\boldsymbol{\kappa}_{0:t}^j,\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\kappa}\right)\, q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)} \\
&= \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)}
\end{aligned}
$$

5: normalize the weight: $W_t^j = \frac{w_t^j}{\sum_{k=1}^J w_t^k}$
6: sample A_t^j from a multinomial distribution with support 1 to J and weights $\boldsymbol{W}_t$
7: **end for**

where

$$
\begin{aligned}
p\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right) &= \int p\left(\boldsymbol{\kappa}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{\kappa}_{t-1}, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t, \boldsymbol{\Theta}\right) d\boldsymbol{\kappa}_{t-1:t}, \\
&= \int w_t\, p\left(\boldsymbol{\kappa}_{t-1} \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right)\, q\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{y}_t,\, \boldsymbol{\kappa}_{t-1},\, \boldsymbol{\Theta}\right) d\boldsymbol{\kappa}_{t-1:t}.
\end{aligned}
$$

The estimated marginal likelihood within the SMC algorithm is given by

$$
\hat{p}\left(\boldsymbol{y}_{1:T} \,|\, \boldsymbol{\Theta}\right) = \hat{p}\left(\boldsymbol{y}_1 \,|\, \boldsymbol{\Theta}\right) \prod_{t=2}^{T} \hat{p}\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right),
$$

where

$$
\hat{p}\left(\boldsymbol{y}_t \,|\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\Theta}\right) \approx \frac{1}{J} \sum_{j=1}^{J} w_t^j.
$$

The likelihood can be used to estimate the unknown parameters by either maximizing the function (i.e., frequentist-based methods) or constructing a posterior distribution of the parameters (i.e., Bayesian-based methods). We will discuss the details of the latter estimation paradigms in the next chapter.

Chapter 4

Bayesian Inference

The values of the unknown parameters within the transition and measurement equations need to be estimated. The estimation can be done by frequentist or Bayesian methods. Because Bayesian and frequentist inference differ in their basic philosophies, the core features of both paradigms are reviewed in the next section.

4.1 Comparison of Frequentist and Bayesian Paradigms

In frequentist estimation, any unknown model parameter is generally assumed constant. The rationale is that even if a parameter cannot be observed, there exists one true value, and randomness stems from natural deviations of anything unknown when experiments are repeated. The fundamental measure of such uncertainty is captured by probability, which is the limit of the relative frequency of an event in a very long, theoretically infinite, sequence of the same experiment conducted independently of each other. The results of a frequentist approach can be represented by a confidence interval or a hypothesis test. Confidence intervals use data from a sample to estimate a population parameter. For example, a $100p\%$ confidence interval includes the true but unknown value with confidence $p \in (0, 1)$. However, it is wrong to state that the unknown parameter lies with this interval with probability p. Similarly, given a statistic to test a null hypothesis H_0 related to the problem, the corresponding frequentist p-value is not the probability that H_0 is true, but rather the probability of observing a result at least as extreme for the outcome under the null distribution in a sequence of similar inferences.

Bayesian inference is different from frequentist methods in multiple ways. First, the probability actually expresses the chance of an event happening in this case. Second, unknown parameters are treated as random variables that can be described with probability distributions. Under the Bayesian paradigm, a probability expresses a degree of belief in an event. Essentially, this methodology starts with a set of prior beliefs based on scientific knowledge of the underlying problem. Then, one needs to update the prior belief in light of the observed data to come up with posterior beliefs. In the end, one analyzes the model fit

and sensitivity with respect to model assumptions. Because the model is set up via a full probabilistic approach, any probabilistic statements can be immediately interpreted as such without relating it to a sequence of independent repetitions. A $100p\%$ probability interval then expresses a range for the quantity of interest with coverage probability p, and a p-value is interpreted as the probability of replicated data being more extreme than observed data evaluated under a specified test statistic.

We perform Bayesian inference conditional on observations to estimate the unknown parameters as we want to assess the parameter uncertainty. Unlike the frequentist methods that yield point estimates of unknown parameters, Bayesian methods yield a posterior distribution of the unknown parameters, which allows us to understand their uncertainty.

4.2 Bayesian Inference

The fundamental usage of Bayesian inference is based on Bayes' theorem; that is, given the probability distribution for the parameters of interest $\boldsymbol{\Theta}$ and the data $\boldsymbol{y}_{1:T}$, the posterior distribution for $\boldsymbol{\Theta}$, on which all inference is based, depends on the observed values. In particular, the likelihood distribution represents the data generating mechanism.

In the case of state-space models, the likelihood has two forms: the marginal likelihood and the complete data likelihood. The marginal likelihood is shown in Equation (3.5). In constructing the marginal likelihood, we consider all possible values of latent variables that can have been generated by observed data. Typically, the marginal likelihood is hard to evaluate in closed form as it involves multidimensional integrals.

The complete data likelihood is constructed assuming that the values of the latent variables are known. Indeed, it is not true, but the value of each latent variables can be imputed as part of the estimation procedure when using Bayesian methods. In our case, the complete data likelihood can be written as

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} | \,\boldsymbol{\Theta}\right) &= p\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right) \prod_{t=1}^{T} p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{\kappa}_{t-1},\, \boldsymbol{\Theta}\right) \prod_{t=1}^{T} p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t,\, \boldsymbol{\Theta}\right), \\ &= p\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right) \underbrace{p\left(\boldsymbol{\kappa}_{1:T} \,|\, \boldsymbol{\Theta}\right)}_{\text{Transition}} \underbrace{p\left(\boldsymbol{y}_{1:T} \,|\, \boldsymbol{\kappa}_{0:T},\, \boldsymbol{\Theta}\right)}_{\text{Measurement}}, \end{aligned} \tag{4.1}$$

where $p\left(\boldsymbol{\kappa}_t \,|\, \boldsymbol{\kappa}_{t-1},\, \boldsymbol{\Theta}\right)$ is the transition density obtained from Equation (3.1), the measurement density $p\left(\boldsymbol{y}_t \,|\, \boldsymbol{\kappa}_t,\, \boldsymbol{\Theta}\right)$ is implied by the measurement equation of Equation (3.2), and $p\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right)$ is the initial distribution for the latent factors. The prior belief on $\boldsymbol{\Theta}$ are converted into the probability distribution $p\left(\boldsymbol{\Theta}\right)$. Then, Bayes' theorem states that

$$p\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right) = \frac{p\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T},\, \boldsymbol{y}_{1:T}\right)}{p\left(\boldsymbol{y}_{1:T}\right)} = \frac{\mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\Theta}\right)}{\int \mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right) p\left(\boldsymbol{\Theta}\right)\, d\boldsymbol{\Theta}},$$

where the denominator

$$\int \mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\Theta}\right)\, d\boldsymbol{\Theta}$$

is a constant, leading to the following expression:

$$\underbrace{p\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right)}_{\text{Posterior}} \propto \underbrace{\mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right)}_{\text{Likelihood}}\, \underbrace{p\left(\boldsymbol{\Theta}\right)}_{\text{Prior}}, \tag{4.2}$$

which summarizes the key elements of Bayesian inference. Each component of Equation (4.2) is discussed in the rest of this section.

4.2.1 The Likelihood

A likelihood function takes the data set as given and gives all of the relevant information to the evaluation of statistical evidence. The likelihood function can be obtained by using Equation (4.1). The transition density can be obtained by using Equation (2.5), and the general form of the measurement part of Equation (4.1) is given in Equation (3.3). Finally, for the CBD-X(3) model, the measurement part of Equation (4.1) is given by:

$$\begin{aligned} &p\left(\boldsymbol{y}_{1:T} \,|\, \boldsymbol{\kappa}_{0:T},\, \boldsymbol{\Theta}\right) \\ &\propto \exp\left(-\sum_{x,t} \frac{D_{x,t}}{2}\left(\log\left(\widehat{m}_{x,t}\right) - \alpha_x - \kappa_t^{(1)} - \kappa_t^{(2)}(x-\overline{x}) - \kappa_t^{(3)}((x-\overline{x})^2 - \sigma_x^2)\right)^2\right). \end{aligned}$$

4.2.2 The Prior

The prior distribution plays a vital role in determining the posterior distribution. In practice, prior distributions are specified using available information, such as experts' opinions or the results of previous studies. In this latter case, the prior distribution is called an informative prior distribution. Similarly, if the prior does not contain any information based on prior beliefs, it is called a non-informative prior distribution.

To obtain the posterior distribution within the CBD-X models, uniform prior distributions are assumed throughout this report, except for the variance (i.e., v_{ii}) and covariance (i.e, v_{ij} for $i \neq j$) parameters of the covariance matrix $\boldsymbol{\Sigma}$ in Equation (2.5).

The variance parameters v_{ii} are assumed to follow the half-normal distribution with a mean parameter equal to 0 and a variance equal to 10 such that the variance parameters v_{ii} have positive values. The variance value is chosen because a large value makes sure that the distribution is proper and yet non-informative. The prior probability density function (pdf) of the variance parameter v_{ii} is

$$f(v_{ii}) \propto e^{\frac{-v_{ii}^2}{20}}, \quad v_{ii} \in (0, \infty).$$

The prior distribution of the covariance parameters v_{ij} follows a truncated-normal distribution with the same mean and variance as the variance parameters, but are assumed to be within the interval $(-\sqrt{v_{ii}v_{jj}}, \sqrt{v_{ii}v_{jj}})$ ensuring the correlation coefficients between the series are within $(-1, 1)$. The prior pdf of the covariance parameter v_{ij} is

$$f\left(v_{ij} \,\middle|\, \{v_{ss}\}_{s=1}^{3}\right) \propto \frac{\phi\left(\frac{v_{ij}}{\sqrt{10}}\right)}{-2\Phi\left(\frac{\sqrt{v_{ii}v_{jj}}}{\sqrt{10}}\right)}, \quad v_{ij} \in (-\sqrt{v_{ii}v_{jj}}, \quad \sqrt{v_{ii}v_{jj}}),$$

where $\phi(x)$ is the standard normal distribution evaluated at x, and $\Phi(x)$ is its cumulative distribution function evaluated at x as well.

4.2.3 The Posterior

The posterior distribution is the probability distribution of unknown parameters, treated as a random variable, conditional on the data. The aim of the Bayesian estimation procedure for the CBD-X models is to identify joint posterior distributions for the latent variables $\boldsymbol{\kappa}_t$ and the parameters $\boldsymbol{\Theta}$. Due to the complexity of both the prior distributions and the likelihood function, no closed-form solutions for this joint posterior distributions can be derived.

4.3 Markov Chain Monte Carlo

Because we cannot derive the closed-form solution of the joint posterior distribution for all the unknown parameters within the CBD-X models, we estimate them by applying a sample-based approach called Markov chain Monte Carlo (MCMC) methods. MCMC methods generate samples from a given probability distribution, which is also called the target distribution, and features of this probability distribution can then be estimated from the samples generated. In our case, the target distribution is the joint posterior distribution $p\left(\boldsymbol{\Theta}, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right)$. However, we sample the latent variables and unknown parameters separately and use different methods. We first sample unknown parameters using traditional MCMC methods; then, we sample latent variables by the SMC method. We already introduced the SMC method in Chapter 3, and in this section, we will introduce the sampling method for the unknown parameters.

MCMC methods combine two main elements. Monte Carlo focuses on drawing a set of independent random samples from a target distribution, and these samples can be used to approximate the target distribution with an empirical distribution as shown in Chapter 3. Markov Chain, on the other hand, means that the next sample is dependent only on the current sample value. Therefore, MCMC is a class of methods aimed at generating successive samples from a target probability distribution. We will now introduce the two

main sampling algorithms used in this study: the Metropolis-Hastings algorithm and the Gibbs sampler.

4.3.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings (MH) algorithm is one of the most useful methods to construct MCMC samples. The Markov chain constructed from this method asymptotically reaches a unique stationary distribution $\pi\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right)$, such that $\pi\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right)$ approaches to the target distribution $p\left(\boldsymbol{\Theta},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}\right)$.

This algorithm has two main ingredients: a proposal distribution and an acceptance probability. The algorithm starts by setting initial parameters $\boldsymbol{\Theta}^{(0)}$. Then, at iteration i and depending on the previous parameter value $\boldsymbol{\Theta}^{(i-1)}$, the algorithm generates a candidate for the new parameter value $\boldsymbol{\Theta}^{(\text{New})}$ from a proposal distribution $q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)$. Same as the proposal distribution in the SMC algorithm, it can be any distribution. We define the transition density $p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)$ as the conditional probability of moving to $\boldsymbol{\Theta}^{(\text{New})}$ from $\boldsymbol{\Theta}^{(i-1)}$.

A sufficient but not necessary condition for the existence of stationary distribution is that the Markov chain be reversible. A Markov chain is reversible if each transition is reversible; that is, for every pair of parameters $\boldsymbol{\Theta}^{(\text{New})}$ and $\boldsymbol{\Theta}^{(i-1)}$, the probability of being in state $\boldsymbol{\Theta}^{(i-1)}$ and transitioning to state $\boldsymbol{\Theta}^{(\text{New})}$ must be equal to the probability of being in state $\boldsymbol{\Theta}^{(\text{New})}$ and transitioning to state $\boldsymbol{\Theta}^{(i-1)}$:

$$p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right) p\left(\boldsymbol{\Theta}^{(i-1)}\right) = p\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{\Theta}^{(\text{New})}\right) p\left(\boldsymbol{\Theta}^{(\text{New})}\right). \tag{4.3}$$

The transition density can be decomposed into

$$p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right) = q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right) r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right),$$

where $r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right)$, called the acceptance probability, is the probability of moving to the proposed value $\boldsymbol{\Theta}^{(\text{New})}$. To ensure the equilibrium in Equation (4.3) is reached in each iteration, the acceptance probability is thus defined as

$$r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right) = \min\left(\frac{p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) q\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{\Theta}^{(\text{New})}\right)}{p\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)},\, 1\right).$$

If the proposal distribution is symmetric, i.e., $q\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{\Theta}^{(\text{New})}\right) = q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)$, then the acceptance probability collapses to

$$r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right) = \min\left(\frac{p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}{p\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)},\, 1\right),$$

and this algorithm is called the Metropolis algorithm. Furthermore, if a random walk proposal is used, i.e.,

$$\boldsymbol{\Theta}^{(\text{New})} = \boldsymbol{\Theta}^{(i-1)} + \boldsymbol{\zeta},$$

where $\boldsymbol{\zeta}$ is a symmetric multivariate distribution centred at 0, then this algorithm is known as a random walk MH algorithm.

After the acceptance probability is calculated, a random number u is generated from the uniform distribution over $(0, 1)$ as probabilities must fall between 0 and 1. If $u \leq r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right)$, then the new parameter value $\boldsymbol{\Theta}^{(\text{New})}$ is accepted and the parameter value moves from the current parameter value $\boldsymbol{\Theta}^{(i-1)}$ to parameter value $\boldsymbol{\Theta}^{(\text{New})}$. Otherwise, $\boldsymbol{\Theta}^{(\text{New})}$ is rejected and the parameter value remains the same.

Assuming the number of iterations required is M, the MH algorithm summarized in Algorithm 2.

4.3.2 Gibbs Sampler

The Gibbs sampler can be treated as a special case of the MH algorithm. It works if closed-from solution of the full conditional posterior distributions of $\Theta_g \in \boldsymbol{\Theta}$, $\boldsymbol{\Theta} = [\Theta_1 \quad ... \quad \Theta_d]$, are available; that is, for $g \in \{1, ..., d\}$, the full conditional probability for Θ_g given all other parameters, $p\left(\Theta_g \,|\, \boldsymbol{\Theta}_{-g}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)$, is available, where the set $\boldsymbol{\Theta}_{-g}$ contain all the parameters except for parameter Θ_g. Just like in the MH algorithm, the Gibbs sampler starts with setting initial values for each unknown parameter. Then, instead of sampling a new candidate from a proposal distribution, the Gibbs sampler generates a candidate directly from the full conditional posterior distribution of this parameter $p\left(\Theta_g \,|\, \boldsymbol{\Theta}_{-g}, \boldsymbol{y}_{1:T}\right)$ and, thus, it always accepts the new candidate because

$$\begin{aligned}
&r\left(\Theta_g^{(\text{New})}, \Theta_g^{(i-1)}\right) \\
&= \min\left(\frac{p\left(\Theta_g^{(\text{New})} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(i-1)} \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}{p\left(\Theta_g^{(i-1)} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}, 1\right) \\
&= \min\left(\frac{p\left(\Theta_g^{(\text{New})} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(i-1)} \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}{p\left(\Theta_g^{(i-1)} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}, 1\right) \\
&= \min\left(\frac{\frac{p\left(\Theta_g^{(\text{New})}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(i-1)}, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}{p(\boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T})\, p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}}{\frac{p\left(\Theta_g^{(i-1)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right) p\left(\Theta_g^{(\text{New})}, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}{p(\boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T})\, p\left(\boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)}}, 1\right) \\
&= \min(1, 1) = 1,
\end{aligned}$$

where at iteration i, $\boldsymbol{\Theta}_{-g}^{(i)}$ contains the parameters' value at iteration i for parameters appearing before Θ_g and the parameters' value at iteration $i-1$ for those appearing after, because the Gibbs sampler updates parameters' value sequentially.

The Gibbs sampler is summarized in Algorithm 3.

Algorithm 2 Metropolis-Hastings Algorithm

1: set initial parameters $\boldsymbol{\Theta}^{(0)}$
2: **for** $i = 1, ..., M$ **do**
3: sample $\boldsymbol{\Theta}^{(\text{New})} \sim q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)$
4: compute the acceptance probability:

$$r\left(\boldsymbol{\Theta}^{(\text{New})}, \boldsymbol{\Theta}^{(i-1)}\right) = \min\left(\frac{p\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)\, q\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{\Theta}^{(\text{New})}\right)}{p\left(\boldsymbol{\Theta}^{(i-1)} \,\middle|\, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)\, q\left(\boldsymbol{\Theta}^{(\text{New})} \,\middle|\, \boldsymbol{\Theta}^{(i-1)}\right)}, 1\right)$$

5: generate $u \sim \text{Uniform}(0, 1)$
6: set

$$\boldsymbol{\Theta}^{(i)} = \begin{cases} \boldsymbol{\Theta}^{(\text{New})} & \text{if } u \leq r\left(\boldsymbol{\Theta}^{(\text{New})}, \Theta^{(i-1)}\right), \\ \boldsymbol{\Theta}^{(i-1)} & \text{otherwise.} \end{cases}$$

7: **end for**

Algorithm 3 Gibbs Sampler

1: set initial parameters $\boldsymbol{\Theta}^{(0)} = [\Theta_1^{(0)} \quad ... \quad \Theta_d^{(0)}]$
2: **for** $i = 1, ..., M$ **do**
3: **for** $g = 1, ..., d$ **do**
4: set $\boldsymbol{\Theta}_{-g}^{(i)} = [\Theta_1^{(i)}, \ldots, \Theta_{g-1}^{(i)}, \Theta_{g+1}^{(i-1)}, \ldots, \Theta_d^{(i-1)}]$
5: sample $\Theta_g^{(i)} \sim \pi\left(\Theta_g \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)}, \boldsymbol{y}_{1:T}, \boldsymbol{\kappa}_{0:T}\right)$
6: **end for**
7: set $\boldsymbol{\Theta}^{(i)} = [\Theta_1^{(i)} \quad ... \quad \Theta_d^{(i)}]$
8: **end for**

4.4 Particle Markov Chain Monte Carlo

To incorporate the dynamics of the period effects within the model estimation procedure and thus reduce the two-stage estimation approach into a single stage, we apply the particle Markov chain Monte Carlo (pMCMC) algorithm. In general, a pMCMC sampling method uses the SMC algorithm to generate a proposal distribution within an MCMC algorithm. The two main algorithms in the pMCMC framework are the particle Metropolis-Hastings sampler and the particle Gibbs (PG) sampler. We will focus on the particle Gibbs sampler in this report.

4.4.1 Conditional Sequential Monte Carlo

The PG sampler, introduced by Andrieu et al. (2010), is an extension of the Gibbs sampler. Instead of generating random variables sequentially at each iteration, the PG sampler generates all random variables at the same time. The main idea of the PG sampler is to run a conditional sequential Monte Carlo (cSMC) algorithm iteratively. The cSMC algorithm is similar to the standard SMC algorithm introduced in Chapter 3, except that a pre-specified path $\boldsymbol{\kappa}_{0:T}^*$ is retained in all the resampling steps, whereas the remaining $J-1$ particles are generated as in the standard SMC algorithm. For simplicity, we set the last particle $\boldsymbol{\kappa}_t^J = \boldsymbol{\kappa}_t^*$, and its ancestor variable $A_t^J = J$ for all t. The cSMC algorithm is summarized in Algorithm 4.

Algorithm 4 Conditional Sequential Monte Carlo

1: sample $\boldsymbol{\kappa}_0^j \sim q\left(\boldsymbol{\kappa}_0 \,|\, \boldsymbol{\Theta}\right)$ for $j = 1, ..., J-1$ and set $\boldsymbol{\kappa}_0^J = \boldsymbol{\kappa}_0^*$
2: set $A_0^j = j$ for $j = 1, ..., J$
3: **for** $t = 1, ..., T$ **do**
4: sample $\boldsymbol{\kappa}_t^j \sim q\left(\boldsymbol{\kappa}_t \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)$
5: set $\boldsymbol{\kappa}_{0:t}^j = \left[\boldsymbol{\kappa}_{0:t-1}^{A_{t-1}^j}, \boldsymbol{\kappa}_t^j\right]$ for $j = 1, ..., J-1$, and set $\boldsymbol{\kappa}_t^J = \boldsymbol{\kappa}_t^*$
6: compute the weight for $j = 1, ..., J$:

$$
\begin{aligned}
w_t^j &= \frac{p\left(\boldsymbol{\kappa}_{0:t}^j,\, \boldsymbol{y}_{1:t}, \boldsymbol{\Theta}\right)}{p\left(\boldsymbol{\kappa}_{0:t-1}^j,\, \boldsymbol{y}_{1:t-1}, \boldsymbol{\kappa}\right)\, q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)} \\
&= \frac{p\left(\boldsymbol{y}_t \,\middle|\, \boldsymbol{\kappa}_t^j, \boldsymbol{\Theta}\right)\, p\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)}{q\left(\boldsymbol{\kappa}_t^j \,\middle|\, \boldsymbol{y}_t;\, \boldsymbol{\kappa}_{t-1}^{A_{t-1}^j}, \boldsymbol{\Theta}\right)}
\end{aligned}
$$

7: normalize the weight: $W_t^j = \frac{w_t^j}{\sum_{k=1}^J w_t^k}$ for $j = 1, ..., J$
8: sample A_t^j from a multinomial distribution with support 1 to $J-1$ and weights $\boldsymbol{W}_t$
9: set $A_t^J = J$
10: **end for**

Once the cSMC algorithm is completed, we get J sets of sampled latent variables $\boldsymbol{\kappa}_{0:T}$. We then select randomly a trajectory of $\boldsymbol{\kappa}_{0:T}$ by using the ancestor variables $A_{0:T}^j$. Recall that the ancestor variables $A_{1:T}^j$ allow us to keep track of the particles. From the cSMC algorithm, we have that $\boldsymbol{\kappa}_{0:t}^j = \left[\boldsymbol{\kappa}_{0:t-1}^{A_{t-1}^j}, \boldsymbol{\kappa}_t^j\right]$ at time t for $j = 1, ..., J$.

To make the tracking process easier, Andrieu et al. (2010) introduce an ancestral lineage B_t^j, which is the index of the particles $\boldsymbol{\kappa}_{0:t}$ with ancestor variable at time t for $j = 1, ..., J$. They define $B_T^j = j$ for $j = 1, ..., J$, and $B_t^j = A_t^{B_{t+1}^j}$ for $t = T-1, \ldots, 1$. As a result, $\boldsymbol{\kappa}_{0:T}^j = \{\boldsymbol{\kappa}_0^{B_0^j}, \ldots, \boldsymbol{\kappa}_T^{B_T^j}\}$ and $B_{0:T}^j = \{B_1^j, \ldots, B_T^j\}$ for any $j = 1, ..., J$. To select a trajectory

of $\boldsymbol{\kappa}_{0:T}$, we sample an index j from a multinomial distribution with weights $\boldsymbol{W}_T$ as the last step does not allow for resampling within the cSMC algorithm.

A set of latent variables $\boldsymbol{\kappa}_{0:T}$ is generated from the cSMC algorithm, and the corresponding ancestor variables $A^j_{0:T}$ are shown in Figure 4.1 for $J = 4$ particles and $T = 2$. Assume the index $j = 2$ is sampled from the multinomial distribution, then the path in darker blue is the selected trajectory. As $B^j_2 = 2$, $B^j_1 = A^{B^j_2}_1 = A^2_1 = 3$, and $B^j_0 = A^{B^j_1}_0 = A^3_0 = 3$, we have that $B^j_{0:2} = \{3, 3, 2\}$ and $\boldsymbol{\kappa}^j_{0:2} = \{\boldsymbol{\kappa}^3_0, \boldsymbol{\kappa}^3_1, \boldsymbol{\kappa}^2_2\}$.

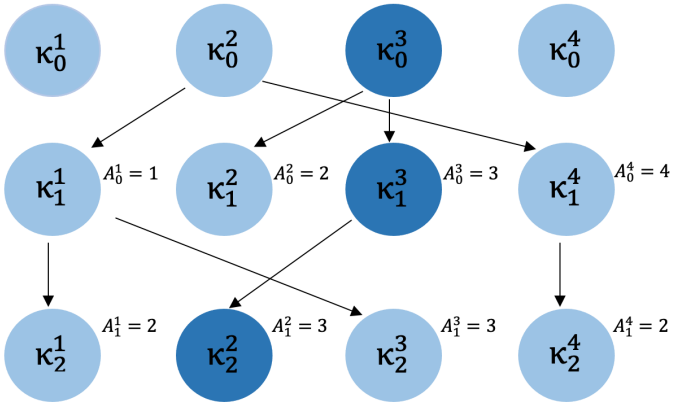

Figure 4.1: Example of ancestral lineages generated by a cSMC algorithm for $J = 4$ and $T = 2$.

4.4.2 Particle Gibbs

The particle Gibbs sampler starts by setting initial values for all the particles $\boldsymbol{\kappa}_{0:T}$ and the unknown parameters $\boldsymbol{\Theta}$. Then, for each iteration, we sample the unknown parameters by using the full conditional posterior distribution as in the Gibbs sampler. The difference with the original Gibbs sampler comes from the last step: the particle Gibbs runs a cSMC targeting $p(\boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$, that is, we sample $\boldsymbol{\kappa}_{0:T} \sim \hat{p}(\boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{y}_{1:T}, \boldsymbol{\Theta})$. Thus, instead of sampling $\boldsymbol{\kappa}_t$ for $t \in \{0, ..., T\}$ one by one as done under the Gibbs sampler, we sample all the $\boldsymbol{\kappa}_{0:T}$ at once. The particle Gibbs algorithm is summarized in Algorithm 5.

4.4.3 pMCMC Strategy for the Estimation of CBD-X Models

We apply cSMC for estimating the latent variables in the CBD-X models and apply an MCMC method to estimate the other parameters $\boldsymbol{\Theta}$. The MCMC method we use is called the Metropolis-within-Gibbs algorithm, where parameters are updated sequentially via the MH algorithm in each iteration, conditional on the rest of the parameters. In the case where the full conditional posterior distribution is available in closed-form, the Gibbs sampler is used.

Algorithm 5 Particle Gibbs

set initial values $\boldsymbol{\Theta}^{(0)}$ and $\boldsymbol{\kappa}_{0:T}^{(0)}$
for $i = 1, ..., M$ **do**
  for $g = 1, ..., d$ **do**
    $\boldsymbol{\Theta}_{-g}^{(i)} = [\Theta_1^{(i)}, \ldots, \Theta_{g-1}^{(i)}, \Theta_{g+1}^{(i-1)}, \ldots, \Theta_d^{(i-1)}]$
    sample $\Theta_g^{(i)} \sim p\left(\Theta_g^{(i)} \,\middle|\, \boldsymbol{\Theta}_{-g}^{(i)},\, \boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T}^{(i-1)}\right)$
  end for
  $\boldsymbol{\Theta}^{(i)} = [\Theta_1^{(i)} \ \ldots \ \Theta_d^{(i)}]$
  sample $\boldsymbol{\kappa}_{0:T}^{(i)} \sim \hat{p}\left(\boldsymbol{\kappa}_{0:T}^{(i)} \,\middle|\, \boldsymbol{y}_{1:T},\, \boldsymbol{\Theta}^{(i)}\right)$ using the cSMC
end for

For the sake of conciseness, we use the CBD-X(3) model. The complete data likelihood function derived in the previous section is

$$\begin{aligned}
&\mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right) \\
&\propto \exp\left(-\sum_{x,t} \frac{D_{x,t}}{2}\left(\log\left(\widehat{m}_{x,t}\right) - \alpha_x - \kappa_t^{(1)} - \kappa_t^{(2)}(x-\overline{x}) - \kappa_t^{(3)}((x-\overline{x})^2 - \sigma_x^2)\right)^2\right) \\
&\quad \prod_{t=1}^{T} \frac{1}{(2\pi)^2\,|\boldsymbol{\Sigma}|} \exp\left(-\frac{1}{2}\left(\boldsymbol{\kappa}_t - \boldsymbol{\mu}_t^*\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\kappa}_t - \boldsymbol{\mu}_t^*\right)\right),
\end{aligned}$$

where

$$\boldsymbol{\kappa}_t = \begin{bmatrix}\kappa_t^{(1)} & \kappa_t^{(2)} & \kappa_t^{(3)}\end{bmatrix}^\top, \quad \boldsymbol{\mu}_t^* = \begin{bmatrix}\kappa_{t-1}^{(1)} - \mu_1 & \kappa_{t-1}^{(2)} - \mu_2 & \kappa_{t-1}^{(3)} - \mu_3\end{bmatrix}^\top,$$

and

$$\boldsymbol{\Sigma} = \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{12} & v_{22} & v_{23} \\ v_{13} & v_{23} & v_{33} \end{bmatrix}.$$

Furthermore, to overcome the identification issue when estimating the CBD-X models, one has to impose a non-unique choice of constraints to restrict the model. In this report, we assume $\kappa_0^{(1)} = 0$, $\kappa_0^{(2)} = 0$, and $\kappa_0^{(3)} = 0$.[2]

We assume that the prior distribution for each α_x is uniform for each age x and that $\boldsymbol{\alpha}_x$ and $\boldsymbol{\kappa}_t$ are independent. Then, the full conditional posterior distribution of α_x for each age x is

$$f(\alpha_x \,|\, \boldsymbol{\Theta}_{-\alpha_x}, \boldsymbol{\kappa}_{0:T}, \boldsymbol{y}_{1:T})$$

[2]Please note that our constraints are different from that in Dowd et al. (2020), but similar to those used in Cairns et al. (2019).

$$
\begin{aligned}
&\propto \exp\left(-\sum_{t=1}^{T}\frac{\hat{D}_{x,t}}{2}\left(\log\left(\widehat{m}_{x,t}\right)-\alpha_x-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)-\kappa_t^{(3)}\left(\left(x-\overline{x}\right)^2-\sigma_x^2\right)\right)^2\right)\\
&\propto \exp\left(-\sum_{t=1}^{T}\frac{\hat{D}_{x,t}}{2}\left(\alpha_x^2-2\alpha_x\Big(\log(\widehat{m}_{x,t})-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)\right.\right.\\
&\qquad\qquad\left.\left.-\kappa_t^{(3)}\left(\left(x-\overline{x}\right)^2-\sigma_x^2\right)\Big)\right)\right)\\
&\propto \exp\left(-\sum_{t=1}^{T}\frac{\hat{D}_{x,t}}{2}\alpha_x^2+2\alpha_x\sum_{t=1}^{T}\frac{\hat{D}_{x,t}}{2}\Big(\log(\widehat{m}_{x,t})-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)\right.\\
&\qquad\qquad\left.-\kappa_t^{(3)}((x-\overline{x})^2-\sigma_x^2)\Big)\right)\\
&\propto \exp\left(\frac{\sum_{t=1}^{T}\hat{D}_{x,t}}{-2}\Big(\alpha_x^2-2\alpha_x\sum_{t=1}^{T}\hat{D}_{x,t}\Big(\log(\widehat{m}_{x,t})-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)\right.\right.\\
&\qquad\qquad\left.\left.-\kappa_t^{(3)}((x-\overline{x})^2-\sigma_x^2)\Big)\frac{1}{\sum_{t=1}^{T}\hat{D}_{x,t}}\right)\right)\\
&\propto \exp\left(\frac{\sum_{t=1}^{T}\hat{D}_{x,t}}{-2}\Big(\alpha_x-\sum_{t=1}^{T}\hat{D}_{x,t}\Big(\log(\widehat{m}_{x,t})-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)\right.\right.\\
&\qquad\qquad\left.\left.-\kappa_t^{(3)}((x-\overline{x})^2-\sigma_x^2)\Big)\frac{1}{\sum_{t=1}^{T}\hat{D}_{x,t}}\right)^2\right),
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\alpha_x \,|\, \boldsymbol{\Theta}_{-\alpha_x}, \boldsymbol{\kappa}_{0:T}, \boldsymbol{y}_{1:T}\\
&\sim \mathcal{N}\left(\frac{\sum_{t=1}^{T}\hat{D}_{x,t}\left(\log(\widehat{m}_{x,t})-\kappa_t^{(1)}-\kappa_t^{(2)}\left(x-\bar{x}\right)-\kappa_t^{(3)}((x-\overline{x})^2-\sigma_x^2)\right)}{\sum_{t=1}^{T}\hat{D}_{x,t}}, \frac{1}{\sum_{t=1}^{T}\hat{D}_{x,t}}\right).
\end{aligned}
$$

Let $\boldsymbol{\Sigma}^{-1}$ be the inverse of the covariance matrix $\boldsymbol{\Sigma}$ such that:

$$
\boldsymbol{\Sigma}^{-1}=\begin{bmatrix} a_{11} & a_{12} & a_{13}\\ a_{12} & a_{22} & a_{23}\\ a_{13} & a_{23} & a_{33}\end{bmatrix}.
$$

Also, let

$$
b_t^{(1)}=\kappa_t^{(1)}-\kappa_{t-1}^{(1)}-\mu_1,\quad b_t^{(2)}=\kappa_t^{(2)}-\kappa_{t-1}^{(2)}-\mu_2,\quad \text{and}\quad b_t^{(3)}=\kappa_t^{(3)}-\kappa_{t-1}^{(3)}-\mu_3.
$$

Then,

$$
(\boldsymbol{\kappa}_t-\boldsymbol{\mu}_t^*)^\top\boldsymbol{\Sigma}^{-1}(\boldsymbol{\kappa}_t-\boldsymbol{\mu}_t^*)=(a_{11}b_t^{(1)}+a_{12}b_t^{(2)}+a_{13}b_t^{(3)})b_t^{(1)}
$$

$$+ (a_{12}b_t^{(1)} + a_{22}b_t^{(2)} + a_{23}b_t^{(3)})b_t^{(2)}$$
$$+ (a_{13}b_t^{(1)} + a_{23}b_t^{(2)} + a_{33}b_t^{(3)})b_t^{(3)},$$

and the full conditional posterior distribution of μ_1 with uniform prior can be written as

$$\begin{aligned}
&f(\mu_1 \,|\, \boldsymbol{\Theta}_{-\mu_1}, \boldsymbol{\kappa}_{0:T}) \\
&\propto \prod_{t=1}^{T} \frac{1}{(2\pi)^2 \,|\boldsymbol{\Sigma}|} \exp\left(-\frac{1}{2}\left(\boldsymbol{\kappa}_t - \boldsymbol{\mu}_t^*\right)^\top \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\kappa}_t - \boldsymbol{\mu}_t^*\right)\right) \\
&\propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T}(a_{11}b_t^{(1)} + a_{12}b_t^{(2)} + a_{13}b_t^{(3)})b_t^{(1)} + (a_{12}b_t^{(1)} + a_{22}b_t^{(2)} + a_{23}b_t^{(3)})b_t^{(2)} \\
&\qquad\qquad + (a_{13}b_t^{(1)} + a_{23}b_t^{(2)} + a_{33}b_t^{(3)})b_t^{(3)}\Bigg) \\
&\propto \exp\left(-\frac{1}{2}\sum_{t=1}^{T} a_{11}\left(b_t^{(1)}\right)^2 + 2a_{12}b_t^{(1)}b_t^{(2)} + 2a_{13}b_t^{(1)}b_t^{(3)}\right) \\
&\propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} a_{11}\left(\mu_1^2 - 2\mu_1\left(\kappa_t^{(1)} - \kappa_{t-1}^{(1)}\right)\right) + 2a_{12}\left(-\mu_1\left(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2\right)\right) \\
&\qquad\qquad + 2a_{13}\Big(-\mu_1\left(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3\right)\Big)\Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2}\sum_{t=1}^{T} a_{11}\mu_1^2 - 2\mu_1\Big(a_{11}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)}) + a_{12}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2) \\
&\qquad\qquad + a_{13}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3)\Big)\Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2}\Big(a_{11}T\mu_1^2 - 2\mu_1\sum_{t=1}^{T} a_{11}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)}) + a_{12}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2) \\
&\qquad\qquad + a_{13}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3)\Big)\Bigg) \\
&\propto \exp\Bigg(-\frac{1}{2\frac{1}{a_{11}T}}\Bigg(\mu_1^2 \\
&\qquad - 2\mu_1\sum_{t=1}^{T}\frac{a_{11}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)}) + a_{12}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2) + a_{13}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3)}{a_{11}T}\Bigg)\Bigg),
\end{aligned}$$

so

$$\begin{aligned}
&\mu_1 |\, \boldsymbol{\Theta}_{-\mu_1}, \boldsymbol{\kappa}_{0:T} \\
&\sim \mathcal{N}\left(\sum_{t=1}^{T}\frac{a_{11}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)}) + a_{12}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2) + a_{13}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3)}{a_{11}T}, \frac{1}{a_{11}T}\right).
\end{aligned}$$

Similarly,

$$\mu_2 \,|\, \boldsymbol{\Theta}_{-\mu_2}, \boldsymbol{\kappa}_{0:T} \sim \mathcal{N}\left(\sum_{t=1}^{T} \frac{a_{22}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)}) + a_{12}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)} - \mu_1) + a_{23}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)} - \mu_3)}{a_{22}T}, \frac{1}{a_{22}T}\right),$$

and

$$\mu_3 \,|\, \boldsymbol{\Theta}_{-\mu_3}, \boldsymbol{\kappa}_{0:T} \sim \mathcal{N}\left(\sum_{t=1}^{T} \frac{a_{33}(\kappa_t^{(3)} - \kappa_{t-1}^{(3)}) + a_{13}(\kappa_t^{(1)} - \kappa_{t-1}^{(1)} - \mu_1) + a_{23}(\kappa_t^{(2)} - \kappa_{t-1}^{(2)} - \mu_2)}{a_{33}T} . \frac{1}{a_{33}T}\right),$$

For the variance (i.e., $v_{11}, v_{22}, \text{and } v_{33}$) and covariance parameters (i.e, v_{12}, $v_{13}, \text{and } v_{23}$) of the covariance matrix $\boldsymbol{\Sigma}$, there is no closed-form solution for the full conditional posterior distribution. We thus apply the MH algorithm to update them with the prior distributions discussed in the previous section. The proposal distributions for the variance parameters are a truncated normal distribution centred at the values from the previous iteration and with lower bound at zero to make sure the value of variance is positive. The proposal distribution for the covariance parameters is a normal distribution centred at the last observation in the chain.

Chapter 5

Empirical Results

In this chapter, we present the results of the fitting procedures discussed above and the models' forecasting performance. First, we describe the data used in model estimation. Second, we discuss the estimation results of the three CBD-X models. Third, we show the 10-year out-of-sample forecasting results for each model. Finally, we apply the deviance information criterion for model selection.

5.1 Data Description

Data are from the Human Mortality Database for the Canadian male population. The Human Mortality Database provides detailed mortality and population data for several countries. Original data for the Canadian population are collected from Statistics Canada except the exposure of population, $E_{x,t}$ for 1961 and 1966, which are census enumerations. The observed number of deaths and the corresponding exposures are available for all ages $x = 0, ..., 109$ and all remaining ages, which are denoted by 110+. The Canadian data are available from the year 1921 to the year 2018.

We are interested in longevity risks, so we select the data from retirement age 65 to age 89. Although data are available for older ages, age at death is often misreported at these old ages, resulting in unreliable estimated death rates. We select the data from the year 1959 to the year 2008. This provides a large enough data set (i.e., covering 50 years) while allowing us to test the out-of-sample models' forecasting performance.

5.2 Estimation Results

The output of a pMCMC algorithm is a sequence of sampled parameters. Each of these sequences is called a chain. A trace plot shows the value of a parameters at each iteration in the algorithm against the iteration number. Usually, the first portion of the chain is discarded. This is called the burn-in period. A burn-in period increases the quality of sample by dropping early values which may be biased by the starting values. As shown in Figure

5.1, the early part of the chain for $\kappa_1^{(1)}$ behaves differently from the remaining part and is discarded.

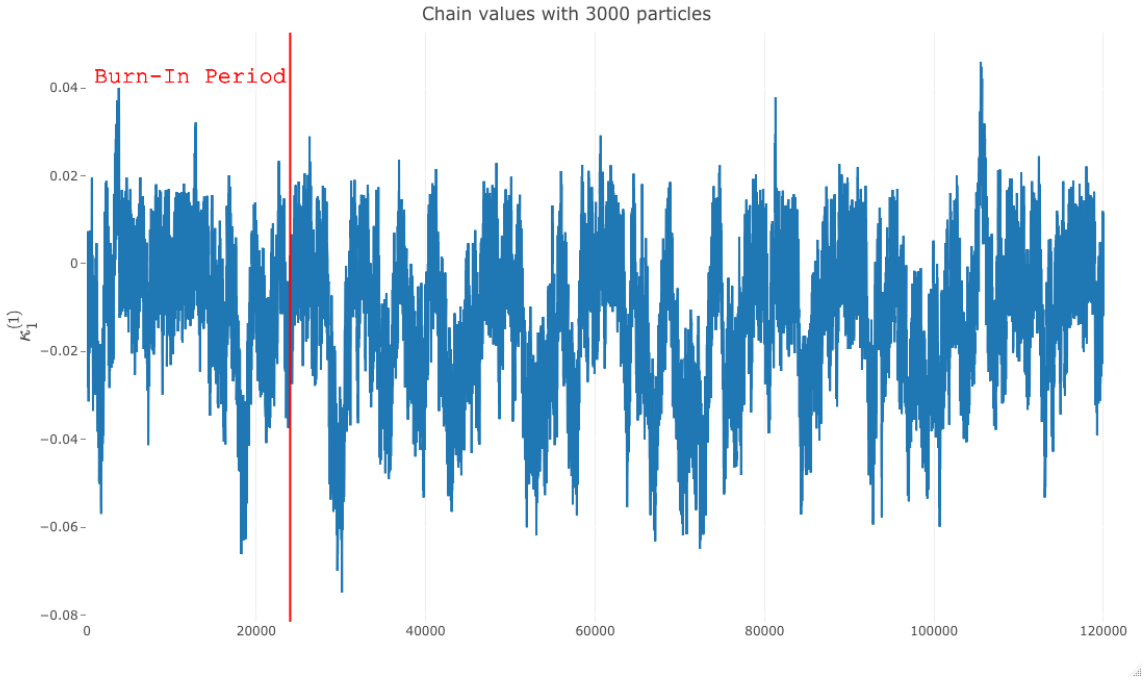

Figure 5.1: Trace plot of $\kappa_1^{(1)}$ with 3,000 particles and 120,000 iterations for the CBD-X(3) model.

The remaining samples in the chain produced by the pMCMC method yield the posterior distribution of each model parameter. For example, the posterior distribution of μ_1 shown in Figure 5.2.

When constructing a pMCMC algorithm, the speed of convergence should be within the practical constraints of time and computational power. The speed of convergence is mainly influenced by the number of iterations and the number of particles. The number of iterations within a pMCMC algorithm represents the total sample size simulated for each parameter, and the number of particles determines how many particles are used in the cSMC algorithm at each iteration. The number of iterations required for convergence varies from application to application. However, several tests can assess convergence. We choose the number of iterations and the number of particles by first looking at the trace plots. We then apply the Gelman-Rubin test to perform further convergence diagnostics. These tests will be discussed in Section 5.2.1.

The algorithm set-up for each model is summarized in Table 5.1. There are more particles in the CBD-X(3) model than for the other two models because it has three period-effect factors. Thus, it needs more particles in the cSMC algorithm to generate consistent approximations. Usually, we will set more iterations for models with more parameters as convergence tends to be slower in these cases. However, we use fewer iterations for the CBD-X(3)

model than the CBD-X(2) model to save on the run-time as we have more particles in the CBD-X(3) model.

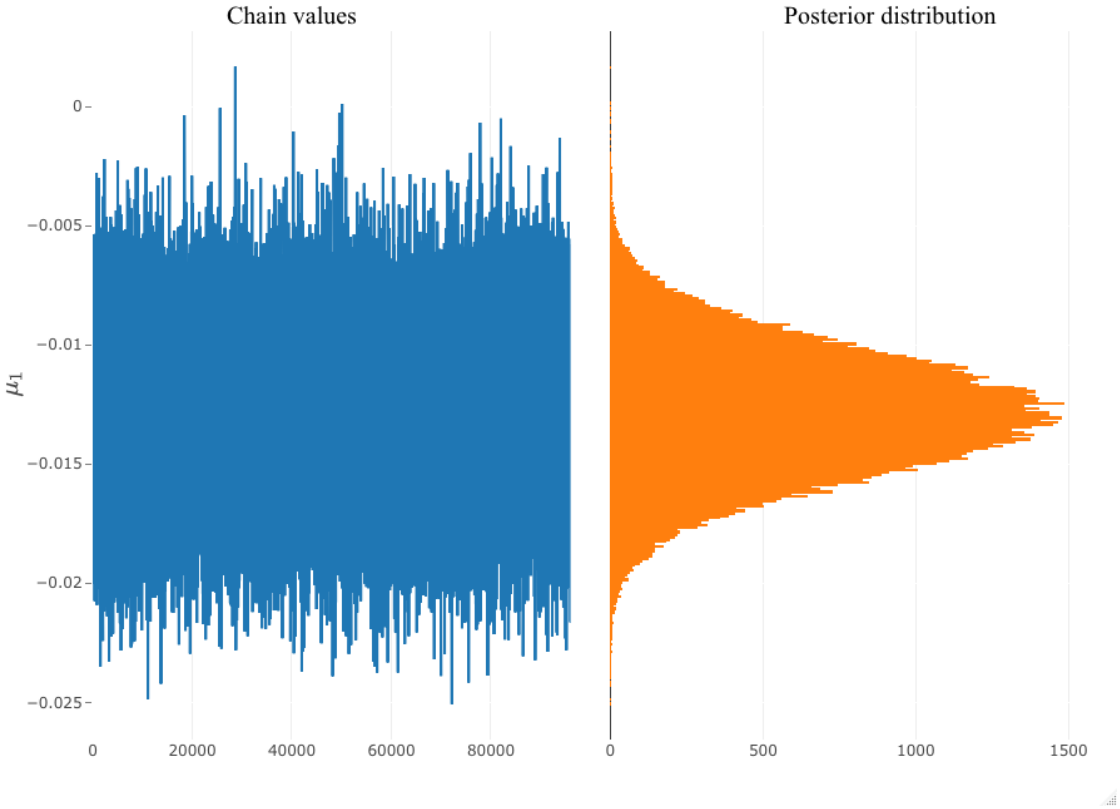

Figure 5.2: Trace plot of μ_1 with 3,000 particles after burn-in period within the CBD-X(3) model (left-hand side) and the corresponding posterior distribution (right-hand side).

Model	Numbers of Iterations (M)	Burn-In Period	Numbers of Particles (J)
CBD-X(1)	30,000	6,000	500
CBD-X(2)	300,000	60,000	500
CBD-X(3)	120,000	24,000	3,000

Table 5.1: Algorithm set-up for each model.

5.2.1 Convergence Diagnostics

As noted above, the outputs of the CBD-X models under the pMCMC algorithm must be diagnosed for convergence before performing any type of statistical inference. To assess convergence, we look at the trace plot first. Trace plots give insight into the behaviour of the Markov chain and point out possible flaws in the algorithm. If the pMCMC chain is stuck in some part of the state space, the trace plots show flat bits indicating slow convergence. Such a trace plot is observed if too many proposed values are rejected, for example, as

shown in Figure 5.3. Therefore, using 1,000 particles is not enough for the CBD-X(3) model to converge, and we increased the number of particles to 3,000 for this model. If too many proposed values are accepted consecutively, trace plots may move slowly and not explore the rest of the state space. If a trace plot exhibits rapid up-and-down variation with no long-term trends or drifts, then the parameter appears to have converged, as in Figure 5.4.

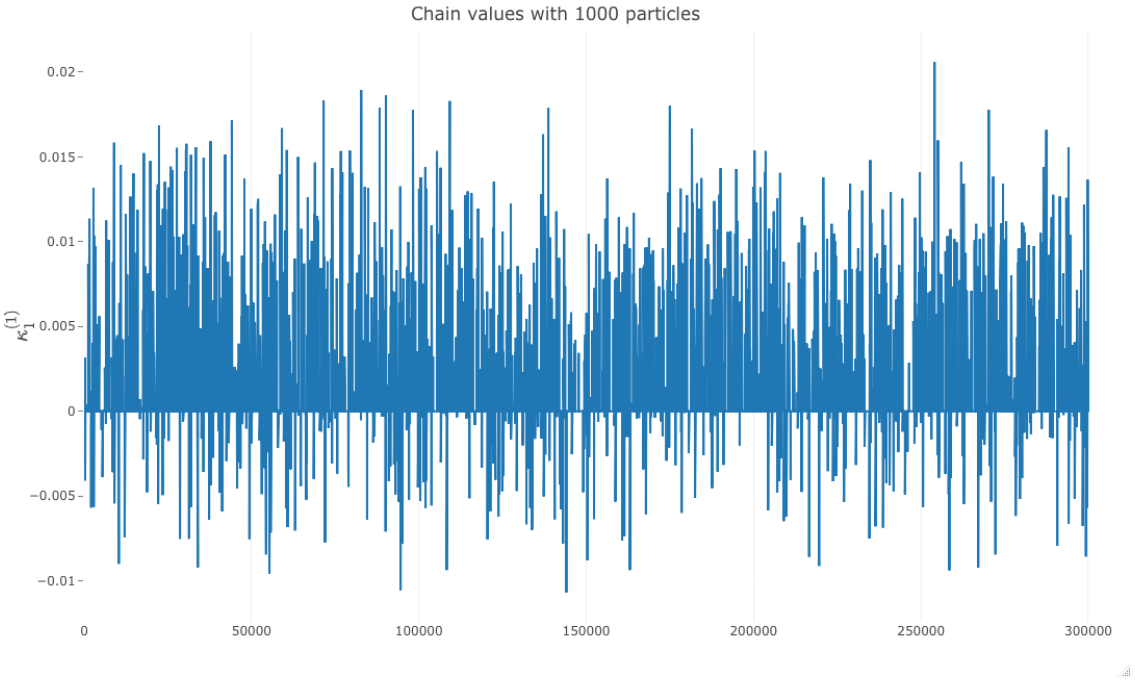

Figure 5.3: Trace plot of $\kappa_1^{(1)}$ with 1,000 particles and 300,000 iterations within the CBD-X(3) model.

Once we are satisfied with visual inspection of the trace plots, we run the pMCMC algorithm with different starting values for each CBD-X model to perform Gelman-Rubin tests. Specifically, Gelman and Rubin (1992) evaluate pMCMC convergence by comparing the estimated between-chain and within-chain variances for each parameter Ψ, where $\boldsymbol{\Psi} \equiv (\boldsymbol{\kappa}_{0:T}, \boldsymbol{\Theta})$. Large differences between these variances indicate nonconvergence because, when running multiple chains with different starting values in parallel with many iterations, they should converge to the same stationary distribution. Suppose we have H chains with different starting values, and each of length M. Let $\bar{\Psi}_h$ and $\hat{\sigma}_h^2$ be the sample posterior mean and variance of the estimated parameter Ψ in h^{th} chain. Let the overall sample posterior mean be $\bar{\Psi} = \frac{1}{H}\sum_{h=1}^{H} \bar{\Psi}_h$. Then, the between-chains variance is given by:

$$\frac{B}{M} = \frac{\sum_{h=1}^{H} \left(\bar{\Psi}_h - \bar{\Psi}\right)^2}{H-1},$$

which measures how the means in each chain vary around the overall mean. The within-chain variance is given by:

$$W = \frac{1}{H} \sum_{h=1}^{H} \hat{\sigma}_h^2,$$

which is the averaged variances of the chains. Under certain stationary conditions, the combined variance

$$\hat{V} = \frac{M-1}{M} W + \frac{H+1}{H} \frac{B}{M}$$

is an unbiased estimator of the marginal posterior variance of Ψ. However, if the chains have converged, the within-chain variance W is also an unbiased estimate of the marginal posterior variance of Ψ. Hence if the chains converged, then for each parameter, the ratio

$$R = \sqrt{\frac{\hat{V}}{W}} \approx 1.$$

By implementing the Gelman-Rubin test, we confirm that all chains for the three models converged under the pMCMC method. Detailed results are shown in Appendix B.

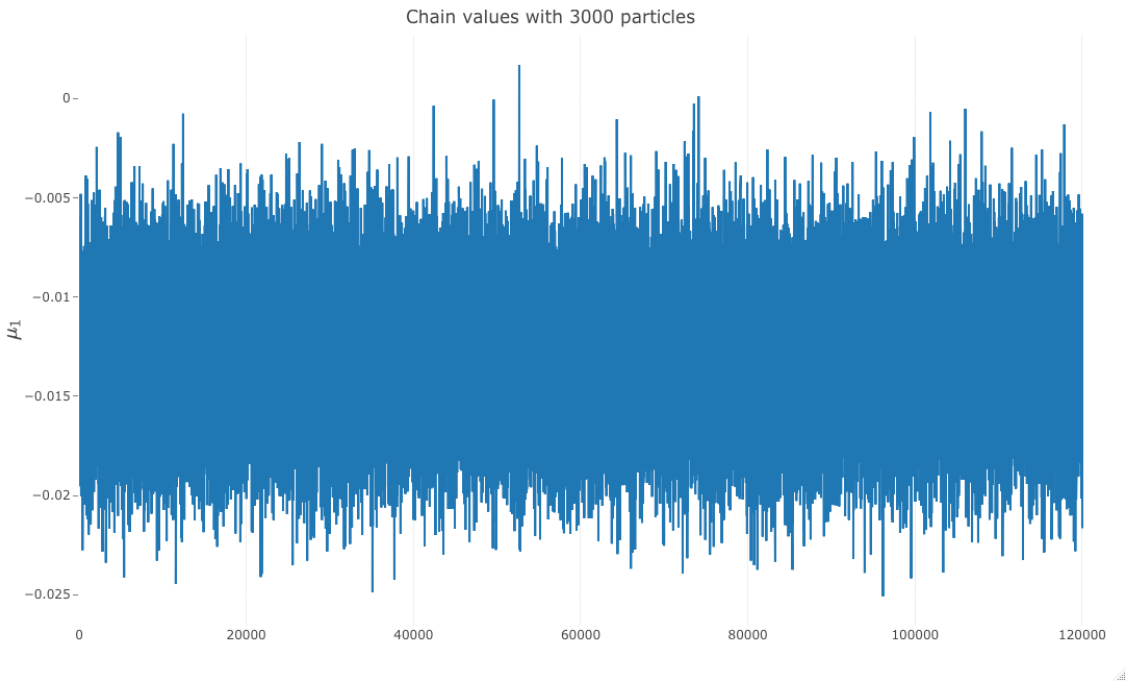

Figure 5.4: Trace plot of μ_1 with 3,000 particles and 120,000 iterations within the CBD-X(3) model.

5.2.2 Estimated Parameters

A few 95% posterior credible intervals for each parameter are shown in Figues 5.5 to 5.8. Credible intervals are an important concept in Bayesian statistics. Their core purpose is to describe and summarize the uncertainty related to the unknown estimated parameters. A credible interval is an interval that contains a value with a certain probability. For instance, a 95% posterior credible intervals for α_x means that α_x is within this interval with 95% probability. For the posterior credible intervals, we assume the mean exists and choose the interval for which the mean is the central point.

Figure 5.5 shows the 95% posterior credible intervals for α_x for all $x \in \{65, \ldots, 89\}$. As discussed in Chapter 2, α_x can be treated as the basic mortality table without any mortality improvement factors. In the plot, we can observe that it increases as age increases. We can also observe that the variance of α_x increases as we add more parameters within the model.

Similar patterns can be observed for $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$: their variance is larger in models with more parameters. Figure 5.6 shows the 95% posterior credible intervals for $\kappa_t^{(1)}$ from 1959 to 2008. Period effect $\kappa_t^{(1)}$ is decreasing with time and more rapidly so in recent years. This may be due to medical improvements and better living conditions. Figures 5.7 and 5.8 show the 95% posterior credible intervals for $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$, respectively, from 1959 to 2008. Period effects $\kappa_t^{(2)}$ and $\kappa_t^{(3)}$ generally have larger variance; consequently, trends are less easy to detect and thus harder to interpret.

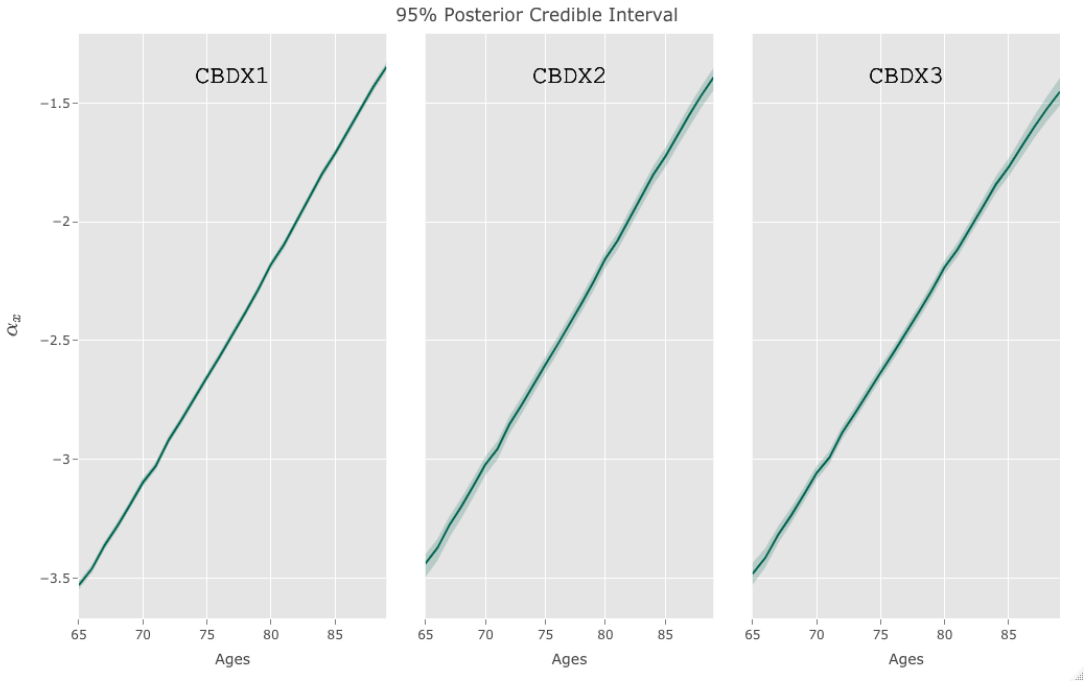

Figure 5.5: 95% posterior credible intervals for α_x for all ages. The solid line is the mean value of α_x.

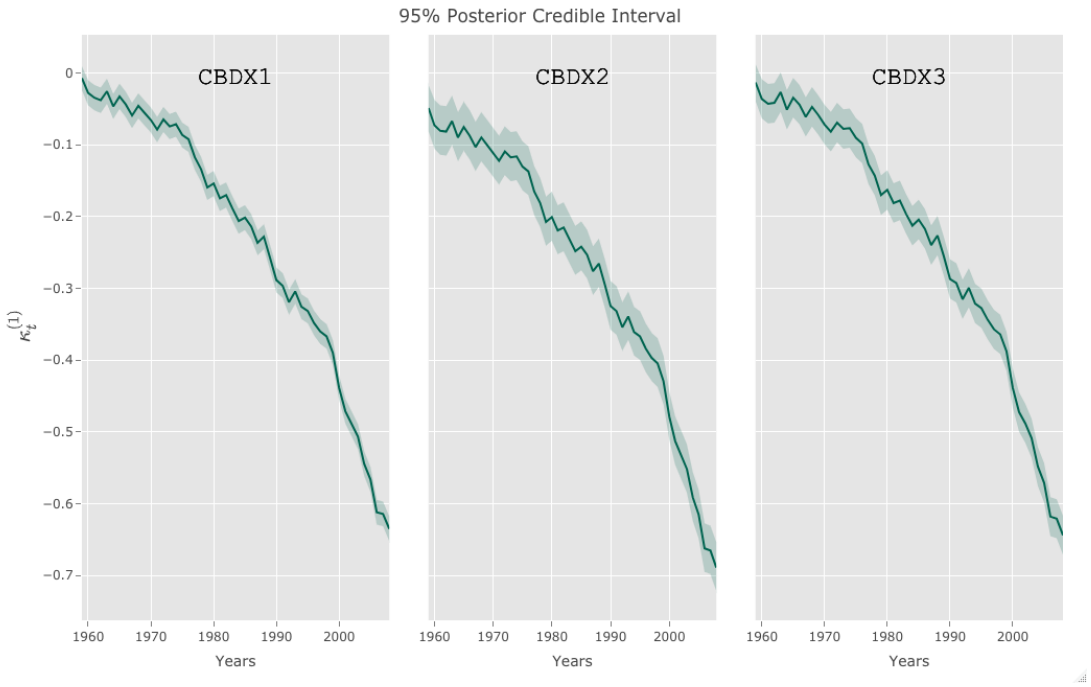

Figure 5.6: 95% posterior credible intervals for $\kappa_t^{(1)}$ from 1959 to 2008. The solid line is the mean value of $\kappa_t^{(1)}$.

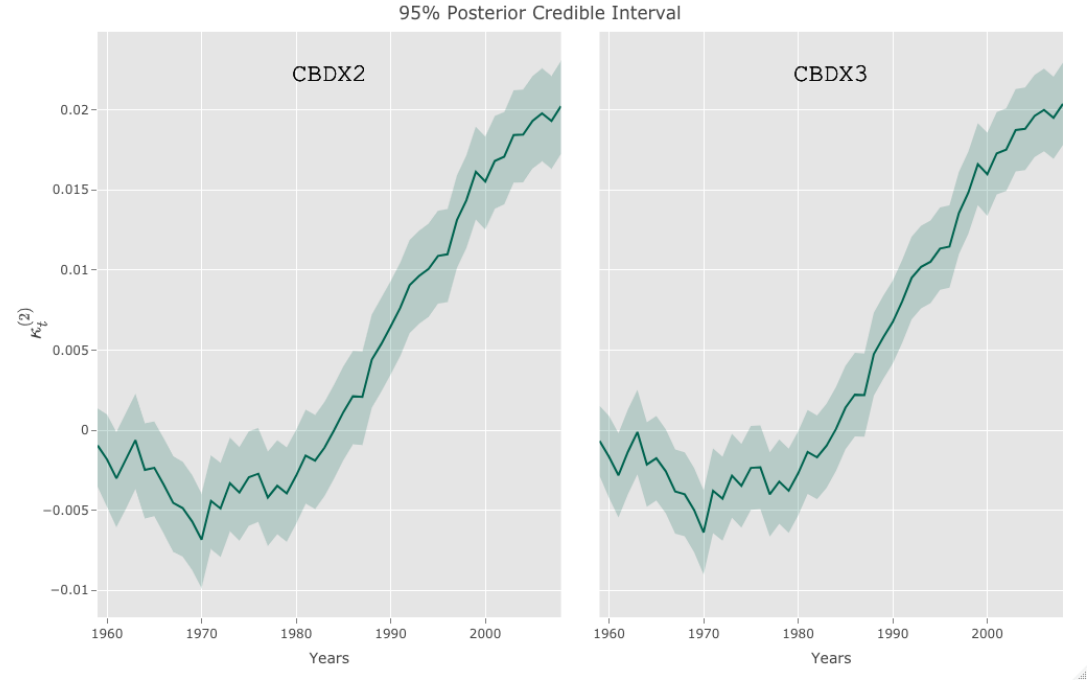

Figure 5.7: 95% posterior credible intervals for $\kappa_t^{(2)}$ from 1959 to 2008. The solid line is the mean value of $\kappa_t^{(2)}$.

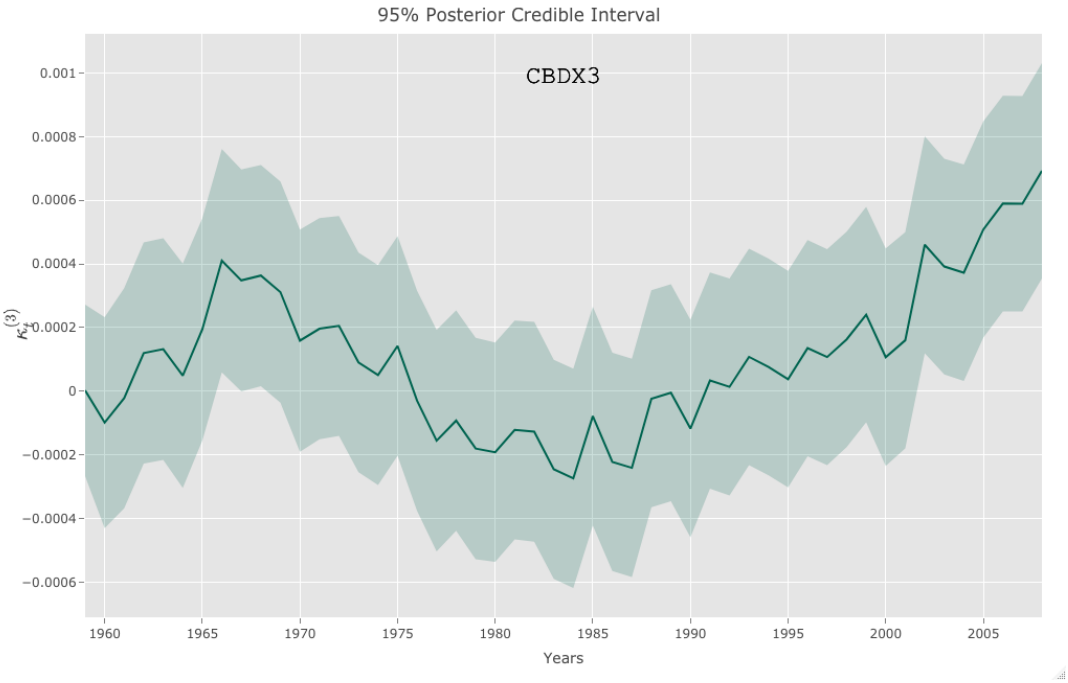

Figure 5.8: 95% posterior credible intervals for $\kappa_t^{(3)}$ from 1959 to 2008. The solid line is the mean value of $\kappa_t^{(3)}$.

5.2.3 Forecasting Death Rates

Once a sufficiently large sample of the posterior distribution of the latent variables $\boldsymbol{\kappa}_t$ and parameters $\boldsymbol{\Theta}$ are simulated, they are used to forecast death rates. We incorporate different sources of uncertainty in forecast intervals by using the following steps. We draw simulated values of the parameters $\boldsymbol{\Theta}$ and the latent variables at the final year (i.e., $\boldsymbol{\kappa}_T$) at random from the pMCMC outputs after discarding the burn-in period. Then, we predict the future period-effect factor $\boldsymbol{\kappa}_{T+h}$, where h are years after T, by using the random walk structure we imposed for the period effect in Equation (2.5). We compute the forecasted death rates $\boldsymbol{m}_{T+h}$ by using the predicted latent variables $\boldsymbol{\kappa}_{T+h}$ and the parameters sampled from the pMCMC output. We repeat this process 10,000 times to get a distribution of the forecasted death rates.

The forecasting performance of the CBD-X models is assessed through the 10-year out-of sample forecasted death rates (i.e., from the year 2009 to the year 2018), that is, $h \in \{1, \ldots, 10\}$. Figures 5.9, 5.10, and 5.11 show the 10-year out-of-sample forecast death rates for the Canadian male population for ages 65, 75, and 85 under the three CBD-X models, respectively. The solid line in each plot is the observed death rate $\hat{m}_{x,t}$ and shaded areas are 95% credible intervals. From the following figures, we can observe that the CBD-X(3) model has the best forecasting performance as the 95% credible intervals capture the observed death rates most of the time.

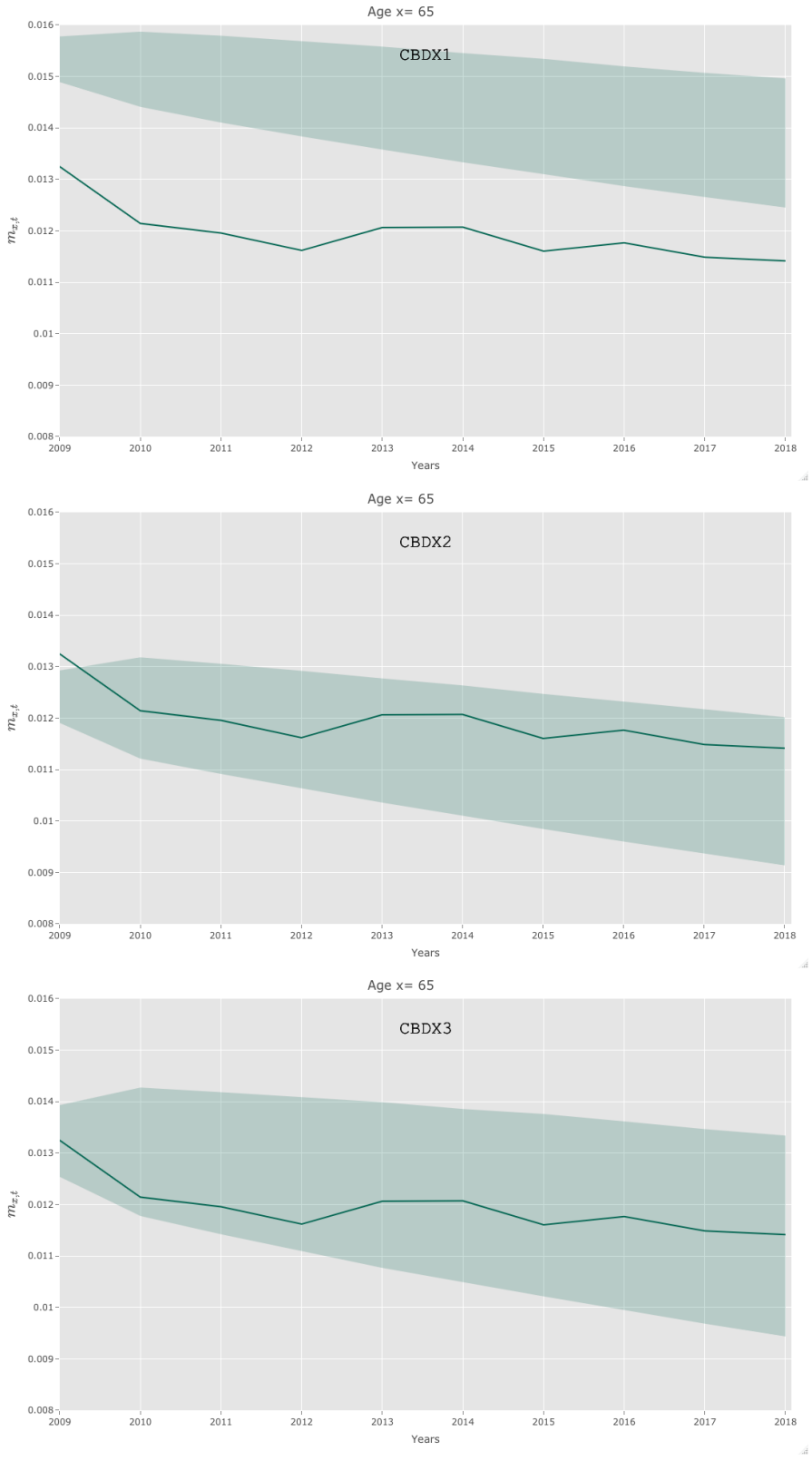

Figure 5.9: 10-year out-of-sample forecasted death rates for the Canadian male population under CBD-X models. The solid line shows the observed $\hat{m}_{x,t}$ for age 65 from 2009 to 2018. Shaded areas are 95% credible intervals.

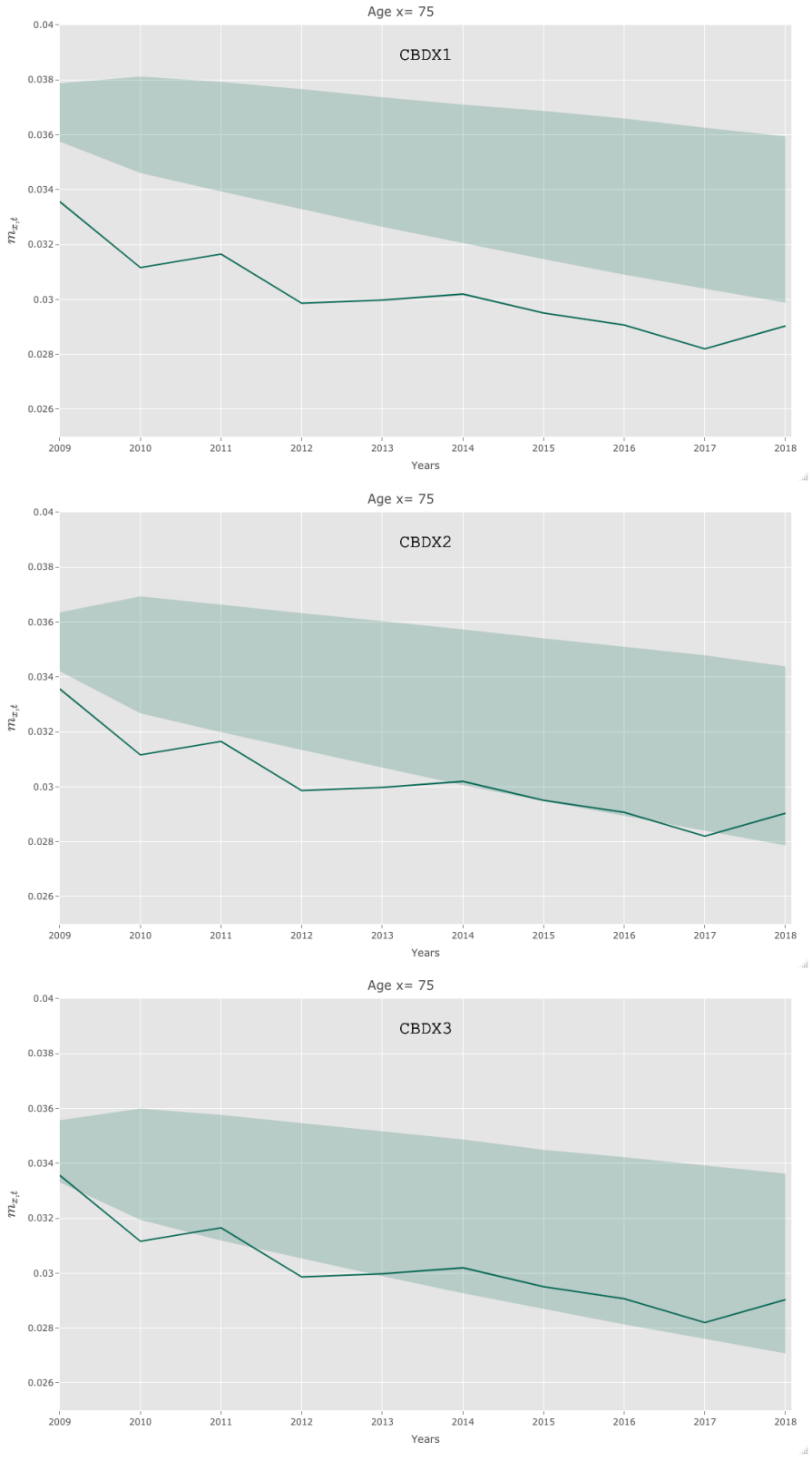

Figure 5.10: 10-year out-of-sample forecasted death rates for the Canadian male population under CBD-X models. The solid line shows the observed $\hat{m}_{x,t}$ for age 75 from 2009 to 2018. Shaded areas are 95% credible intervals.

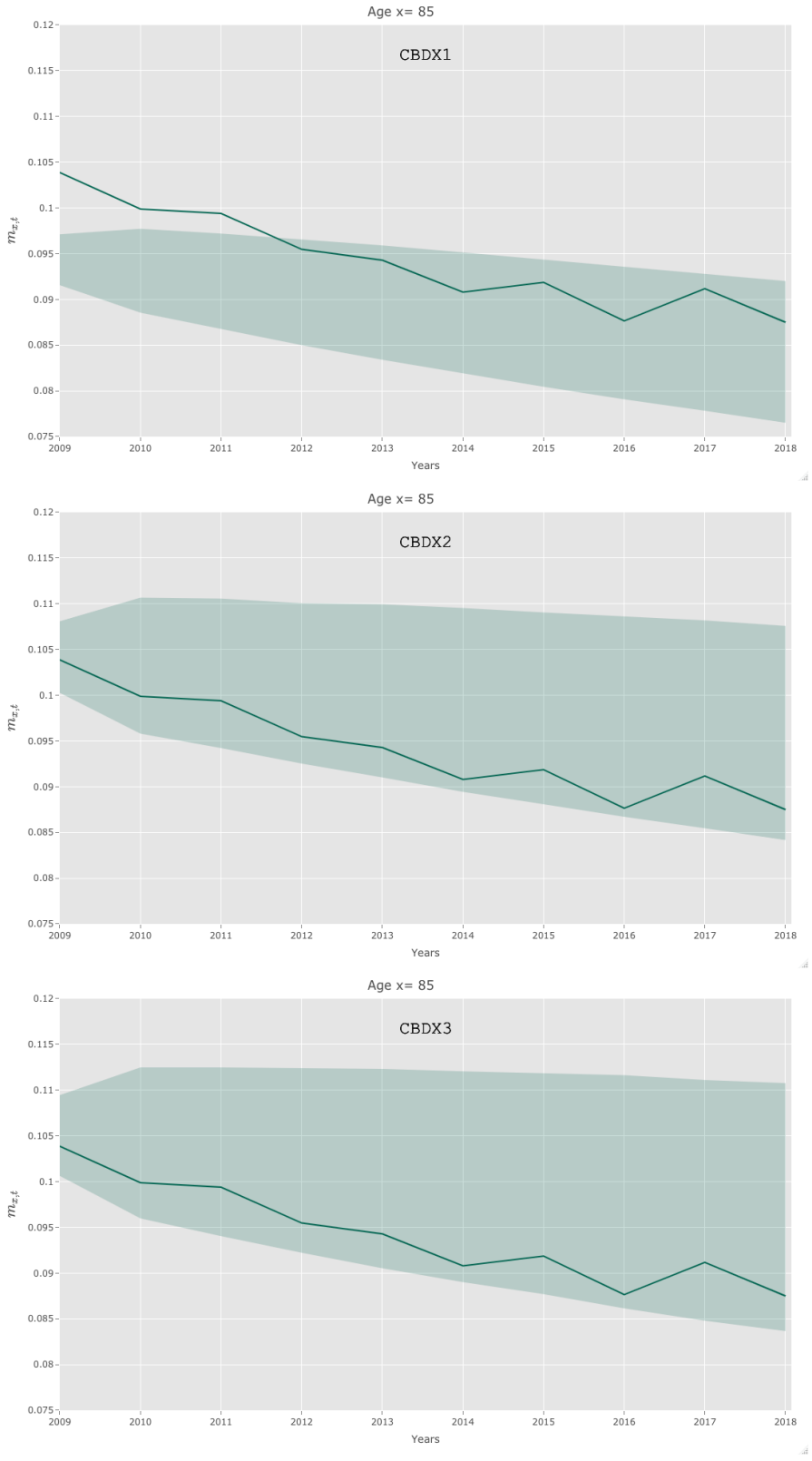

Figure 5.11: 10-year out-of-sample forecasted death rates for the Canadian male population under CBD-X models. The solid line shows the observed $\hat{m}_{x,t}$ for age 85 from 2009 to 2018. Shaded areas are 95% credible intervals.

5.2.4 Model Selection

The problem of model selection requires us to consider two competing notions: one is a measure of model fit that promotes selecting more accurate models, and the other is a measure of model complexity. The deviance information criterion (DIC) proposed by Spiegelhalter et al. (2002) is a useful method in Bayesian model selection. Under the DIC, the goodness of fit is measured by the posterior mean deviance, denoted by $\overline{\text{Deviance}(\boldsymbol{\Psi})}$. The deviance is defined as

$$\text{Deviance}(\boldsymbol{\Psi}) = -2\log\left(\mathcal{L}\left(\boldsymbol{y}_{1:T},\, \boldsymbol{\kappa}_{0:T} \,|\, \boldsymbol{\Theta}\right)\right) + C,$$

where C is a constant that can be canceled out when comparing different models as it is the same for all the models, and the likelihood function comes from Equation (4.1).

The model complexity in DIC is represented by the effective number of parameters of the model being evaluated, denoted by p_D. Spiegelhalter et al. (2002) define p_D as

$$p_D = \overline{\text{Deviance}(\boldsymbol{\Psi})} - \text{Deviance}(\bar{\boldsymbol{\Psi}}),$$

where $\bar{\boldsymbol{\Psi}}$ is a point estimate of parameters $\boldsymbol{\Psi}$. Although the posterior mean of the parameters is often used as the point estimate, other values can be substituted such as the median or the mode.

Gelman et al. (2004) propose an alternative way of calculating p_D. In their version, p_D is estimated as

$$p_D = \frac{1}{2}\text{Var}\left(\text{Deviance}(\boldsymbol{\Psi})\right),$$

where $\text{Var}\left(\text{Deviance}(\boldsymbol{\Psi})\right)$ is the overall variance of $\text{Deviance}(\boldsymbol{\Psi})$ and thus, p_D will not depend on the point estimate.

We use the method of Gelman et al. (2004) to calculate the effective number of parameters p_D. Combining the two components discussed above, we obtain the DIC as:

$$\begin{aligned} DIC &= p_D + \overline{\text{Deviance}(\boldsymbol{\Psi})} \\ &= \frac{1}{2}\text{Var}\left(\text{Deviance}(\boldsymbol{\Psi})\right) + \overline{\text{Deviance}(\boldsymbol{\Psi})}. \end{aligned}$$

The value of $\overline{\text{Deviance}(\boldsymbol{\Psi})}$ decreases as the number of parameters in a model increases, and models with smaller DIC values should be preferred.

The DIC values for the three CBD-X models are shown in Table 5.2. The CBD-X(3) is the best model as it has the smallest DIC value.

Model	DIC	p_D	Posterior Mean Deviance
CBD-X(1)	12,547.210	87.718	12,459.490
CBD-X(2)	1,563.776	184.093	1,392.704
CBD-X(3)	766.372	387.507	378.864

Table 5.2: Deviance information criterion for CBD-X models.

Chapter 6

Forecasting Performance In Different Estimation Methods

One of the advantages of using the Bayesian method is that we can incorporate all the uncertainty within our model estimation and forecasting. This chapter investigates the impact of including the uncertainty by comparing the forecasting performance of the CBD-X(3) model estimated under the maximum likelihood method to that estimated under the pMCMC method. We only compare the performance in the CBD-X(3) model because it has the best performance in both model selection and forecasting, as shown in the previous chapter.

The maximum likelihood method in this chapter is different from the two-stage maximum likelihood estimation method we introduced at the beginning of Chapter 3. Instead of estimating all the parameters (i.e., $\boldsymbol{\Theta}$ and $\boldsymbol{\kappa}_{0:T}$) by using a frequentist method as in the two-stage maximum likelihood estimation method, the maximum likelihood method in this chapter is only used to estimate the unknown parameters $\boldsymbol{\Theta}$ by means of the PF. That is to say, the mortality model is still in a state-space representation as in Chapter 3 but instead of applying the Bayesian method to estimate the unknown parameters $\boldsymbol{\Theta}$, we use a frequentist method and specifically, the MLE. First, we use the mean of the posterior distribution of $\boldsymbol{\Theta}$ from pMCMC as the starting value to maximize the likelihood function obtained in the SMC algorithm (i.e., Equation (3.5)) of the CBD-X(3) model by using the Nelder and Mead (1965) method. Then, the latent variables $\boldsymbol{\kappa}_{0:T}$ are estimated from the SMC algorithm with the optimal value of $\boldsymbol{\Theta}$ obtained from MLE.

Similar to the forecasting algorithm in the pMCMC estimation introduced in Chapter 5, the latent variables $\boldsymbol{\kappa}_{T+h}$ are predicted using the random walk model for the period effect; that is, Equation (2.5). Then, we compute the forecasted death rate $\boldsymbol{m}_{T+h}$ by using the predicted latent variables $\boldsymbol{\kappa}_{T+h}$ and the constant parameters $\boldsymbol{\Theta}$ obtained by MLE. We repeat this process 10,000 times to get a distribution of the forecasted death rate.

The 95% confidence intervals under MLE and the 95% credible intervals under pMCMC of the 10-year out-of-sample forecasted death rates for the Canadian male population at

ages 65, 70, 75, 80, 85, and 89 are shown in Figures 6.1 to 6.6. Overall, the 95% confidence intervals under MLE are narrower than the 95% credible intervals under pMCMC. This means, if we estimate mortality models with the frequentist method, there is a part of uncertainty that will be missed, which may have a negative financial impacts on certain companies.

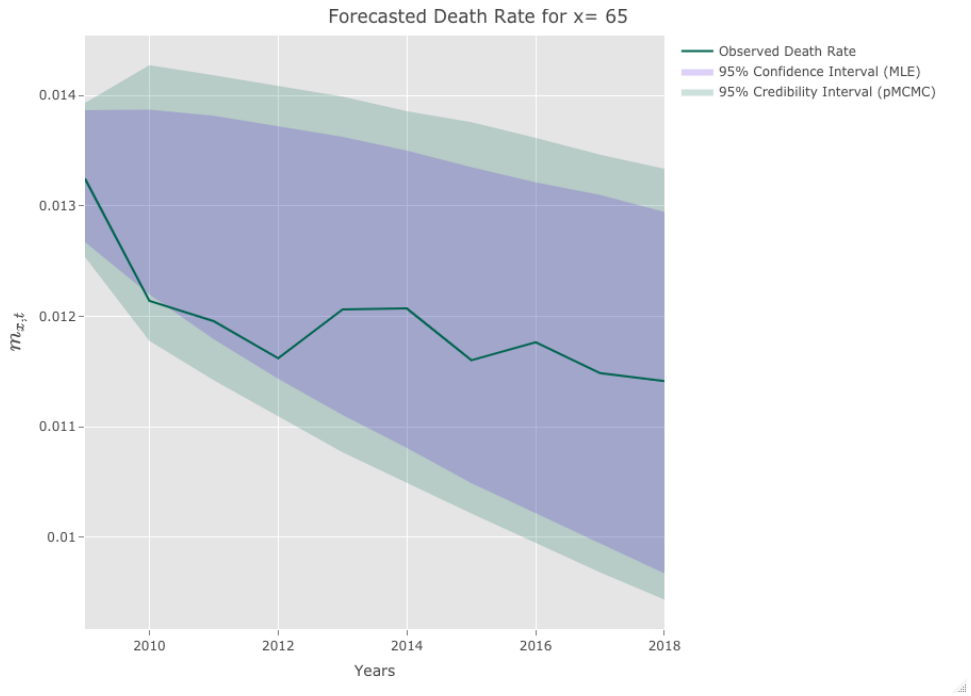

Figure 6.1: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 65 from 2009 to 2018.

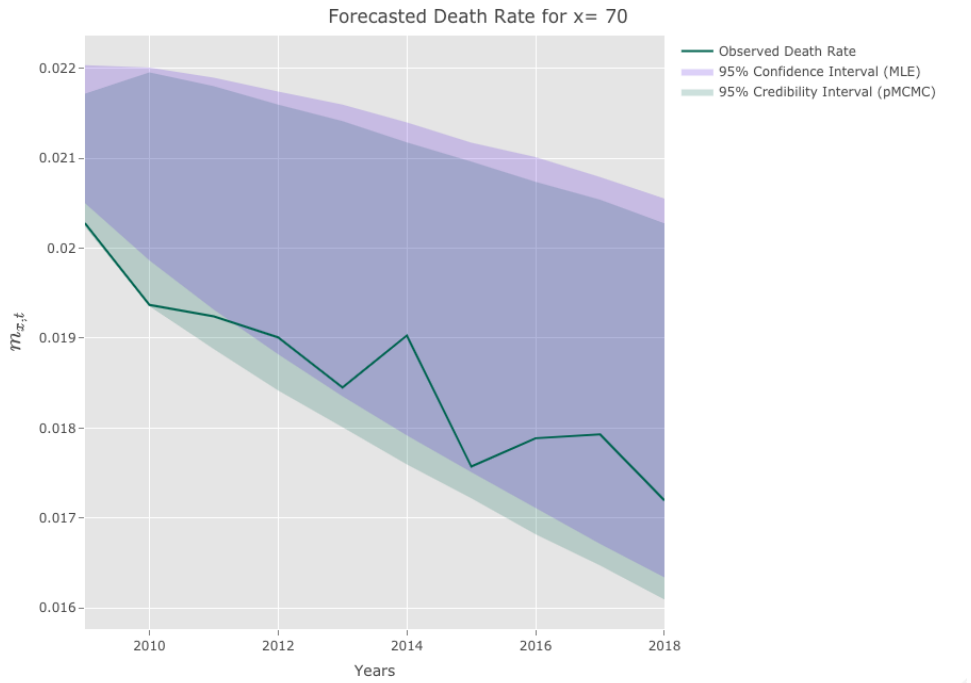

Figure 6.2: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 70 from 2009 to 2018.

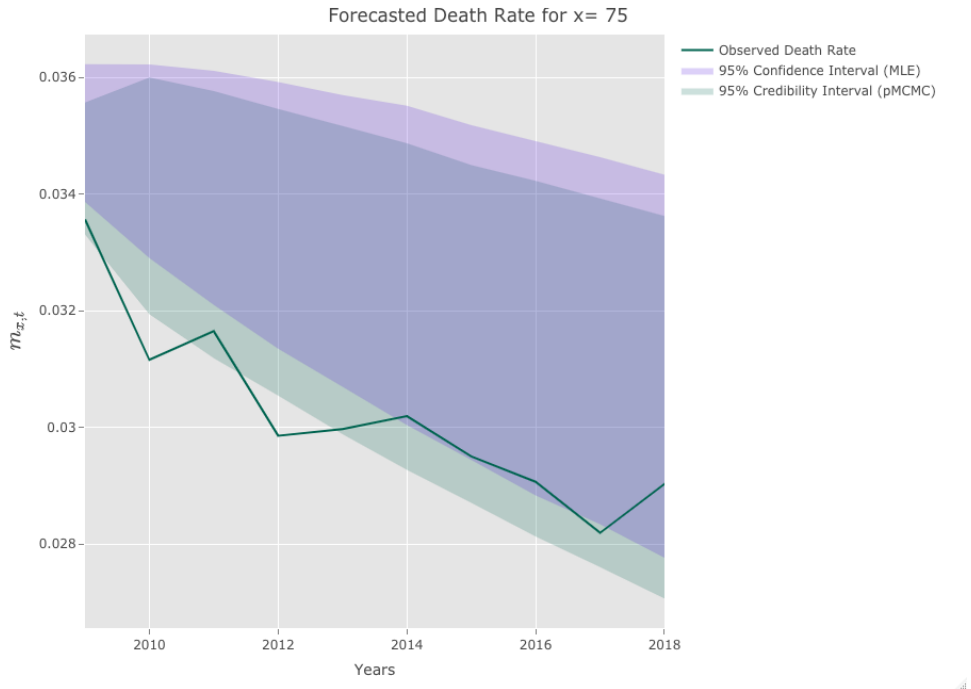

Figure 6.3: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 75 from 2009 to 2018.

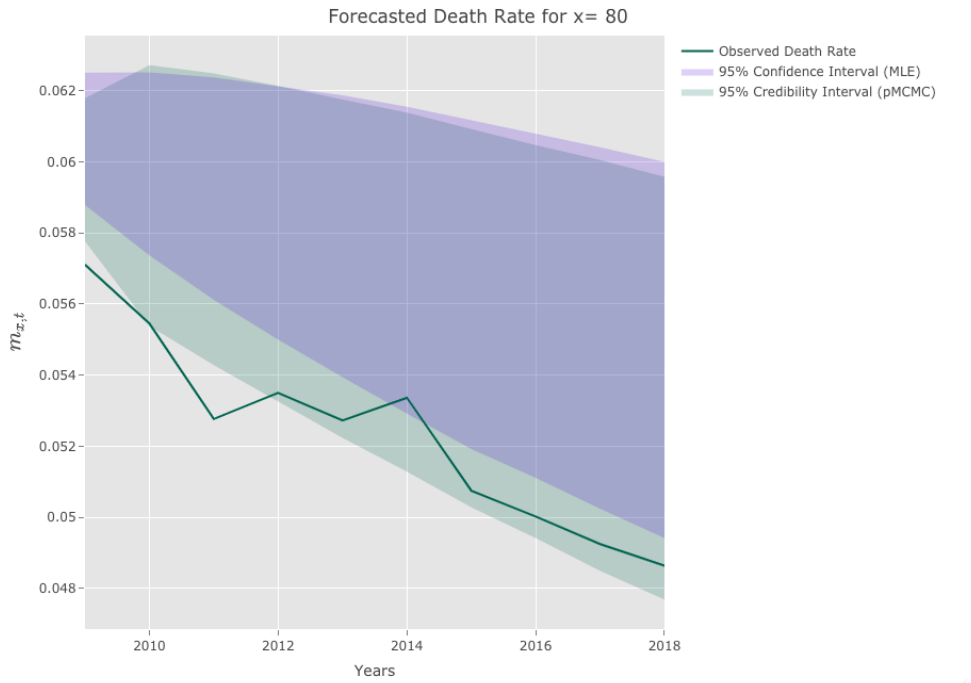

Figure 6.4: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 80 from 2009 to 2018.

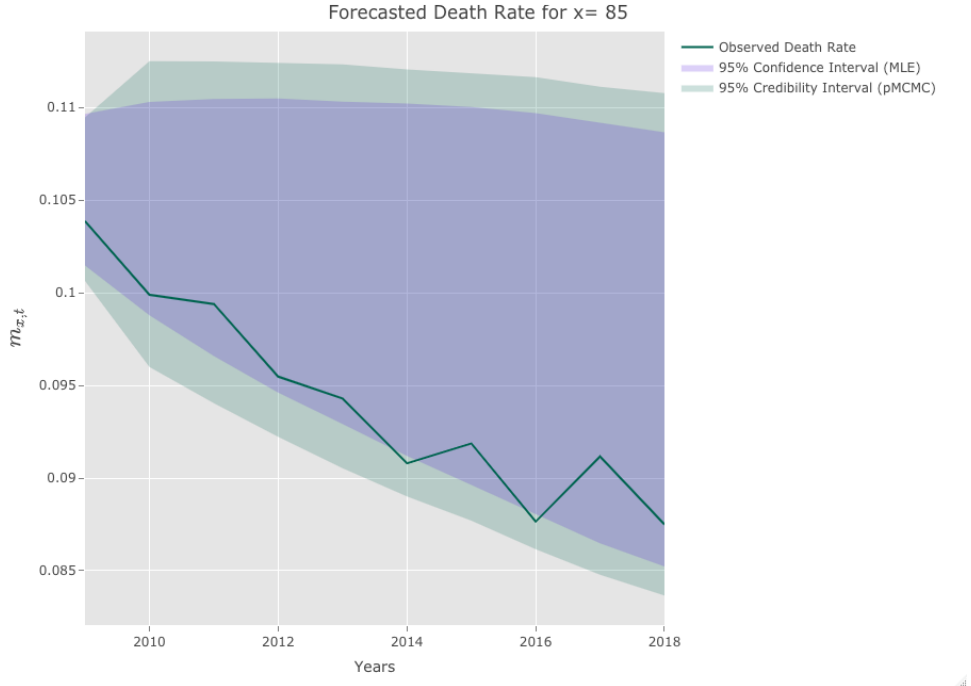

Figure 6.5: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 85 from 2009 to 2018.

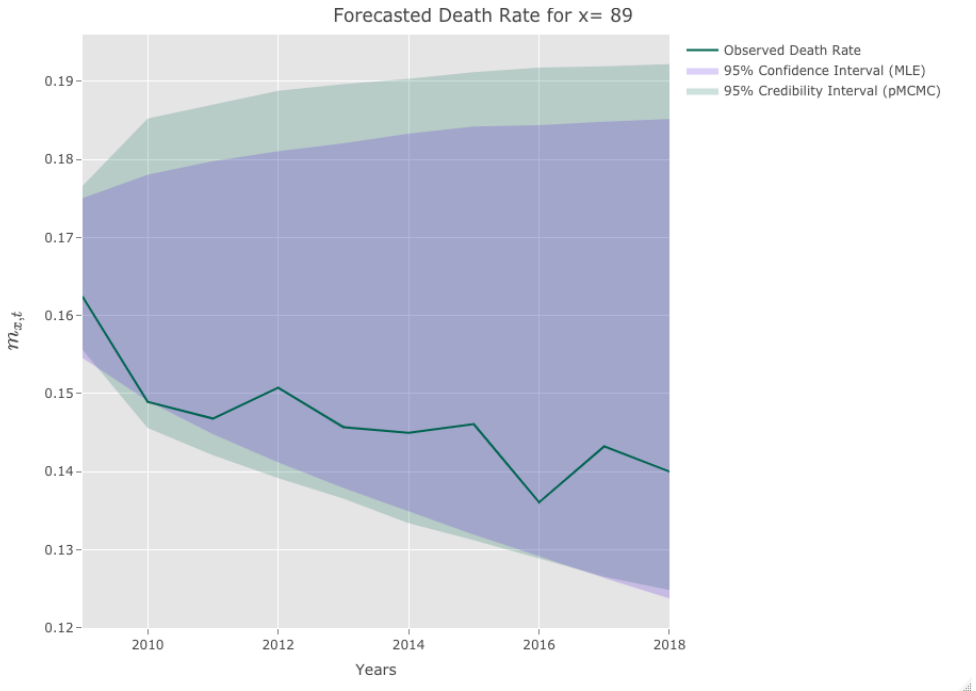

Figure 6.6: 10-year out-of-sample forecasted death rates for the Canadian male population under MLE method (shaded purple area) and pMCMC method (shaded green area). The solid line shows the observed $\hat{m}_{x,t}$ for age 89 from 2009 to 2018.

Chapter 7

Conclusion

This report presented a Bayesian method—specifically the pMCMC method—to estimate CBD-X models combining features of the LC and CBD models. By using this method, the dynamics of the period effect $\boldsymbol{\kappa}_{0:T}$ can be incorporated within the model estimation. Also, parameter uncertainty can be obtained readily and used in calculating forecasting intervals.

First, the CBD-X model was recast into state-space formulation. By doing so, we were able to incorporate the dynamics of period effects into the model estimation. Although the dynamics of the period effects in our models are linear and Gaussian, we still applied the SMC method to allow for more flexibility in choosing the structure of period effects in the future.

Second, a Bayesian approach was used to estimate the unknown parameters $\boldsymbol{\Theta}$ and their uncertainty, as Bayesian approaches yield posterior distributions of model parameters as well as mortality rates. A sampling-based approach called MCMC was used because we cannot derive the closed-from solution of joint posterior distributions for unknown parameters. The models were estimated based on Canadian male mortality data.

After fitting the models, we were able to perform model comparison and predictions. The CBD-X(3) model is the best model as it leads to the smallest DIC values and the best forecasting performance. Its 95% credibility intervals of the 10-year out-of-sample forecasted death rate capture the observed death rates in most cases.

To assess how including the parameter uncertainty influenced the forecasting performance, we compared the 95% confidence intervals obtained with pMCMC to those obtained with MLE for 10-year out-of-sample forecasted death rates. We observed that the 95% confidence intervals are generally smaller than the 95% credibility intervals. Thus, in contrast to frequentist estimation methods, the Bayesian approach captures more uncertainty in forecasting—consistently with the fact that mortality models' parameter uncertainty is large.

The Bayesian-based estimation approach proposed in this report is flexible and easy to implement as it can be applied to other mortality models. For future extensions, one possible direction would be applying this estimation approach for multi-population mortality

models. Due to globalization, different populations' mortality are closely linked together. Therefore, developing multi-population mortality models to analyze and forecast the mortality of more than one population in a joint way, such as modelling population between different countries, is of paramount importance. Another possible extension of our work would include the addition of cohort effects within model estimation as it might improve the model performance. Changing the structure of period effects might also be another interesting avenue for future research.

Bibliography

Andrieu, C., A. Doucet, and R. Holenstein. 2010. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72:269–342.

Brouhns, N., M. Denuit, and J. K. Vermunt. 2002. A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics* 31:373–393.

Cairns, A. J. 2011. Modelling and management of longevity risk: Approximations to survivor functions and dynamic hedging. *Insurance: Mathematics and Economics* 49:438–453.

Cairns, A. J. G., D. Blake, and K. Dowd. 2006. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance* 73:687–718.

Cairns, A. J. G., D. Blake, K. Dowd, G. D. Coughlan, and M. Khalaf-Allah. 2011. Bayesian stochastic mortality modelling for two populations. *ASTIN Bulletin* 41:29–59.

Cairns, A. J. G., D. Blake, K. Dowd, G. D.Coughlan, D. Epstein, A. Ong, and I. Balevich. 2009. A quantitative comparison of stochastic mortality models using data from England and Wales and the United States. *North American Actuarial Journal* 13:1–35.

Cairns, A. J. G., D. Blake, K. Dowd, and A. R. Kessler. 2016. Phantoms never die: Living with unreliable population data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179:975–1005.

Cairns, A. J. G., M. Kallestrup-Lamb, C. Rosenskjold, D. Blake, and K. Dowd. 2019. Modelling socio-economic differences in mortality using a new affluence index. *ASTIN Bulletin* 49:555–590.

Chan, W., J. S. Li, and J. Li. 2014. The CBD mortality indexes: Modeling and applications. *North American Actuarial Journal* 18:38–58.

Currie, S. J. Richards, J. G. Kirkby, and I. D. 2006. The importance of year of birth in two-dimensional mortality data. *British Actuarial Journal* 12:5–38.

Czado, C., A. Delwarde, and M. Denuit. 2005. Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics* 36:260–284.

Dowd, K., Cairns, A. J. G., and D. Blake. 2020. CBDX: A workhorse mortality model from the Cairns-Blake-Dowd family. *Annals of Actuarial Science* 14:445–460.

Fung, M. C., G. W. Peters, and P. V. Shevchenko. 2017. A unified approach to mortality modelling using state-space framework: Characterisation, identification, estimation and forecasting. *Annals of Actuarial Science* 11:343–389.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC.

Gelman, A., and D. B. Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457 – 472.

Georgios, M., Cairns, A. J. G., S. George, and K. Torsten. 2017. Stochastic mortality modelling: Key drivers and dependent residuals. *North American Actuarial Journal* 21:1–26.

Gompertz, B. 1825. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London* 115:513–583.

Haqqi Anna Zili, A., S. Mardiyati, and D. Lestari. 2018. Forecasting Indonesian mortality rates using the Lee-Carter model and ARIMA method. *AIP Conference Proceedings* 2023:020212.

Hunt, A., and D. Blake. 2020. Identifiability in age/period mortality models. *Annals of Actuarial Science* 14:461–499.

Hunt, A., and D. Blake. 2021. On the structure and classification of mortality models. *North American Actuarial Journal* 25:215–234.

Kamaruddin, H., and I. Noriszura. 2018. Forecasting selected specific age mortality rate of Malaysia by using Lee-Carter model. *Journal of Physics: Conference Series* 974:012003.

Koissi, M.-C., A. F. Shapiro, and G. Högnäs. 2006. Evaluating and extending the Lee–Carter model for mortality forecasting: Bootstrap confidence interval. *Insurance: Mathematics and Economics* 38:1–20.

Lee, R. D., and L. R. Carter. 1992. Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87:659–671.

Lundström, H., and J. Qvist. 2004. Mortality forecasting and trend shifts: An application of the Lee-Carter Model to Swedish mortality data. *International Statistical Review* 72:37–50.

Nelder, J. A., and R. Mead. 1965. A simplex method for function minimization. *The Computer Journal* 7:308–313.

Pedroza, C. 2006. A Bayesian forecasting model: Predicting U.S. male mortality. *Biostatistics* 7:530–550.

Plat, R. 2009. On stochastic mortality modeling. *Insurance: Mathematics and Economics* 45:393–404.

Reichmuth, W. H., and S. Sarferaz. 2008. Bayesian demographic modeling and forecasting: An application to U.S. mortality. *Working paper*.

Renshaw, A. E., and S. Haberman. 2006. A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics* 38:556–570.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:583–639.

Appendix A

Distribution of the Error Term

The standard actuarial approach in mortality modelling assumes that, conditional on the model central death rate $m_{x,t}$ and the observed exposures $\hat{E}_{x,t}$, the observed number of deaths $\hat{D}_{x,t}$ has a Poisson distribution with mean and variance both equal to $m_{x,t}\hat{E}_{x,t}$.[1] Cairns et al. (2016) assumes that for each t and x, the observed number of deaths $\hat{D}_{x,t}$ is conditionally independent and has a lognormal distribution; that is, $\log(\hat{D}_{x,t})$ has a normal distribution with mean μ_d and variance σ_d^2. The lognormal distribution is chosen for computational convenience. As in Cairns et al. (2016), we equate the mean and variance of the lognormal distribution to the mean and variance of a matching Poisson distribution. By matching the mean, we obtain:

$$m_{x,t}\hat{E}_{x,t} = \exp\left(\mu_d + \frac{\sigma_d^2}{2}\right),$$

which is equivalent to

$$\mu_d = \log(m_{x,t}\hat{E}_{x,t}) - \frac{\sigma_d^2}{2}.$$

By matching the variance, on the other hand, we have

$$m_{x,t}\hat{E}_{x,t} = \left(\exp(\sigma_d^2) - 1\right)\exp\left(2\mu_d + \sigma_d^2\right),$$

so

$$\begin{aligned}\left(\exp(\sigma_d^2) - 1\right) &= \frac{m_{x,t}\hat{E}_{x,t}}{\exp\left(2\log(m_{x,t}\hat{E}_{x,t}) - \sigma_d^2 + \sigma_d^2\right)} \\ &= \frac{1}{m_{x,t}\hat{E}_{x,t}} = \frac{1}{m_{x,t}E_{x,t}}\end{aligned}$$

[1] In this report, the observed and theoretical exposures are assumed to be the same, i.e., $E_{x,t} = \hat{E}_{x,t}$

$$= \frac{1}{D_{x,t}}.$$

By using the first-order Taylor series expansion about zero, we have that:

$$\exp(x) - 1 \approx x,$$

thus, $\sigma_d^2 \approx \frac{1}{D_{x,t}}$. Since $D_{x,t}$ is unobserved, we use the observed number of deaths $\hat{D}_{x,t}$ to approximate it, meaning that $\sigma_d^2 \approx \frac{1}{\hat{D}_{x,t}}$. This approximation is accurate if $\hat{D}_{x,t}$ is large. Also, assuming that $\hat{D}_{x,t}$ is large enough to make the above approximation accurate, then the term $\frac{\sigma_d^2}{2}$ in μ_d will be smaller enough to ignore it, as done in Cairns et al. (2019) .

We follow the same logic and assume the estimated central death rate $\hat{m}_{x,t}$ is lognormal. Hence,

$$\log(\hat{D}_{x,t}) = \log(\hat{m}_{x,t}\,\hat{E}_{x,t}) \sim \mathcal{N}\left(\log(m_{x,t}\,E_{x,t}), \frac{1}{\hat{D}_{x,t}}\right),$$

which implies that

$$\log(\hat{m}_{x,t}) \sim \mathcal{N}\left(\log(m_{x,t}), \frac{1}{\hat{D}_{x,t}}\right).$$

Appendix B

Summary Tables

The Gelman-Rubin test for each CBD-X models are summarized in the following tables.

CBD-X(1) Model	Mean	Standard Deviation	Gelman-Rubin Test
α_1	-3.529	0.012	1.014
α_2	-3.461	0.012	1.014
α_3	-3.362	0.012	1.015
α_4	-3.280	0.012	1.014
α_5	-3.190	0.011	1.014
α_6	-3.096	0.011	1.015
α_7	-3.027	0.011	1.015
α_8	-2.919	0.011	1.014
α_9	-2.834	0.011	1.014
α_{10}	-2.741	0.011	1.014
α_{11}	-2.653	0.011	1.014
α_{12}	-2.565	0.011	1.014
α_{13}	-2.476	0.011	1.015
α_{14}	-2.381	0.011	1.015
α_{15}	-2.284	0.011	1.015
α_{16}	-2.178	0.011	1.014
α_{17}	-2.097	0.011	1.015
α_{18}	-1.995	0.011	1.014
α_{19}	-1.898	0.012	1.014
α_{20}	-1.796	0.012	1.014

CBD-X(1) Model	Mean	Standard Deviation	Gelman-Rubin Test
α_{21}	-1.712	0.012	1.014
α_{22}	-1.616	0.012	1.014
α_{23}	-1.520	0.012	1.014
α_{24}	-1.431	0.012	1.014
α_{25}	-1.348	0.012	1.014
$\kappa_1^{(1)}$	-0.008	0.011	1.014
$\kappa_2^{(1)}$	-0.029	0.012	1.013
$\kappa_3^{(1)}$	-0.035	0.012	1.013
$\kappa_4^{(1)}$	-0.039	0.012	1.013
$\kappa_5^{(1)}$	-0.027	0.012	1.014
$\kappa_6^{(1)}$	-0.048	0.012	1.013
$\kappa_7^{(1)}$	-0.034	0.012	1.014
$\kappa_8^{(1)}$	-0.045	0.012	1.014
$\kappa_9^{(1)}$	-0.060	0.012	1.014
$\kappa_{10}^{(1)}$	-0.047	0.012	1.013
$\kappa_{11}^{(1)}$	-0.057	0.012	1.014
$\kappa_{12}^{(1)}$	-0.067	0.012	1.014
$\kappa_{13}^{(1)}$	-0.080	0.012	1.013
$\kappa_{14}^{(1)}$	-0.066	0.012	1.014
$\kappa_{15}^{(1)}$	-0.075	0.012	1.013
$\kappa_{16}^{(1)}$	-0.072	0.012	1.014
$\kappa_{17}^{(1)}$	-0.087	0.012	1.014
$\kappa_{18}^{(1)}$	-0.093	0.012	1.014
$\kappa_{19}^{(1)}$	-0.118	0.012	1.013
$\kappa_{20}^{(1)}$	-0.135	0.012	1.014
$\kappa_{21}^{(1)}$	-0.160	0.012	1.014
$\kappa_{22}^{(1)}$	-0.155	0.012	1.014
$\kappa_{23}^{(1)}$	-0.176	0.012	1.014
$\kappa_{24}^{(1)}$	-0.171	0.012	1.014
$\kappa_{25}^{(1)}$	-0.190	0.012	1.014
$\kappa_{26}^{(1)}$	-0.207	0.012	1.013
$\kappa_{27}^{(1)}$	-0.203	0.012	1.014
$\kappa_{28}^{(1)}$	-0.215	0.012	1.014

CBD-X(1) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{29}^{(1)}$	-0.238	0.012	1.014
$\kappa_{30}^{(1)}$	-0.229	0.012	1.014
$\kappa_{31}^{(1)}$	-0.259	0.012	1.013
$\kappa_{32}^{(1)}$	-0.290	0.012	1.014
$\kappa_{33}^{(1)}$	-0.297	0.012	1.014
$\kappa_{34}^{(1)}$	-0.320	0.012	1.013
$\kappa_{35}^{(1)}$	-0.305	0.012	1.014
$\kappa_{36}^{(1)}$	-0.327	0.012	1.014
$\kappa_{37}^{(1)}$	-0.333	0.012	1.014
$\kappa_{38}^{(1)}$	-0.349	0.012	1.014
$\kappa_{39}^{(1)}$	-0.361	0.012	1.014
$\kappa_{40}^{(1)}$	-0.368	0.012	1.014
$\kappa_{41}^{(1)}$	-0.391	0.012	1.014
$\kappa_{42}^{(1)}$	-0.440	0.012	1.014
$\kappa_{43}^{(1)}$	-0.472	0.012	1.014
$\kappa_{44}^{(1)}$	-0.490	0.012	1.014
$\kappa_{45}^{(1)}$	-0.508	0.012	1.014
$\kappa_{46}^{(1)}$	-0.546	0.012	1.014
$\kappa_{47}^{(1)}$	-0.567	0.012	1.014
$\kappa_{48}^{(1)}$	-0.613	0.012	1.014
$\kappa_{49}^{(1)}$	-0.615	0.012	1.014
$\kappa_{50}^{(1)}$	-0.636	0.012	1.014
μ_1	-0.013	0.003	1.000
v_{11}	0.000	0.001	1.000

CBD-X(2) Model	Mean	Standard Deviation	Gelman-Rubin Test
α_1	-3.438	0.025	1.000
α_2	-3.372	0.024	1.000
α_3	-3.276	0.023	1.000
α_4	-3.198	0.022	1.000
α_5	-3.111	0.021	1.000
α_6	-3.021	0.021	1.000
α_7	-2.956	0.020	1.000
α_8	-2.852	0.019	1.000
α_9	-2.772	0.019	1.000
α_{10}	-2.683	0.018	1.000
α_{11}	-2.600	0.018	1.000
α_{12}	-2.518	0.018	1.000
α_{13}	-2.434	0.018	1.000
α_{14}	-2.345	0.018	1.000
α_{15}	-2.254	0.018	1.000
α_{16}	-2.155	0.018	1.000
α_{17}	-2.080	0.018	1.000
α_{18}	-1.985	0.018	1.000
α_{19}	-1.895	0.019	1.000
α_{20}	-1.801	0.019	1.000
α_{21}	-1.725	0.020	1.000
α_{22}	-1.636	0.020	1.000
α_{23}	-1.549	0.021	1.000
α_{24}	-1.467	0.022	1.000
α_{25}	-1.393	0.023	1.000
$\kappa_1^{(1)}$	-0.049	0.017	1.000
$\kappa_2^{(1)}$	-0.073	0.018	1.000
$\kappa_3^{(1)}$	-0.081	0.018	1.000
$\kappa_4^{(1)}$	-0.082	0.018	1.000
$\kappa_5^{(1)}$	-0.068	0.018	1.000
$\kappa_6^{(1)}$	-0.090	0.018	1.000
$\kappa_7^{(1)}$	-0.076	0.018	1.000
$\kappa_8^{(1)}$	-0.088	0.018	1.000

CBD-X(2) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_9^{(1)}$	-0.104	0.018	1.000
$\kappa_{10}^{(1)}$	-0.091	0.018	1.000
$\kappa_{11}^{(1)}$	-0.101	0.018	1.000
$\kappa_{12}^{(1)}$	-0.113	0.018	1.000
$\kappa_{13}^{(1)}$	-0.123	0.018	1.000
$\kappa_{14}^{(1)}$	-0.110	0.018	1.000
$\kappa_{15}^{(1)}$	-0.118	0.018	1.000
$\kappa_{16}^{(1)}$	-0.117	0.018	1.000
$\kappa_{17}^{(1)}$	-0.131	0.018	1.000
$\kappa_{18}^{(1)}$	-0.138	0.018	1.000
$\kappa_{19}^{(1)}$	-0.166	0.018	1.000
$\kappa_{20}^{(1)}$	-0.182	0.018	1.000
$\kappa_{21}^{(1)}$	-0.208	0.018	1.000
$\kappa_{22}^{(1)}$	-0.201	0.018	1.000
$\kappa_{23}^{(1)}$	-0.220	0.018	1.000
$\kappa_{24}^{(1)}$	-0.216	0.018	1.000
$\kappa_{25}^{(1)}$	-0.233	0.018	1.000
$\kappa_{26}^{(1)}$	-0.249	0.018	1.000
$\kappa_{27}^{(1)}$	-0.243	0.018	1.000
$\kappa_{28}^{(1)}$	-0.254	0.018	1.000
$\kappa_{29}^{(1)}$	-0.277	0.018	1.000
$\kappa_{30}^{(1)}$	-0.266	0.018	1.000
$\kappa_{31}^{(1)}$	-0.295	0.018	1.000
$\kappa_{32}^{(1)}$	-0.325	0.018	1.000
$\kappa_{33}^{(1)}$	-0.333	0.018	1.000
$\kappa_{34}^{(1)}$	-0.355	0.018	1.000
$\kappa_{35}^{(1)}$	-0.340	0.018	1.000
$\kappa_{36}^{(1)}$	-0.362	0.018	1.000
$\kappa_{37}^{(1)}$	-0.368	0.018	1.000
$\kappa_{38}^{(1)}$	-0.385	0.018	1.000
$\kappa_{39}^{(1)}$	-0.398	0.018	1.000
$\kappa_{40}^{(1)}$	-0.405	0.018	1.000
$\kappa_{41}^{(1)}$	-0.430	0.018	1.000

CBD-X(2) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{42}^{(1)}$	-0.479	0.018	1.000
$\kappa_{43}^{(1)}$	-0.513	0.018	1.000
$\kappa_{44}^{(1)}$	-0.532	0.018	1.000
$\kappa_{45}^{(1)}$	-0.553	0.018	1.000
$\kappa_{46}^{(1)}$	-0.592	0.018	1.000
$\kappa_{47}^{(1)}$	-0.616	0.018	1.000
$\kappa_{48}^{(1)}$	-0.663	0.018	1.000
$\kappa_{49}^{(1)}$	-0.666	0.018	1.000
$\kappa_{50}^{(1)}$	-0.690	0.018	1.000
$\kappa_{1}^{(2)}$	-0.001	0.001	1.000
$\kappa_{2}^{(2)}$	-0.002	0.001	1.000
$\kappa_{3}^{(2)}$	-0.003	0.001	1.000
$\kappa_{4}^{(2)}$	-0.002	0.001	1.000
$\kappa_{5}^{(2)}$	-0.001	0.001	1.000
$\kappa_{6}^{(2)}$	-0.002	0.001	1.000
$\kappa_{7}^{(2)}$	-0.002	0.001	1.000
$\kappa_{8}^{(2)}$	-0.003	0.001	1.000
$\kappa_{9}^{(2)}$	-0.005	0.001	1.000
$\kappa_{10}^{(2)}$	-0.005	0.001	1.000
$\kappa_{11}^{(2)}$	-0.006	0.001	1.000
$\kappa_{12}^{(2)}$	-0.007	0.001	1.000
$\kappa_{13}^{(2)}$	-0.004	0.001	1.000
$\kappa_{14}^{(2)}$	-0.005	0.001	1.000
$\kappa_{15}^{(2)}$	-0.003	0.001	1.000
$\kappa_{16}^{(2)}$	-0.004	0.001	1.000
$\kappa_{17}^{(2)}$	-0.003	0.001	1.000
$\kappa_{18}^{(2)}$	-0.003	0.001	1.000
$\kappa_{19}^{(2)}$	-0.004	0.001	1.000
$\kappa_{20}^{(2)}$	-0.003	0.001	1.000
$\kappa_{21}^{(2)}$	-0.004	0.001	1.000
$\kappa_{22}^{(2)}$	-0.003	0.001	1.000
$\kappa_{23}^{(2)}$	-0.002	0.001	1.000
$\kappa_{24}^{(2)}$	-0.002	0.001	1.000

CBD-X(2) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{25}^{(2)}$	-0.001	0.001	1.000
$\kappa_{26}^{(2)}$	-0.000	0.001	1.000
$\kappa_{27}^{(2)}$	0.001	0.001	1.000
$\kappa_{28}^{(2)}$	0.002	0.001	1.000
$\kappa_{29}^{(2)}$	0.002	0.001	1.000
$\kappa_{30}^{(2)}$	0.004	0.001	1.000
$\kappa_{31}^{(2)}$	0.005	0.001	1.000
$\kappa_{32}^{(2)}$	0.006	0.001	1.000
$\kappa_{33}^{(2)}$	0.008	0.001	1.000
$\kappa_{34}^{(2)}$	0.009	0.001	1.000
$\kappa_{35}^{(2)}$	0.010	0.001	1.000
$\kappa_{36}^{(2)}$	0.010	0.001	1.000
$\kappa_{37}^{(2)}$	0.011	0.001	1.000
$\kappa_{38}^{(2)}$	0.011	0.001	1.000
$\kappa_{39}^{(2)}$	0.013	0.001	1.000
$\kappa_{40}^{(2)}$	0.014	0.001	1.000
$\kappa_{41}^{(2)}$	0.016	0.001	1.000
$\kappa_{42}^{(2)}$	0.016	0.001	1.000
$\kappa_{43}^{(2)}$	0.017	0.001	1.000
$\kappa_{44}^{(2)}$	0.017	0.001	1.000
$\kappa_{45}^{(2)}$	0.018	0.001	1.000
$\kappa_{46}^{(2)}$	0.018	0.001	1.000
$\kappa_{47}^{(2)}$	0.019	0.001	1.000
$\kappa_{48}^{(2)}$	0.020	0.001	1.000
$\kappa_{49}^{(2)}$	0.019	0.001	1.000
$\kappa_{50}^{(2)}$	0.020	0.001	1.000
v_{11}	0.000	0.000	1.000
v_{22}	0.000	0.000	1.000
v_{12}	0.000	0.000	1.000
μ_1	-0.013	0.003	1.000
μ_2	0.000	0.000	1.000

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
α_1	-3.482	0.026	1.010
α_2	-3.415	0.023	1.010
α_3	-3.318	0.021	1.010
α_4	-3.238	0.019	1.010
α_5	-3.150	0.017	1.010
α_6	-3.059	0.017	1.009
α_7	-2.993	0.016	1.008
α_8	-2.889	0.016	1.006
α_9	-2.808	0.016	1.005
α_{10}	-2.719	0.016	1.004
α_{11}	-2.636	0.017	1.003
α_{12}	-2.553	0.017	1.003
α_{13}	-2.470	0.017	1.002
α_{14}	-2.381	0.017	1.002
α_{15}	-2.291	0.017	1.002
α_{16}	-2.192	0.018	1.002
α_{17}	-2.119	0.018	1.002
α_{18}	-2.025	0.019	1.002
α_{19}	-1.937	0.020	1.002
α_{20}	-1.845	0.021	1.001
α_{21}	-1.772	0.023	1.001
α_{22}	-1.686	0.025	1.001
α_{23}	-1.602	0.028	1.001
α_{24}	-1.525	0.032	1.001
α_{25}	-1.454	0.036	1.001
$\kappa_1^{(1)}$	-0.012	0.015	1.006
$\kappa_2^{(1)}$	-0.035	0.016	1.006
$\kappa_3^{(1)}$	-0.042	0.016	1.006
$\kappa_4^{(1)}$	-0.040	0.017	1.006
$\kappa_5^{(1)}$	-0.026	0.016	1.006
$\kappa_6^{(1)}$	-0.050	0.016	1.006
$\kappa_7^{(1)}$	-0.033	0.016	1.006
$\kappa_8^{(1)}$	-0.043	0.016	1.006

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_9^{(1)}$	-0.060	0.016	1.006
$\kappa_{10}^{(1)}$	-0.046	0.017	1.006
$\kappa_{11}^{(1)}$	-0.057	0.016	1.006
$\kappa_{12}^{(1)}$	-0.071	0.016	1.006
$\kappa_{13}^{(1)}$	-0.081	0.016	1.006
$\kappa_{14}^{(1)}$	-0.068	0.016	1.006
$\kappa_{15}^{(1)}$	-0.077	0.016	1.006
$\kappa_{16}^{(1)}$	-0.076	0.016	1.006
$\kappa_{17}^{(1)}$	-0.089	0.016	1.006
$\kappa_{18}^{(1)}$	-0.097	0.016	1.006
$\kappa_{19}^{(1)}$	-0.126	0.017	1.006
$\kappa_{20}^{(1)}$	-0.142	0.017	1.006
$\kappa_{21}^{(1)}$	-0.169	0.016	1.006
$\kappa_{22}^{(1)}$	-0.162	0.017	1.006
$\kappa_{23}^{(1)}$	-0.180	0.017	1.006
$\kappa_{24}^{(1)}$	-0.177	0.017	1.006
$\kappa_{25}^{(1)}$	-0.195	0.016	1.006
$\kappa_{26}^{(1)}$	-0.212	0.016	1.006
$\kappa_{27}^{(1)}$	-0.203	0.017	1.006
$\kappa_{28}^{(1)}$	-0.216	0.016	1.006
$\kappa_{29}^{(1)}$	-0.239	0.016	1.006
$\kappa_{30}^{(1)}$	-0.225	0.016	1.006
$\kappa_{31}^{(1)}$	-0.254	0.016	1.006
$\kappa_{32}^{(1)}$	-0.286	0.016	1.006
$\kappa_{33}^{(1)}$	-0.291	0.016	1.006
$\kappa_{34}^{(1)}$	-0.314	0.016	1.006
$\kappa_{35}^{(1)}$	-0.298	0.016	1.006
$\kappa_{36}^{(1)}$	-0.320	0.016	1.006
$\kappa_{37}^{(1)}$	-0.326	0.016	1.006
$\kappa_{38}^{(1)}$	-0.343	0.016	1.006
$\kappa_{39}^{(1)}$	-0.356	0.016	1.006
$\kappa_{40}^{(1)}$	-0.363	0.016	1.006
$\kappa_{41}^{(1)}$	-0.387	0.016	1.006

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{42}^{(1)}$	-0.438	0.016	1.006
$\kappa_{43}^{(1)}$	-0.471	0.016	1.006
$\kappa_{44}^{(1)}$	-0.487	0.016	1.006
$\kappa_{45}^{(1)}$	-0.508	0.016	1.006
$\kappa_{46}^{(1)}$	-0.547	0.016	1.006
$\kappa_{47}^{(1)}$	-0.570	0.016	1.006
$\kappa_{48}^{(1)}$	-0.617	0.016	1.006
$\kappa_{49}^{(1)}$	-0.620	0.016	1.006
$\kappa_{50}^{(1)}$	-0.643	0.016	1.006
$\kappa_{1}^{(2)}$	-0.001	0.001	1.002
$\kappa_{2}^{(2)}$	-0.002	0.002	1.002
$\kappa_{3}^{(2)}$	-0.003	0.002	1.002
$\kappa_{4}^{(2)}$	-0.001	0.002	1.002
$\kappa_{5}^{(2)}$	-0.000	0.002	1.002
$\kappa_{6}^{(2)}$	-0.002	0.002	1.002
$\kappa_{7}^{(2)}$	-0.002	0.002	1.002
$\kappa_{8}^{(2)}$	-0.002	0.002	1.002
$\kappa_{9}^{(2)}$	-0.004	0.002	1.002
$\kappa_{10}^{(2)}$	-0.004	0.002	1.002
$\kappa_{11}^{(2)}$	-0.005	0.002	1.002
$\kappa_{12}^{(2)}$	-0.006	0.002	1.002
$\kappa_{13}^{(2)}$	-0.004	0.002	1.002
$\kappa_{14}^{(2)}$	-0.004	0.002	1.002
$\kappa_{15}^{(2)}$	-0.003	0.002	1.002
$\kappa_{16}^{(2)}$	-0.003	0.002	1.002
$\kappa_{17}^{(2)}$	-0.002	0.002	1.002
$\kappa_{18}^{(2)}$	-0.002	0.002	1.002
$\kappa_{19}^{(2)}$	-0.004	0.002	1.002
$\kappa_{20}^{(2)}$	-0.003	0.002	1.002
$\kappa_{21}^{(2)}$	-0.004	0.002	1.002
$\kappa_{22}^{(2)}$	-0.003	0.002	1.002
$\kappa_{23}^{(2)}$	-0.001	0.002	1.002
$\kappa_{24}^{(2)}$	-0.002	0.002	1.002

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{25}^{(2)}$	-0.001	0.002	1.002
$\kappa_{26}^{(2)}$	0.000	0.002	1.002
$\kappa_{27}^{(2)}$	0.001	0.002	1.002
$\kappa_{28}^{(2)}$	0.002	0.002	1.002
$\kappa_{29}^{(2)}$	0.002	0.002	1.002
$\kappa_{30}^{(2)}$	0.005	0.002	1.002
$\kappa_{31}^{(2)}$	0.006	0.002	1.002
$\kappa_{32}^{(2)}$	0.007	0.002	1.002
$\kappa_{33}^{(2)}$	0.008	0.002	1.002
$\kappa_{34}^{(2)}$	0.010	0.002	1.002
$\kappa_{35}^{(2)}$	0.010	0.002	1.002
$\kappa_{36}^{(2)}$	0.011	0.002	1.002
$\kappa_{37}^{(2)}$	0.011	0.002	1.002
$\kappa_{38}^{(2)}$	0.012	0.002	1.002
$\kappa_{39}^{(2)}$	0.014	0.002	1.002
$\kappa_{40}^{(2)}$	0.015	0.002	1.002
$\kappa_{41}^{(2)}$	0.017	0.002	1.002
$\kappa_{42}^{(2)}$	0.016	0.002	1.002
$\kappa_{43}^{(2)}$	0.017	0.002	1.002
$\kappa_{44}^{(2)}$	0.018	0.002	1.002
$\kappa_{45}^{(2)}$	0.019	0.002	1.002
$\kappa_{46}^{(2)}$	0.019	0.002	1.002
$\kappa_{47}^{(2)}$	0.020	0.002	1.002
$\kappa_{48}^{(2)}$	0.020	0.002	1.002
$\kappa_{49}^{(2)}$	0.020	0.002	1.002
$\kappa_{50}^{(2)}$	0.020	0.002	1.002
$\kappa_{1}^{(3)}$	0.000	0.000	1.001
$\kappa_{2}^{(3)}$	-0.000	0.000	1.001
$\kappa_{3}^{(3)}$	-0.000	0.000	1.001
$\kappa_{4}^{(3)}$	0.000	0.000	1.001
$\kappa_{5}^{(3)}$	0.000	0.000	1.001
$\kappa_{6}^{(3)}$	0.000	0.000	1.001
$\kappa_{7}^{(3)}$	0.000	0.000	1.001

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_8^{(3)}$	0.000	0.000	1.001
$\kappa_9^{(3)}$	0.000	0.000	1.001
$\kappa_{10}^{(3)}$	0.000	0.000	1.001
$\kappa_{11}^{(3)}$	0.000	0.000	1.001
$\kappa_{12}^{(3)}$	0.000	0.000	1.001
$\kappa_{13}^{(3)}$	0.000	0.000	1.001
$\kappa_{14}^{(3)}$	0.000	0.000	1.001
$\kappa_{15}^{(3)}$	0.000	0.000	1.001
$\kappa_{16}^{(3)}$	0.000	0.000	1.001
$\kappa_{17}^{(3)}$	0.000	0.000	1.001
$\kappa_{18}^{(3)}$	−0.000	0.000	1.001
$\kappa_{19}^{(3)}$	−0.000	0.000	1.001
$\kappa_{20}^{(3)}$	−0.000	0.000	1.001
$\kappa_{21}^{(3)}$	−0.000	0.000	1.001
$\kappa_{22}^{(3)}$	−0.000	0.000	1.001
$\kappa_{23}^{(3)}$	−0.000	0.000	1.001
$\kappa_{24}^{(3)}$	−0.000	0.000	1.001
$\kappa_{25}^{(3)}$	−0.000	0.000	1.001
$\kappa_{26}^{(3)}$	−0.000	0.000	1.001
$\kappa_{27}^{(3)}$	−0.000	0.000	1.001
$\kappa_{28}^{(3)}$	−0.000	0.000	1.001
$\kappa_{29}^{(3)}$	−0.000	0.000	1.001
$\kappa_{30}^{(3)}$	−0.000	0.000	1.001
$\kappa_{31}^{(3)}$	−0.000	0.000	1.001
$\kappa_{32}^{(3)}$	−0.000	0.000	1.001
$\kappa_{33}^{(3)}$	0.000	0.000	1.002
$\kappa_{34}^{(3)}$	0.000	0.000	1.001
$\kappa_{35}^{(3)}$	0.000	0.000	1.001
$\kappa_{36}^{(3)}$	0.000	0.000	1.001
$\kappa_{37}^{(3)}$	0.000	0.000	1.001
$\kappa_{38}^{(3)}$	0.000	0.000	1.001
$\kappa_{39}^{(3)}$	0.000	0.000	1.001
$\kappa_{40}^{(3)}$	0.000	0.000	1.001

CBD-X(3) Model	Mean	Standard Deviation	Gelman-Rubin Test
$\kappa_{41}^{(3)}$	0.000	0.000	1.001
$\kappa_{42}^{(3)}$	0.000	0.000	1.002
$\kappa_{43}^{(3)}$	0.000	0.000	1.001
$\kappa_{44}^{(3)}$	0.000	0.000	1.001
$\kappa_{45}^{(3)}$	0.000	0.000	1.001
$\kappa_{46}^{(3)}$	0.000	0.000	1.001
$\kappa_{47}^{(3)}$	0.001	0.000	1.001
$\kappa_{48}^{(3)}$	0.001	0.000	1.001
$\kappa_{49}^{(3)}$	0.001	0.000	1.001
$\kappa_{50}^{(3)}$	0.001	0.000	1.001
v_{11}	0.000	0.000	1.000
v_{22}	0.000	0.000	1.000
v_{33}	0.000	0.000	1.000
v_{12}	0.000	0.000	1.000
v_{13}	0.000	0.000	1.000
v_{23}	0.000	0.000	1.000
μ_1	-0.013	0.003	1.000
μ_2	0.000	0.000	1.000
μ_3	0.000	0.000	1.000