Modeling Non-linearity in Mortality Data: Application to Longevity Bond Pricing

by

Huijing Li

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

Faculty of Management

Warren Centre for Actuarial Studies and Research

University of Manitoba

Winniprg

Contents

List of Figures

List of Tables

Modeling Non-linearity in Mortality Data: Application to Longevity Bond Pricing

Huijing Li

September 15, 2015

Abstract

Human mortality has been improving faster than expected over the past few decades. This unprecedented improvement has caused significant financial stress to pension plan sponsors and annuity providers. To better model and forecast mortality rates, we examine the nonlinearity in mortality data from England and Wales with a sample period of 1900-2011. More specifically, we consider four nonlinear time series models: threshold autoregressive model, Markov regime switching model, structural change model, and auto-regressive conditional heteroskedasticity model. We then compare their goodness of fit and forecasting performance. Finally, we study the impact of different nonlinear models on longevity bond pricing.

1 Introduction

1.1 Background

In recent years, life expectancy has been increasing and mortality improvement has been accelerating. Medical advances and a strengthening health system have contributed significantly to this rising longevity. Mortality rates have been trending downwards over at least 100 years for all age groups in 11 countries (Mitchell et al., 2013), and average life expectancy has increased from 47 to 75 from 1900 to 1988 in US (Lee & Carter, 1992). Longevity improvement is good news to human beings; however, it causes financial risk to pension plan sponsors and annuity providers. This risk, which is known as longevity risk, is the risk of human mortality improving faster than expected. For pension plan sponsors, when pensioners live longer than expected, pension benefit payouts will be higher than expected. For annuity providers, as life expectancy increases, more annuity liabilities than anticipated will be paid out.

Pension plan administrators and life annuity providers are facing the challenge of how to reduce their exposure to longevity risk. There are several ways to hedge longevity risk. The traditional way is to buy reinsurance; however, the capital markets have more potential of dealing with unexpected catastrophe losses than the reinsurance market (Lin & Cox, 2005). Therefore, mortality-linked securitization is a sufficient way to hedge longevity risk. Mortality-linked securities, which are financial derivatives that link the payoffs to certain mortality indices, consist of mortality derivatives and longevity derivatives. Mortality derivatives are designed to protect extreme losses caused by catastrophic events such as wars or severe epidemics. Longevity derivatives are designed to against long-term unexpected long life expectancy (Zhou et al., 2015). For example, a longevity bond with annual coupons proportional to cohort survival rates can be used to hedge longevity risk (Levantesi et al., 2008).

When mortality improves faster than expected, the longevity bond holders will receive higher coupon rates. The first morality derivative, which was issued by Swiss Re and Vita Capital in December 2003, is a three-year floating-rate bond which links the return of principal to a mortality index. Another example of mortality derivatives is survivor swap. A survivor swap is a financial contract in which two parties agree to exchange cash flows, based on certain mortality indices, in the future. Survivor swaps can help insurers manage mortality risk.

The first longevity bond was issued by European Investment Bank (EIB) and structured by BNP Paribas in 2005. The EIB/BNP bond is a 25-year maturity bond with an annual payment of realized cohort survival rate for men in England and Wales. This longevity bond is designed to protect bondholders against unexpected mortality improvement. However, EIB/BNP bonds did not generate enough demand. Some reasons that EIB/BNP bonds failed might have resulted from limited issuing size preventing a liquid market, large upfront capital leaving no capital to hedge other risks, and fund managers not welcoming the bonds because they thought that the longevity bonds would not make money (Cairns et al., 2005). In 2010, Kortis Capital Ltd issued a longevity bond and successfully transferred 50 million USD of longevity trend risk to the capital markets. This bond was triggered when a large divergence occurs in the mortality improvements between males aged 75-85 in England and Wales and males aged 55-65 in the US.

1.2 Literature Review

In order to successfully price and trade longevity bonds, a suitable mortality model is necessary. Researchers have recently developed various mortality models. For example, Lee and Carter (1992) proposed a stochastic mortality model which

consists of two age-specific factors and one time-varying factor. The time-varying factor is also called the mortality index, since it can capture the mortality trend. The Lee-Carter mortality model has been widely used, and it has been employed as a benchmark statistical methodology for US Census Bureau population to estimate the long-run forecast of US life expectancy (Hollmann et al., 2000). Renshaw and Haberman (2005) extended the Lee-Carter model by adding a cohort effect. The authors stated that this extension of the Lee-Carter model can "capture characteristic and systematic mortality patterns, directly attributable to age, period and cohort effect" (p.569). Furthermore, since mortality improvement has been proven to be more significant at higher ages, Cairns et al. (2006) introduced a two-factor stochastic model to capture mortality improvement for higher age groups. The mortality-rate dynamics for all ages are commonly effected by the first factor, while the second factor has more influences on the mortality for the higher ages. Mitchell et al. (2013) proposed a mortality model which was a transformation of the Lee-Carter model. The authors modeled mortality rate changes rather than mortality rate itself to capture the mortality improvement level at different age groups. In this thesis, we adopt the Lee-Carter model and use a two-step fitting method. First, we fit the Lee-Carter model to obtain age-specific factors and mortality index. Second, we further model the estimated mortality index by time-series models.

In recent literature, mortality indices are often modeled by Autoregressive Integrated Moving Average (ARIMA) models, such as random walk with drift or Autoregressive (AR) model. However, standard ARIMA models assume a linear correlation structure in a mortality index and cannot capture nonlinear patterns. Our analysis of historical data shows evidence of non-linearity in the mortality index. Therefore, a nonlinear time series model is more appropriate.

Some researchers have considered nonlinear models for modeling mortality index.

Milidonis et al. (2010) applied regime switching geometric Brownian motion to US population mortality index. Regime switching models allow mortality to jump between different states. Each regime has different mean and variance, so changes in both mean and volatility can be captured in the mortality index. The authors have demonstrated that regime switching model performed better than random walk with drift in terms of fitting; moreover, significant improvement in mortality risk modeling was provided.

In addition, Hainaut (2012) applied regime switching model to extend the Renshaw and Haberman (2003) model and showed that the regime switching model has a significantly higher log-likelihood than the original Renshaw-Haberman model.

Li et al. (2011) used broken-trend stationary model on the Lee-Carter model. This model only allowed one break point in the mortality index. The authors concluded that mortality index was better explained by a broken-trend stationary model with a decline trending in the 1970s than by a random walk model. Sweeting (2011) modeled structural change based on the original Cairns-Blake-Dowd (CBD) model by using Chow Test to detect multiple structural changes in the mortality index. Berkum et al. (2013) considered random walk with a piece-wise constant drift on the Lee-Carter model with application to mortality in The Netherlands and Belgium. The authors allowed for multiple changes in the mortality index. The number of structural changes was found by optimizing Bayesian Information Criterion (BIC). The authors concluded that mortality rate projections based on structural change modeling were less sensitive to the calibration period compared to ARIMA model, and the projections for the mortality index better followed the observed trends.

Hong (2011) extended the Lee-Carter model with AutoRegressive Conditional Heteroskedasticity (ARCH)/Generalized ARCH (GARCH) model to capture volatility clustering in mortality improvement. In addition, Chen et al. (2014) applied

ARMA-GARCH model to mortality data, and they then modeled the residuals by a copula model to discover the mortality dependence of the residuals. ARCH/GARCH models developed by Engle (1982) and Bollerslev (1986) respectively were used to characterize the non-linearity of variance in financial time series. By modeling the conditional changing of volatility in mortality improvement, Hong concluded that ARCH/GARCH models can effectively capture the volatility clustering and can better describe and forecast mortality improvement in comparison to standard ARIMA models.

No-arbitrage pricing methods are often used in mortality-linked securities. The most important step in implementing no-arbitrage pricing method is to choose an appropriate risk-neutral measure. However, this step is difficult based on the limited transactions in the mortality market (Zhou et al., 2015). Chen et al. (2014) considered CAT bond pricing methodology in longevity bond pricing. The authors priced longevity bond based on the expected loss and a risk loading parameter. Zhou et al. (2015) applied a t$\hat{a}$tonnement approach to price longevity bond based on the primary economic concept of Demand and Supply. The authors regarded the pricing framework as a t$\hat{a}$tonnement process, in which market equilibrium is achieved through an auction. In this thesis, we adopt this economic pricing method, and determine the equilibrium of supply and demand by maximizing the expected utility level for the bond buyer and seller.

1.3 Objectives and Outlines

All the nonlinear models mentioned above provide better fitting than ordinary ARIMA models. Milidonis et al. (2010) stated that Markov regime switching models perform better in mortality fitting, since regime switching models not only provide

transparent representation of the status of structural changes in mortality, but also capture different mortality dynamics from the parameter estimates of the mean and volatility in mortality. However, Berkum et al. (2013) argued that structural change models are more appropriate to project mortality index than regime switching models. The authors reasoned that structural changes caused by medical advances and health system improvements were permanent, since medical advances and health improvements cannot be reversed. Different nonlinear mortality models may provide wide variations in forecasting. It is difficult to identify which model is more appropriate from the existing literature. Therefore, in this thesis we will examine the fitting and forecasting performance of four nonlinear models: Threshold Autoregressive model (TAR), Markov regime switching model, structural change model, and ARCH/GARCH model. This thesis is the first to apply TAR model to mortality data. TAR model proposed by Tong (1978) was developed to predict stock price movement, and is an extension of autoregressive model. TAR model admits multiple regimes to describe different dynamics of time series, and the movements between regimes are governed by threshold variables.

The remainder of this thesis is organized as follows: Section 2 introduces the Lee-Carter mortality model and fits this model to EW data. Section 3 describes four nonlinear models, and fits these models to the mortality index. This section also examines the goodness-of-fit and forecasting performance of different models. Section 4 applies economic pricing approach to longevity bond, and studies the impact of different nonlinear models on the bond price. Finally, Section 5 concludes the paper.

2 Mortality Model

2.1 Lee-Carter Model

In this thesis, we will use the Lee-Carter model to describe the dynamics of mortality rates. Lee-Carter model is adopted not only because it has been widely used in literature, but also because its simplicity allows us to focus on our main objectives. This model was designed to make a long-run forecast for US mortality. It was fitted to US mortality data from 1933 to 1987 using Singular Value Decomposition (SVD) method. Lee-Carter model can also be fitted by Maximum Log-likelihood Estimation (MLE) method (Wilmoth 1993).

Let $m_{x,t}$ be the central death rate of an individual at age x in year t. The central death rate is the number of deaths divided by the average number who were living at age x in year t. $m_{x,t}$ is modeled as:

$$\ln(\hat{m}_{x,t}) = a_x + b_x k_t \tag{1}$$

where a_x and b_x are age-specific effects, and k_t is the time-varying index. b_x measures the age-specific response to changes in k_t. k_t is also called the mortality index. It captures mortality changes over time.

In Lee and Carter (1992), the mortality index k_t is further modeled as a random walk with drift term:

$$k_t = \alpha + k_{t-1} + \epsilon_t \tag{2}$$

where α is a constant term, which measures the average change of mortality index each year. ϵ_t is an error term that follows a normal distribution with mean 0 and variance σ^2.

2.2 Data Description

The historical mortality data are obtained from Human Mortality Database (2014). The mortality data consists of death counts and exposure to death risk among the population of England and Wales (EW) with a sample age range of 0-110+ and a sample period of 1900-2011.

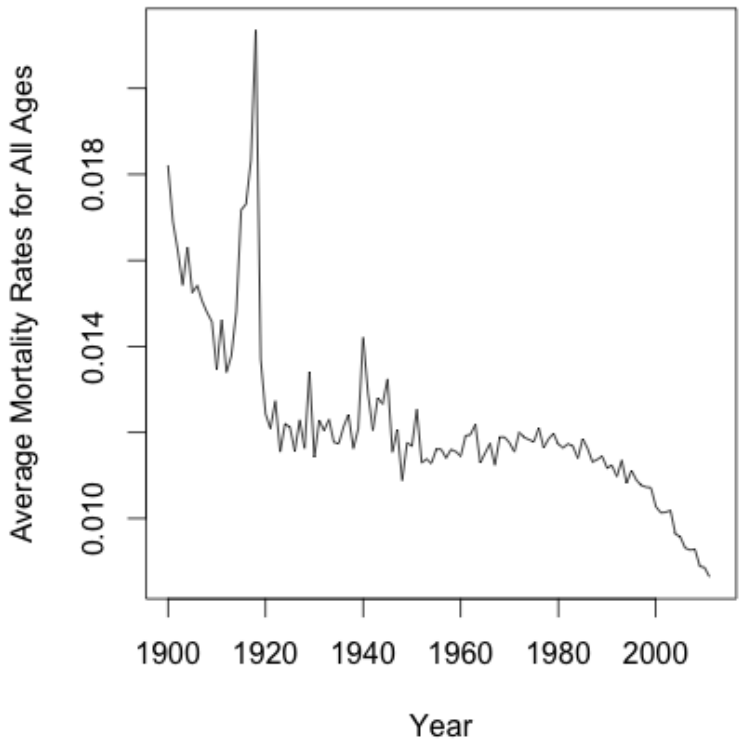


Figure 1: Average Mortality Rates of All Age Groups for EW Population.

Let the sample period be $[t_0, t_1]$, and the sample age range be $[x_0, x_1]$. The average mortality rates of all age groups is defined as $\frac{1}{x_1-x_0+1}\sum_{x=x_0}^{x_1} m_{x,t}$. Figure 1 depicts the average mortality rates of all age groups for EW population in the sample period. Figure 3 clearly shows that mortality rates have been decreasing over time. Mortality

rates were quite volatile before 1950, but have been less volatile in recent years. In addition, mortality rates have been decreasing concavely over the last few decades, which indicates that mortality improvement has been accelerating.

2.3 Fitting the Model

In the following subsection, we use MLE method to estimate the parameters in equation (1). We define $D_{x,t}$ as the random variable that represents the number of deaths at age x in year t, and $E_{x,t}$ as the observed number of exposures to risk. To derive the log-likelihood function, Wilmoth (1993) assumed that

$$D_{x,t} \sim Poisson(\hat{m}_{x,t} E_{x,t})$$

The log-likelihood function can then be expressed by

$$\sum_{x=x_0}^{x_1} \sum_{t=t_0}^{t_1} [D_{x,t} \ln(\hat{m}_{x,t} E_{x,t}) - \hat{m}_{x,t} E_{x,t} - \ln(D_{x,t}!)] \tag{3}$$

where $\ln(D_{x,t}!)$ is a constant term. There are $2x_1 - 2x_0 + t_1 - t_0 + 3$ parameters in Lee-Carter model, which contains $(x_1 - x_0 + 1)$ of a_x, $(x_1 - x_0 + 1)$ of b_x and $(t_1 - t_0 + 1)$ of k_t. In addition, the number of observations is $(x_1 - x_0 + 1)(t_1 - t_0 + 1)$. To find the maximum of this function, we use the Newton-Raphson iterative process, and update a_x, b_x and k_t one at a time. The details for the derivation of log-likelihood function and the Newton-Raphson iterative method are attached in Appendix A. To ensure the parameter uniqueness, we set two parameter constraints (1) $\sum_{x=x_0}^{x_1} b_x = 1$ and (2) $k_{t_1} = 0$.

Figure 2 plots the estimates of the age-specific effects a_x and b_x. a_x is high for $x \leq 1$. The major cause of high infant mortality rates is sudden infant death

syndrome. a_x then increases with age with a drop in the 20s. The higher mortality rates among adolescents are caused by three main reasons: vehicular deaths, weapon-related violence and suicide. The plot of b_x indicates the speed of mortality rate decline for different ages in response to changes in k_t. The plot shows that the mortality rates generally decline faster in younger age groups than older age groups. In other words, when the mortality rate decreases over time, the mortality at younger age groups improves more significantly than that at older age groups.

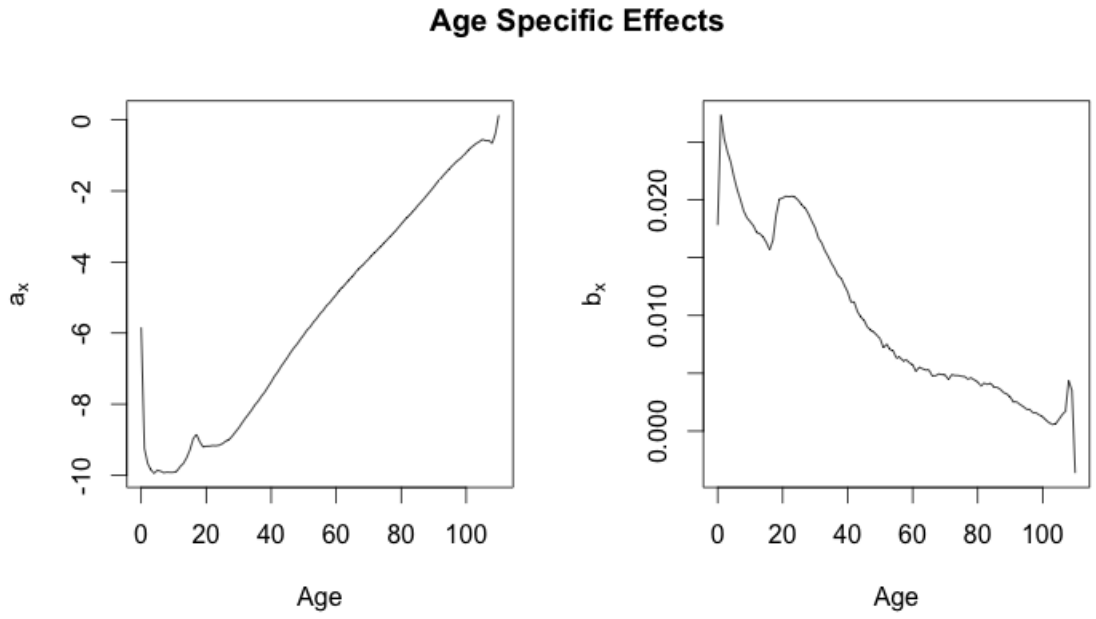


Figure 2: Estimates of Age-Specific Parameters for EW population.

Let $\Delta k_t = k_t - k_{t-1}$, and Δk_t indicate the mortality improvement at year t. The estimates of mortality index k_t and mortality improvement Δk_t are shown in Figure 2. The left panel in Figure 3 shows that the mortality index is trending downward over time, which implies that mortality rates are decreasing in the sample period. We also notice that there are two obvious mortality jumps around 1920 and 1940. The first

Figure 3: Estimate of Mortality Index and Mortality Improvement for EW Population.

mortality jump was caused by the First World War and the influenza pandemic in 1918, while the second mortality jump was caused by the Second World War. These catastrophic events caused a sudden increase in mortality rates. Figure 3 shows that the mortality index has concaved downward over the last few decades. The concave downward curve implies the accelerating mortality improvement.

The right panel of Figure 3 shows the first difference of the mortality index. This plot shows that Δk_t was very volatile before 1950, and relatively stable after 1950. We also notice a slightly downward trend in mortality improvement after year 2000. Both plots in Figure 3 provide evidence of non-linearity in mortality trend.

2.4 Modeling the Mortality Index

In the original paper, Lee and Carter (1992) modeled mortality index by random walk with drift. They provided two suggestions about how to consider the 1918 epidemic. One solution is to consider it to be an unusual event which makes disproportionate influences on the forecasting. Thus, we can set a dummy variable for year 1918 in order to remove its influence. Another solution is to treat the epidemic as an event that could possibly occur again in the future. In other words, the epidemic may provide valuable information for mortality forecasting. In this thesis, we adopt the latter suggestion of treating the observation in 1918 like any other.

We use Equation (2) to fit the mortality index k_t. Conditional sum of squares method is used to find the initial values for the parameter, then maximum likelihood method is applied to obtain the parameters. The fitted model can be expressed as follows:

$$k_t = -2.2343 + k_{t-1} + \epsilon_t \qquad \epsilon_t \sim N(0,\ 55.02) \tag{4}$$

The intercept coefficient, -2.2343, is the average annual change of mortality index. It means that the mortality index decreases by 2.2343 every year, and long term mortality improves in the sample period. The error term is independent and identically distributed (i.i.d.) normal with mean 0 and variance 55.02.

Antolin (2010) considered linear AR (1) for modeling and forecasting the mortality improvement Δk_t. This is equivalent to model k_t in ARIMA (1,1,0). The parameters are obtained by maximum likelihood method. The fitted linear AR (1) model for Δk_t is as follows:

$$\Delta k_t = -2.2122 - 0.2241\Delta k_{t-1} + \epsilon_t, \qquad \epsilon_t \sim N(0,\ 51.76) \tag{5}$$

Equation (5) also indicates that mortality improvement is trending linearly downward over time. The long term unconditional mean of Δk_t is -1.8072, which indicates that the mortality rate is improving significantly. The residuals not captured by the linear model have a variance of 51.76.

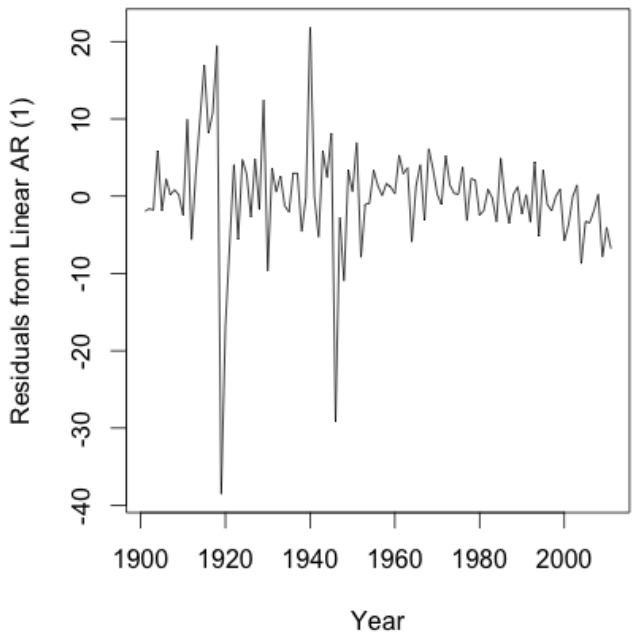


Figure 4: Residuals from Linear AR (1) Model

Figure 4 plots the residuals obtained from the fitted linear AR (1) model to mortality improvement. Figure 4 shows a downward trend after year 2000, and a cluster of volatility is also shown in the residuals. Therefore, the non-linearity exists in either the time series of mortality improvement or the residuals from a fitted AR (1) model.

3 Nonlinear Time Series Models

3.1 Determining AR and MA terms

In this section, we first introduce four nonlinear time series models: TAR model, Markov regime switching model, structural change model, and ARCH/GARCH model. We then fit the four nonlinear models to mortality improvement, Δk_t, to examine the fitting performance.

Before we introduce nonlinear models, we need to determine an ARIMA model that can appropriately capture the auto-correlation in the time series. We plot the Auto-Correlation Functions (ACF) and the Partial Auto-Correlation Functions (PACF) for the mortality improvement to determine the ARIMA lags. ACF provides a set of correlation coefficients between the time series and its lags over time, while PACF provides a set of partial correlation coefficients. The partial correlation is the amount of correlation that is not captured by their mutual correlations. For example, if $\Delta k_t = \alpha + \beta_1 \Delta k_{t-1} + \beta_2 \Delta k_{t-2}$, then the partial correlation of Δk_t and Δk_{t-2} is the amount of correlation that not explained by their common correlation from Δk_t and Δk_{t-1}.

Figure 5 shows the ACF and PACF plots for mortality improvement. The lag one auto-correlation is statistically significant. From the ACF plot, it appears that the ACF is cutting off at lag one. Therefore, MA term may be necessary. To further determine the AR and MA orders, we compare Bayesian Information Criterion (BIC) for models AR (1), MA (1), and ARIMA (1), and then select the model with the best BIC. BIC is a criterion used for model selection which penalizes additional parameters and hence helps to overcome any over-fitting problem that may present in complex

models. The formula for BIC is

$$\text{BIC} = -2 \ln \hat{L} + k \ln(n)$$

where $\hat{L}$ is the maximized value of the likelihood function of the model, k is the number of parameters to be estimated, and n is the number of observations. The reason that we only consider BIC in model selection is that BIC gives a stronger penalty to parameters. Since over fitting is our major concern, it is more appropriate to use BIC to select model in this thesis.

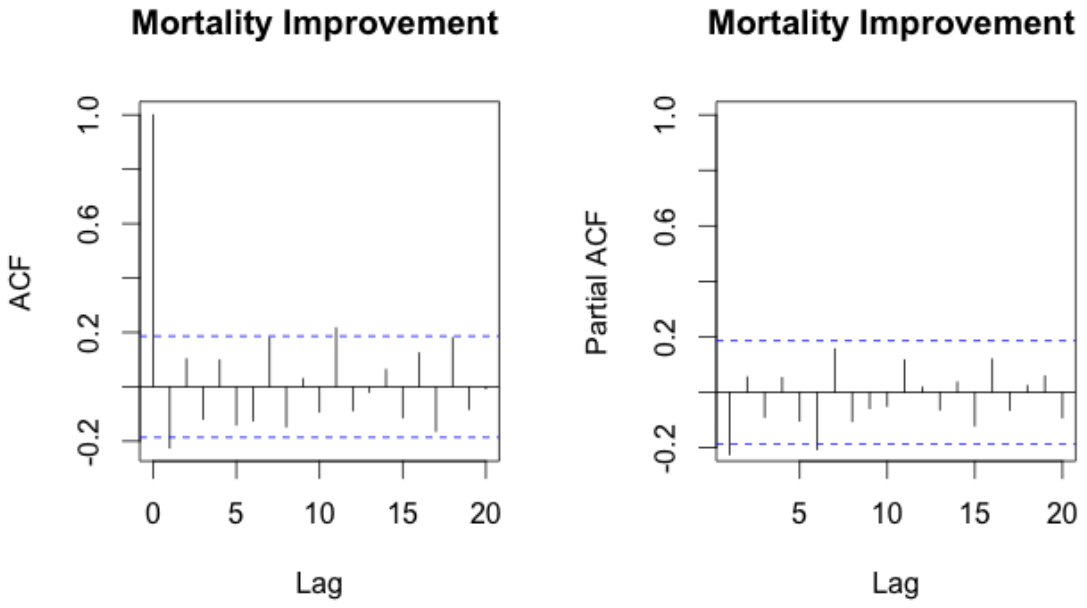


Figure 5: ACF and PACF of mortality improvement for EW population.

Table 1 summarizes the BIC results for different time series models, and AR (1) model generates the lowest BIC.

	AR (1)	MA (1)	ARMA (1)
BIC	762.5391	763.1591	766.3486

Table 1: BIC Values for Different Linear Models

PACF displays a significant spike at lag one, and a sharp cutoff after lag one. This indicates that all the higher order auto-correlations are effectively explained by the lag one auto-correlation. Both Table 1 and PACF plot suggest a first-order AR model. Therefore, AR (1) model will be used as the basic structure for mortality improvement for EW populations in this thesis.

3.2 Threshold Autoregressive (TAR) Model

3.2.1 Model Specification

TAR model was first presented and proposed by Tong (1978) to describe the non-linear movement of stock prices in financial market. Later, comprehensive discussion and extensions of TAR model were made by Tong and Lim (1980) and Tong (1983). TAR model uses piece-wise AR models to improve linear approximation. AR models are estimated separately in each time series segment as defined by threshold variables. A threshold AR (1) model with multiple threshold variables can be expressed as follows:

$$\Delta k_t = \begin{cases} \alpha^{(1)} + \beta^{(1)}\Delta k_{t-1} + \epsilon_t^{(1)}, & \text{if } \Delta k_{t-1} \leq r_1 \\ \alpha^{(2)} + \beta^{(2)}\Delta k_{t-1} + \epsilon_t^{(2)}, & \text{if } r_1 < \Delta k_{t-1} \leq r_2 \\ \vdots & \\ \alpha^{(n+1)} + \beta^{(n+1)}\Delta k_{t-1} + \epsilon_t^{(n+1)}, & \text{if } \Delta k_{t-1} > r_n \end{cases} \tag{6}$$

where r_i is the i th threshold variable, and $r_1 < r_2 < \cdots < r_m$. $\alpha^{(i)}$ and $\beta^{(i)}$ are AR parameters for regime i. $\epsilon_t^{(i)}$ is the error term for regime i, and $\epsilon_t^{(i)}$ follows normal distribution with mean 0 and variance $\sigma_t^{(i)2}$. TAR model allows the AR structure to change based on the regimes determined by threshold variables. By doing this, complex non-linearity dynamics can be captured. In Equation (6), m threshold variables classify the observations into $m+1$ regimes. Δk_t switches between regimes based on the value of Δk_{t-1} and the threshold variables.

3.2.2 Model Fitting

To fit the TAR model and obtain estimates of parameters, we use a two-step procedure. In the first step, we use MLE method to fit TAR model with different numbers of threshold variables. A minimal percentage of observations in each regime is set to be 10%. In the second step, we calculate the log-likelihood value and BIC for each fitted model and choose the model with the best BIC.

	AR	TAR-2R	TAR-3R	TAR-4R
BIC	762.5391	755.8349	734.9055	745.8303

*2R, 3R, and 4R indicate 2 regimes, 3 regimes and 4 regimes, respectively

Table 2: BIC Values for TAR Model

Table 2 summarizes the BIC values for linear AR (1) model and TAR (1) models with two regimes, three regimes, and four regimes. The TAR model with three regimes has the lowest BIC, meaning that TAR model with two thresholds can better describe the mortality improvement.

The fitted TAR (1) model with two thresholds can be written as follows:

$$\Delta k_t = \begin{cases} 2.2951 + 0.1795\Delta k_{t-1} + \epsilon_t^{(1)}, & \Delta k_{t-1} \leq -4.58426 \\ -2.6544 - 0.2871\Delta k_{t-1} + \epsilon_t^{(2)}, & -4.58426 < \Delta k_{t-1} \leq 1.84435 \\ -0.2803 - 0.9070\Delta k_{t-1} + \epsilon_t^{(3)}, & \Delta k_{t-1} > 1.84435 \end{cases}$$

where the error terms are i.i.d. following a normal distribution with mean 0 and variance 17.9570 in the first regime, 20.6113 in the second regime and 136.7667 in the third regime. The two thresholds are -4.58426 and 1.84435, and two thresholds divided Δk_t into three regimes. There are 26.12 % of observations in the first regime, 50.45 % of observations in the second regime, and 23.42 % of observations in the third regime.

Figure 6 shows the historical mortality improvement with two thresholds obtained from TAR model. The two horizontal dashed lines show the location of the two thresholds. Also, this figure shows that regime two has more observations than the other two regimes.

We apply t statistic to test the significance of the parameter estimates. The purpose is to remove insignificant parameter estimates. The null hypothesis of the test is that parameter estimate of AR term is not significant, which is $H_o : \beta^{(i)} = 0$. The alternative hypothesis will be $H_a : \beta^{(i)} \neq 0$. The t-value is calculated by the ratio of estimated parameter and its standard error. Table 3 gives p values of the t statistic.

Using 5% significance level, Table 3 shows that only the AR coefficient for regime three is significant. However, the AR parameter estimate for regime one and two are

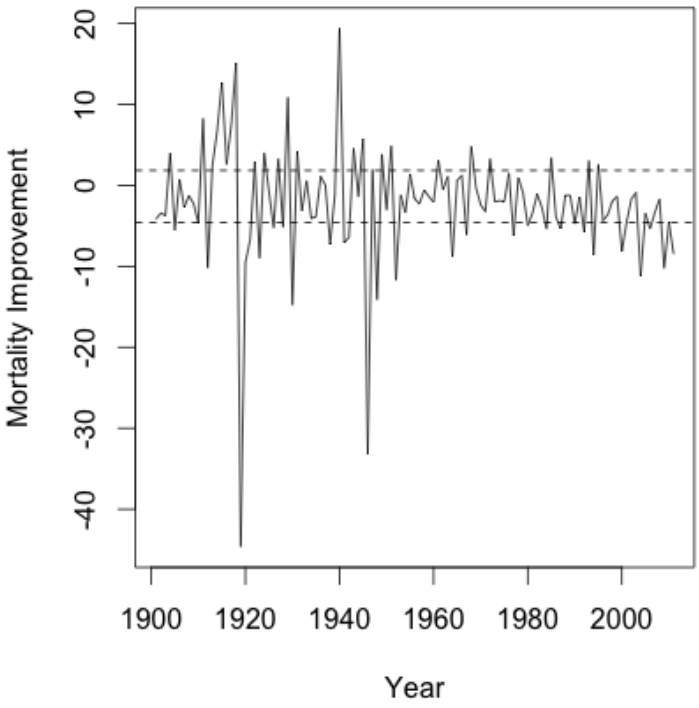


Figure 6: Fitting TAR Model with Two Threshold Variables.

not significant. Therefore, we refit the first two regimes by random walk with drift. The newly fitted model is expressed as

$$\Delta k_t = \begin{cases} 10.3902 + \Delta k_{t-1} + \epsilon_t^{(1)}, & \Delta k_{t-1} \leq -4.58426 \\ -0.3936 + \Delta k_{t-1} + \epsilon_t^{(2)}, & -4.58426 < \Delta k_{t-1} \leq 1.84435 \\ -0.2803 - 0.9070 \Delta k_{t-1} + \epsilon_t^{(3)}, & \Delta k_{t-1} > 1.84435 \end{cases}$$

where error terms are i.i.d. that follow normal distribution with mean 0 and variance 66.3521 in the first regime, 25.2181 in the second regime, and 136.7667 in the third regime.

We then compare BIC values for reduced model and the TAR (1) model fitted ear-

	$\beta^{(1)}$	$\beta^{(2)}$	$\beta^{(3)}$
P value	0.2422	0.6082	0.0064

Table 3: p-value of AR Coefficients for TAR-3R model

lier to examine if the parameter reduction is appropriate. Table 4 shows that the BIC value for the original TAR (1) model is much lower. Even through the original TAR (1) model contains insignificant parameter estimates, it does not necessarily mean these estimates are irrelevant to the model. Removing these parameter estimates results in changing the model dynamic which causes the lost of relevant information, thus we get a undesirable model fitting. Therefore, we choose to use the non-reduced TAR (1) model.

	TAR-3R	TAR-Reduced
BIC	755.8349	779.9320

Table 4: BIC Values for TAR (1) Model and its Reduced Version

3.2.3 Model Interpretation

The fitted TAR model with three regimes can be interpreted as follows.

when $\Delta k_{t-1} < -4.58426$, Δk_t will be in the first regime, and $E(\Delta k_t \mid \Delta k_{t-1}) < 1.4722$. There is a higher than normal mortality improvement at time t-1. If $\Delta k_{t-1} < -12.7861$, $E(\Delta k_t \mid \Delta k_{t-1}) < 0$, meaning there will be mortality improvement at time t, but the level is much smaller than time t-1. A very negative Δk_{t-1} may imply that the population recovered from a mortality jump at time t-1. The recovery may last more than one year, and hence time t may still

see a small improvement. If $-12.7861 < \Delta k_{t-1} < -4.58426$, then $0 < E(\Delta k_t \mid \Delta k_{t-1}) < 1.4722$, mortality deteriorates slightly.

When $-4.58426 \leq \Delta k_{t-1} < 1.84435$, Δk_t will be in the second regime, and $-3.1839 < E(\Delta k_t \mid \Delta k_{t-1}) < -1.3388$. Mortality is expected to improve moderately at time t. This regime has the most observations and it represents the typical mortality improvement dynamic in a normal year.

When $\Delta k_{t-1} > 1.84435$, Δk_t will be in the third regime, and $E(\Delta k_t \mid \Delta k_{t-1}) < -1.9531$. A very high positive Δk_{t-1} implies that there may be a mortality jump at time t-1. At time t, mortality returns to the normal level, resulting in a very high mortality improvement. If the mortality deterioration at time t-1 is caused by random fluctuation, the mortality is expected to improve as usual. The recovery from the mortality jump causes the high variance in this regime.

3.3 Markov Regime Switching Model

3.3.1 Model Specification

Markov regime switching model proposed by Hamilton (1989) admits different structures at different state levels, and the switching is based on a state variable which follows the first-order Markov Chain. Milidonis et al. (2010) state that Markov regime switching model can identify the time of shock for the underlying mortality variable, and this model can capture both a temporary mortality shock or permanent mortality improvements.

Suppose that the historical mortality improvement can be described by AR (1) model. The n-state Markov switching model can be expressed as follows:

$$\Delta k_t = \begin{cases} \alpha^{(1)} + \beta^{(1)}\Delta k_{t-1} + \epsilon_t^{(1)}, & S_t = 1 \\ \alpha^{(2)} + \beta^{(2)}\Delta k_{t-1} + \epsilon_t^{(2)}, & S_t = 2 \\ \vdots & \\ \alpha^{(n)} + \beta^{(n)}\Delta k_{t-1} + \epsilon_t^{(n)}, & S_t = n \end{cases} \tag{7}$$

where S_t is the state variable, which indicates the state level at time t. The error term $\epsilon_t^{(i)}$ are i.i.d. that follows normal distribution with mean 0 and variance $\sigma^{(i)2}$. When S_t changes, Δk_t switches from one AR (1) mode to another.

S_t is assumed to follow a first order Markov chain. Let p_{ij} denote the transition probability from state i to state j. The Markovian probability transition matrix M that describes the random switching between different regimes can be written as

$$\begin{aligned} M &= \begin{bmatrix} P(S_t = 1 \mid S_{t-1} = 1) & P(S_t = 2 \mid S_{t-1} = 1) & \cdots & P(S_t = n \mid S_{t-1} = 1) \\ P(S_t = 1 \mid S_{t-1} = 2) & P(S_t = 2 \mid S_{t-1} = 2) & \cdots & P(S_t = n \mid S_{t-1} = 2) \\ \vdots & & & \\ P(S_t = 1 \mid S_{t-1} = n) & P(S_t = 2 \mid S_{t-1} = n) & \cdots & P(S_t = n \mid S_{t-1} = n) \end{bmatrix} \\ &= \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{n1} \\ p_{21} & p_{22} & \cdots & p_{n2} \\ \vdots & \vdots & \vdots & \vdots \\ p_{n1} & p_{2n} & \cdots & p_{nn} \end{bmatrix} \end{aligned} \tag{8}$$

3.3.2 Model Fitting

Let Θ represent the set of parameters in the Markov switching model, and Ω_t denote the vector of $(\Delta k_{t_1}, \Delta k_{t_1-1}, \cdots, \Delta k_{t_0+1})$. The log-likelihood function of the Markov switching model based on the observations of mortality improvement can be expressed as:

$$\ln[L(\Theta \mid \Delta k_{t_1}, \Delta k_{t_1-1}, \cdots, \Delta k_{t_0+1})] = \ln f(\Delta k_{t_0+1} \mid \Theta) + \sum_{t=t_0+2}^{t_1} \ln f(\Delta k_t \mid \Omega_{t-1}, \Theta) \quad (9)$$

where f is the probability density function. We maximize the log-likelihood to obtain parameter estimates, and the parameter estimation was proposed by Milidonis et al. (2010). The details are attached in Appendix B.

To determine the number of regimes, we first fit the model with two, three, and four regimes, respectively, and then choose the model with the lowest BIC.

	AR	MRSW-2R	MRSW-3R	MRSW-4R
BIC	762.5391	720.0849	736.5640	772.8863

*2R, 3R, and 4R indicate 2 regimes, 3 regimes and 4 regimes, respectively

Table 5: BIC Values for Markov Regime Switching Model

Table 5 summarizes the BIC for AR (1) model and the three Markov regime switching models with two regimes, three regimes, and four regimes. The regime switching model with two regimes has the lowest BIC; hence, it is selected for our data. The fitted model is written as follows:

$$\Delta k_t = \begin{cases} -2.6722 - 0.1232\Delta k_{t-1} + \epsilon_t^{(1)}, & S_t = 1 \\ -3.2395 - 0.4768\Delta k_{t-1} + \epsilon_t^{(2)}, & S_t = 2 \end{cases}$$

where $\epsilon_t^{(1)}$ are i.i.d. following normal distribution with mean 0 and variance 223.7507, and $\epsilon_t^{(2)}$ are i.i.d. following normal distribution with mean 0 and variance 13.2227. The transition matrix is

$$M = \begin{bmatrix} 0.8515185 & 0.1484815 \\ 0.03027188 & 0.96972812 \end{bmatrix}$$

The transition probabilities can be explained as follows: when Δk_{t-1} is in regime one, the probability that Δk_t remains in the first regime is 0.8515, and the probability that the Δk_t jumps to the second regime is 0.1485. When Δk_{t-1} is in the second regime, the probability that Δk_t jumps to regime one is 0.0303, and the probability that Δk_t remains in regime two is 0.9697.

Figure 7 provides the plot of Δk_t and the probability that Δk_t is in each regime. The lower two panels of Figure 7 depict the smoothing probability of Δk_t in regime one and regime two, respectively, given its observed value. The top two panels shadow the period that Δk_t stays in regime one and regime two. This figure indicates that given its observed value, Δk_t has a high probability in regime one during 1910 to 1920 and 1940 to 1950, and Δk_t has a higher probability in regime two during other time periods.

We use t-statistic to test the significance of AR coefficients in the fitted Markov regime switching model. Table 6 shows the p-value of the t-statistic test. At 5 % significance level, the estimate of AR coefficient in regime one is not significant, while that in regime two is significant.

Therefore, we reduce the structure for the first regime to random walk with drift and recalculate the BIC for the reduced model. The fitted reduced model is written

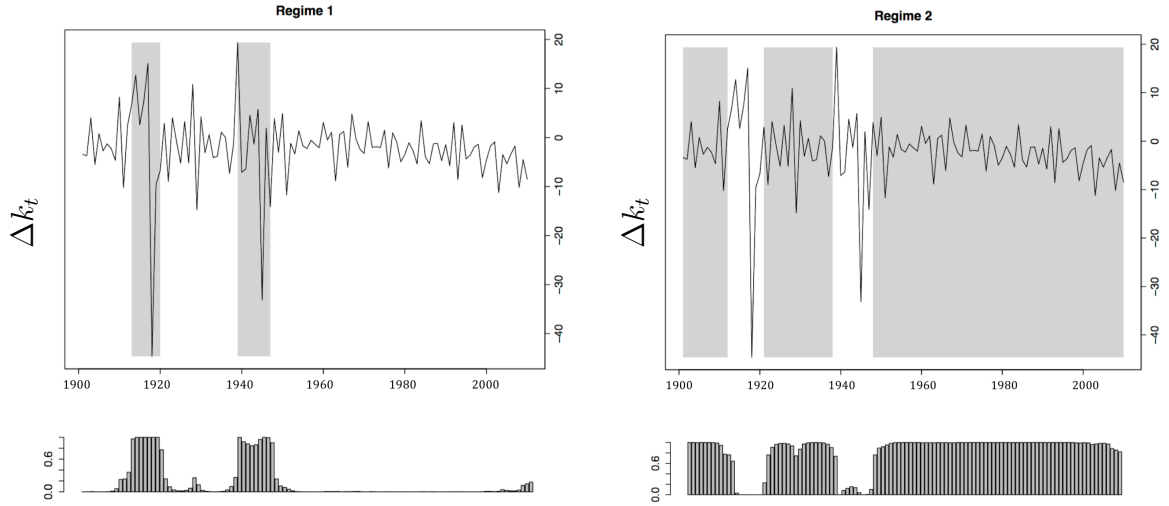


Figure 7: Δk_t Associated to Different Regimes versus the Smoothing Probabilities.

	$\beta^{(1)}$	$\beta^{(2)}$
p-value	0.5935	4.454e-06

Table 6: p-value of AR Coefficients for Markov Regime Switching Model with Two Regimes

as follows:

$$\Delta k_t = \begin{cases} -2.4554 + \Delta k_{t-1} + \epsilon_t^{(1)}, & S_t = 1 \\ -3.2395 - 0.4768\Delta k_{t-1} + \epsilon_t^{(2)}, & S_t = 2 \end{cases}$$

where $\epsilon_t^{(1)}$ are i.i.d. normal with mean 0 and variance 228.6334, and $\epsilon_t^{(2)}$ are i.i.d. normal with mean 0 and variance 13.2227.

As shown in Table 7, the BIC of the reduced model is lower. Therefore, we shall continue with the reduced model.

	MRSM-2R	MRSM-2R Reduced
BIC	720.0849	715.7067

Table 7: BIC Values for MRSM-2R and its Reduced Version

3.3.3 Model Interpretation

From the fitted two regime Markov switching model, the mortality improvement, Δk_t, switches between two regimes. The unconditional means of Δk_t in the two regimes are -2.4554 and -2.1936, respectively, which implies that mortality is improving significantly in both regimes. Therefore, mortality improving speed is not the primary reason for dividing Δk_t into two regimes. The variance of the error terms in regime one is 228.6334, while the variance in regime two is 13.2227. The much higher variance in regime one suggests that the Δk_t tends to have large shocks in regime one, while it is more stable in regime two. Figure 7 provides the evidence that the change of volatility in mortality improvement makes a huge contribution to the switching.

From Figure 7, given its observed value, the smoothed probability that Δk_t is in regime one is very high during the 1920s and the 1950s. As mentioned in section 2, wars and severe flu occurred during these two decades, and the mortality variation is unstable during these two periods. The transition probability suggests that if Δk_{t-1} experiences a shock, Δk_t has a very high probability of remaining in regime one. However, when Δk_{t-1} is in the stable regime two, Δk_t has a high probability of remaining in regime two.

3.4 Structural Change Model

3.4.1 Model Specification

Structural change model was originally presented by Lewis (1955). This model was developed to demonstrate the structural transformation from the rural agricultural sector to the urban industrial sector due to economic growth. Recently, some researchers have considered structural change model in mortality modeling, such as Sweeting (2011) and Berkum et al. (2013). Unlike TAR model and Markov regime switching model, the changes in structural change model are permanent. In other words, the mortality improvement, Δk_t, will follow a new model after each change point, and it may never switch back to the previous model.

In a structural change model, changes may occur in the mean, variance, and dependence structure. In this thesis, we consider changes in both mean and variance. Assume that Δk_t follows a piece-wise AR (1) model. A structural change model with m breaking points can be written as:

$$\Delta k_t = \begin{cases} \alpha^{(1)} + \beta^{(1)}\Delta k_{t-1} + \epsilon_t^{(1)}, & t \leq t_1 \\ \alpha^{(2)} + \beta^{(2)}\Delta k_{t-1} + \epsilon_t^{(2)}, & t_1 < t \leq t_2 \\ \vdots & \\ \alpha^{(n+1)} + \beta^{(n+1)}\Delta k_{t-1} + \epsilon_t^{(n+1)}, & t > t_n \end{cases} \tag{10}$$

where t_j is the jth detected breaking point in Δk_t, and the error term $\epsilon_t^{(i)}$ are i.i.d.s that follow normal distribution with mean 0 and variance $\sigma^{(i)2}$. In Equation (10), n change points divide the time series into $n+1$ regimes.

3.4.2 Model Fitting

The following five-step procedure is applied to find the optimal number and locations of change points. A minimum number of 10% observations in each regime is required in the procedure.

1. Assume that there is only one change point. Find the location of change point that minimizes BIC.

2. Assume that there are two change points. Find the location of change points that minimize BIC.

3. Assume that there are three change points. Find the location of change points that minimize BIC.

4. Assume that there are four change points. Find the location of change points that minimize BIC.

5. Compare the BIC of the above four fitted structural change model, and choose the one with minimum BIC.

Table 8 presents the BIC for the structural change models with one, two, three, and four change points. The model with two change points has the lowest BIC; hence, it best fits the EW mortality data. The location of the two optimal change points for this model are 1952 and 1999.

The fitted model is as follows.

$$\Delta k_t = \begin{cases} -1.9361 - 0.2251\Delta k_{t-1} + \epsilon_t^{(1)}, & \text{if } \ t \leq 1952 \\ -1.7376 - 0.4554\Delta k_{t-1} + \epsilon_t^{(2)}, & \text{if } \ 1952 < t \leq 1999 \\ -5.1888 - 0.3145\Delta k_{t-1} + \epsilon_t^{(3)}, & \text{if } \ 1999 < t \leq 2011 \end{cases}$$

	AR (1)	1 Change Point	2 Change Points	3 Change Points	4 Change Points
BIC	762.5391	728.4645	726.3325	728.9403	733.7645

Table 8: BIC Values for Structural Change Models

where $\epsilon_t^{(1)}$ is i.i.d. that follows normal distribution with mean 0 and variance 97.6500, $\epsilon_t^{(2)}$ is i.i.d. normal with mean 0 and variance 7.3090. and $\epsilon_t^{(3)}$ is i.i.d. that following normal distribution with mean 0 and variance 7.643.

Figure 8 displays the unconditional mean for each regime as horizontal lines, and also shows the location of the two change points. The first regime has 46.85% observations, the second regime has 42.34% observations, and the third regime has 10.81% observations. Each regime has a different mean and a different variance. In particular, the third regime has a much lower mean compared to the other two regimes. The first regime has much higher variance because of catastrophic events that occurred in the first half of the 20th century.

In Table 9, we show the p-value of t-statistic testing the significance of AR coefficient. At 5% significance level, the AR coefficients for regime one and regime three are not significant, while that for regime two is significant.

	$\beta^{(1)}$	$\beta^{(2)}$	$\beta^{(3)}$
p-value	0.0997	0.0012	0.2064

Table 9: p-value for AR coefficients in Structural Change Model with Two Change Points

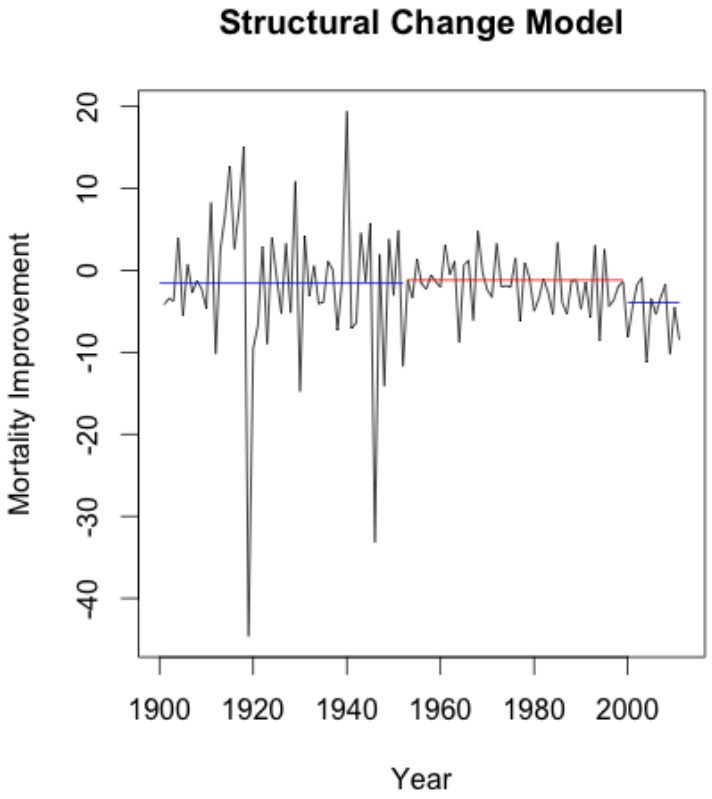


Figure 8: Fitting Structural Change Model for EW population.

Therefore, we reduce the structure of Δk_t in regimes one and three to random walk with drift. The fitted model is expressed by

$$\Delta k_t = \begin{cases} -0.1476 + \Delta k_{t-1} + \epsilon_t^{(1)}, & \text{if } \ t \leq 1952 \\ -1.7376 - 0.4554\Delta k_{t-1} + \epsilon_t^{(3)}, & \text{if } \ 1952 < t \leq 1999 \\ -0.0267 + \Delta k_{t-1} + \epsilon_t^{(4)}, & \text{if } \ 1999 < t \leq 2011 \end{cases}$$

where $\epsilon_t^{(1)}$ is i.i.d. normal with mean 0 and variance 261.4, $\epsilon_t^{(2)}$ is i.i.d. normal with mean 0 and variance 7.309, and $\epsilon_t^{(3)}$ is i.i.d. that follows normal distribution with mean 0 and variance 31.89.

Table 10 gives BIC for the original structural change model with three regimes and the reduced model, and the original structural change model has a lower BIC.

	SC-3R	SC-3R Reduced
BIC	726.3325	759.9648

Table 10: BIC Values for Structural Change Model and its Reduced Version

3.4.3 Model Interpretation

In the fitted structural change model, mortality improvement Δk_t are divided into three segments. The unconditional means for the three regimes are -1.5803, -1.193905, and -3.947362, respectively. The unconditional means imply that mortality improves in all three segments. In addition, the mortality improvement in regime three is the highest, which coincides with our observations of accelerating mortality improvement in the past decade. Before 1952, Δk_t is in the first regime, and mortality improvement is very volatile since it includes two major mortality jumps in 1920 and 1940. From 1952 to 1998, Δk_t is in the second regime, and it is relatively stable, as the second world war has ended and no events that have great impact on mortality occurred in this period. After 1999, Δk_t changes to the third regime, which has a much lower unconditional mean than the other two segments. Mortality improvement has accelerated in this regime due to medical advances and the strengthening health system.

3.5 ARCH/GARCH Model

3.5.1 Model Specification

ARCH proposed by Engle (1982) and GARCH model by Bollerslev (1986) are developed to characterize the nonlinear volatility of financial time series. ARCH/GARCH model "provides parsimonious parameterization for the conditional variance" (p.308), which is modeled as the linear regression of its long run mean, previous error terms, and the conditional variance in previous periods (Bera & Higgens, 1993). By modeling mortality improvement with ARCH/GARCH model, the time-varying volatility of mortality improvement can be captured. To examine whether it is appropriate to use ARCH/GARCH model, we first plot the squared residuals from fitted AR (1) model and study whether there have been some changes in variance.

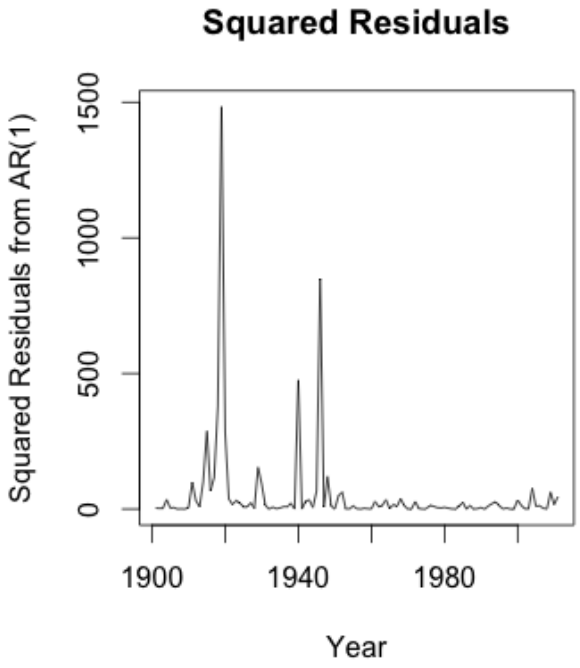


Figure 9: Squared Residuals from Fitted AR (1) Model for EW population.

Figure 9 displays the plot of the squared residuals obtained from fitting Δk_t to a AR (1) model. This figure shows a cluster of volatility in the squared residuals. Therefore, it is appropriate to use ARCH/GARCH model for the volatility of Δk_t.

Let ϵ_t be the error term in the AR (1) model, and $\epsilon_t = z_t\sigma_t$, where σ_t is the time-varying standard deviation, and z_t is a sequence of identically and independently distributed normal random variables with mean zero and variance one. An ARCH (q) model can be expressed as:

$$\sigma_t^2 = \alpha + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}^2 \tag{11}$$

where $\alpha > 0$, and $\beta_i > 0$. In an ARCH model, the variance of the error term at time t is a function of the squared error term at time $t-1$. As a result, a large error term trends to be followed by another large error, and a small error term tends to be followed by another small error. The order q is an indication of the length of time that a catastrophe influences the variance of subsequent error terms (Bera & Higgens, 1993).

A GARCH (p, q) model can be expressed as

$$\sigma_t^2 = \alpha + \sum_{i=1}^{q} \beta_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \gamma_i \sigma_{t-i}^2 \tag{12}$$

where $\alpha > 0$, $\beta_i > 0$ and $\gamma_i > 0$. The variance of error term at time t is conditional on the previous error terms and the previous variances. The order p is also an indication of the length of time that a variance of error term influences the variance of subsequent error terms.

Before fitting the mortality data by ARCH/GARCH model, we plot the ACF and PACF for the squared residuals to determine if residuals are correlated. Also, ACF

and PACF can identify the order of arch lags q. Figure 10 shows that both ACF and PACF have a spike at lag one. The significance level of ACF decays slowly after lag one, and that of PACF cuts off at lag one. Both suggest that the squared residuals are following an AR (1), meaning that the squared error term at time t can be effectively explained by the squared error term at time $t-1$. Therefore, $q=1$ should be used in mortality modeling.

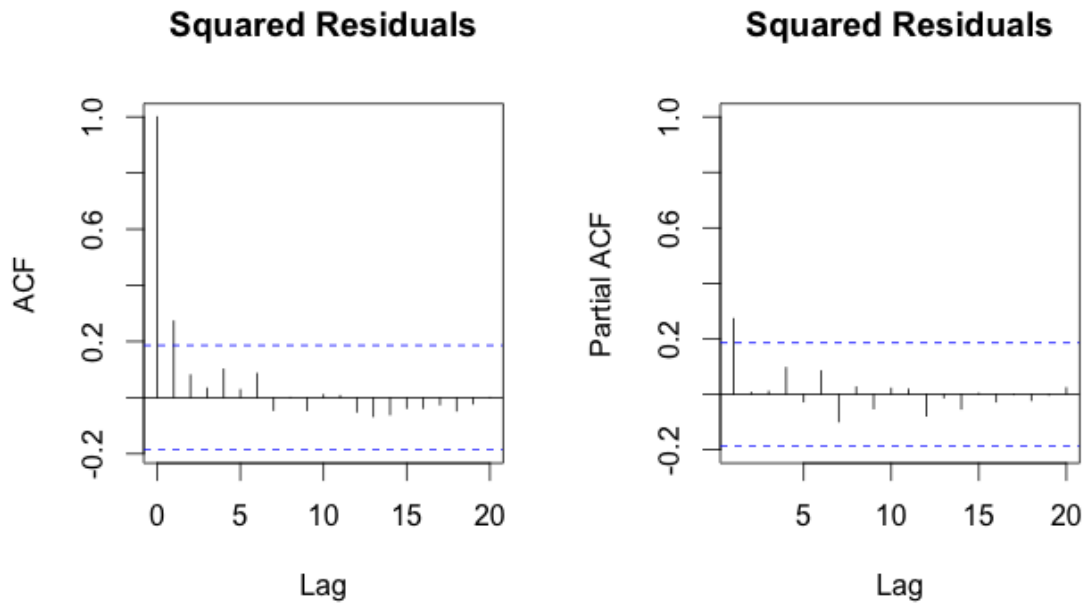


Figure 10: ACF and PACF of Squared Residuals from Fitted AR (1) Model for EW population.

3.5.2 Model Fitting

We obtain the parameter estimates of ARCH and GARCH models by MLE method with Quasi-Newton optimization algorithm. Since both ACF and PACF plots suggest q=1, we fit ARCH (1) and GARCH (1, 1) models. We choose GARCH term p=1 because it is the most commonly used. Generally, a low order GARCH model is preferred since it provides "parsimony and better numerical stability of estimation"

(p.126), and a higher order GARCH model may generate many local maxima and minima (Zivot, 2009). Let σ_t^2 be the conditional variance of Δk_t. The fitted ARCH (1) model is expressed as follows:

$$\sigma_t^2 = 29.2350 + 0.7906\epsilon_{t-1}^2$$

and the fitted GARCH (1, 1) model is written as

$$\sigma_t^2 = 23.58163 + 0.83414\epsilon_{t-1}^2 + 0.06261\sigma_{t-1}^2$$

Table 11 compares the BIC values of the two models. it shows that ARCH (1) has a lower BIC, and thus it fits EW mortality data better.

	ARCH (1)	GARCH (1,1)
BIC	749.2005	757.9698

Table 11: BIC Values for ARCH (1) and GARCH (1, 1) Model

We apply t-statistics to test the significance of parameter estimates for both models. For ARCH (1) model, the null hypothesis is β=0. For GARCH (1, 1) model, the null hypothesis is $\beta = 0$ and/or $\gamma = 0$.

p value	β	γ
ARCH(1)	4.96e-07	
GARCH(1,1)	0.0037	0.3617

Table 12: Parameter Significance for ARCH/GARCH Model

Table 12 presents p-values of the test. The p-value of ARCH (1) parameter is close

to 0, indicating that the null hypothesis is rejected and the ARCH term coefficient is significant. The p-value of the ARCH term in GARCH (1,1) is small, indicating that coefficient of ARCH term is significant. However, the p-value of GARCH term coefficient is high. Therefore, the GARCH term coefficient is not significant. Table 11 provides evidence that ARCH (1) model is more appropriate than GARCH (1, 1) model. This result is consistent with the results obtained from comparison of BIC values.

3.5.3 Model Interpretation

The ARCH model can be explained as:

> The constant term, 29.2350, indicates that the variance of the error term is on average 29.235. The variance of the error term at time t is influenced by the squared error term at time $t-1$. A large variance of the error term caused by a catastrophe tends to be followed by another large error term. For example, a large error term in 1918 that was caused by the flu pandemic gave rise to the higher variance of the error term in 1919. However, ARCH (1) model can only capture the fluctuation in variance, but not the change in the mean of mortality improvement.

3.6 Model Comparison

3.6.1 Goodness-of-fit

Table 13 summarizes the BIC values from fitting different time series models to EW mortality data. This table shows that all four nonlinear models have lower BIC

than linear AR (1) model. This suggests that nonlinear models fit significantly better than AR (1) model, and that the nonlinear trend in mortality improvement should be properly modeled.

BIC	One Regime	Two Regimes	Three Regimes	Four Regimes	Five Regimes
AR	762.5391				
ARCH	749.2005				
GARCH	757.9698				
TAR		755.8349	734.9055	745.8303	
MRSW		*715.7067*	736.5640	772.8863	
SC		728.4645	726.3325	728.9403	733.7645

*Number in italic indicates the lowest BIC value

Table 13: Goodness-of-fit: BIC Values for Linear and Nonlinear Models

Out of all the models we fitted, the Markov regime switching model with two regimes fits considerably better than the other models. There are several reasons that Markov regime switching model fits best. Markov regime switching models provide transparent representation of the status of structural changes in mortality. According to Milidonis et al. (2011), "Markov regime switching model has the flexibility to choose through the maximum likelihood, the switching time, and the duration as well as the value of the parameter estimate" (p. 267). All of these properties of Markov regime switching models contribute to good fitting for mortality data.

Interestingly, ARCH/GARCH model shows the least improvement among the nonlinear models we considered because this model only considers the time-varying property for volatility but not for mean. The accelerating mortality improvement in the past decade means that the mean value of Δk_t has changed. Unfortunately, ARCH/GARCH model cannot capture this change. Therefore, its goodness of fit is not as desirable as that of other models.

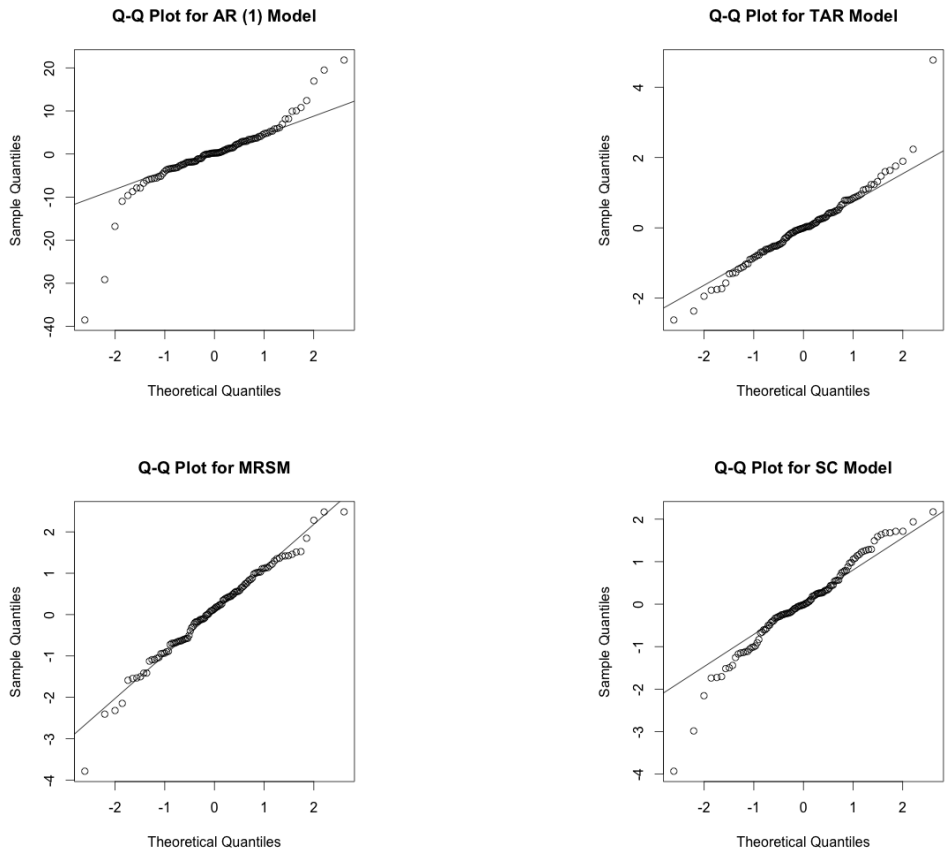


Figure 11: Q-Q plots of residuals from Different Fitted Models

To examine the normality of residuals from fitted nonlinear models, we draw Q-Q plots for different models respectively. Figure 11 shows the comparison of Q-Q plots of residuals for linear and nonlinear models. The Q-Q plots for nonlinear models show a stronger linear pattern than AR (1) model, which suggest that the residuals obtained from nonlinear models are more normally distributed than from AR (1) model.

3.6.2 Forecasting Performance

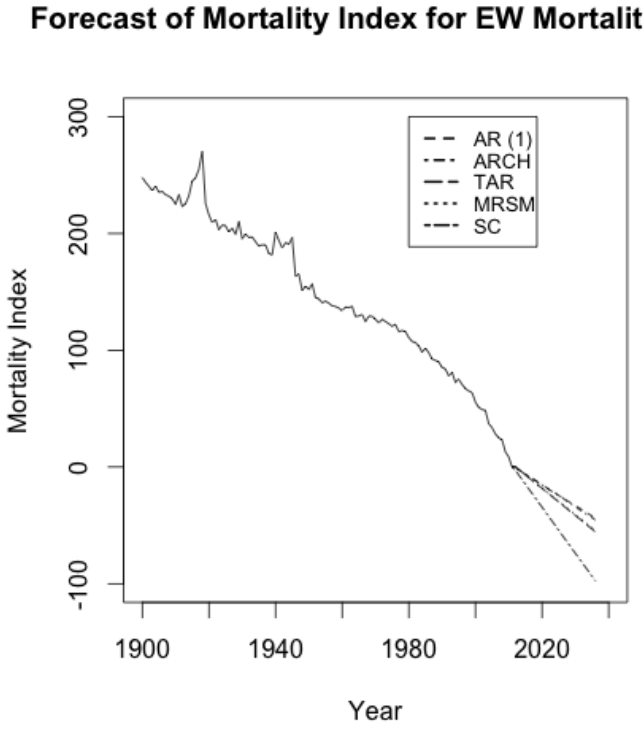
To compare the forecasting performance of different models, we use Monte Carlo simulation to obtain multiple future mortality paths.

The following procedure is used to forecast a 25 year mortality path:

1. Set $t = t_1 + 1$.
2. Simulate 5000 normal random numbers for the error term at time t.
3. Calculate the 5000 simulated mortality index k_t based on the time series model used.
4. Calculate the central death rates $m_{x,t}$ using Equation (1) together with fitted age parameters a_x, b_x, and simulated k_t for age x and year t.
5. Repeat Step 2-4 for $t = t_1 + 2, t_1 + 3, \cdots, t_1 + 25$

Figure 12 presents the mean forecasts of mortality index from different models. This figure shows that the mean forecast of ARCH model is much higher than that of other nonlinear models, because ARCH model cannot capture the change on the mean of Δk_t.

Figure 13 shows the 95% confidence interval of the forecasts from each model. Comparing panels 1-5, we find that ARCH model has the widest confidence intervals because this model has the property of volatility clustering, which causes higher variance in the mortality index forecasts.



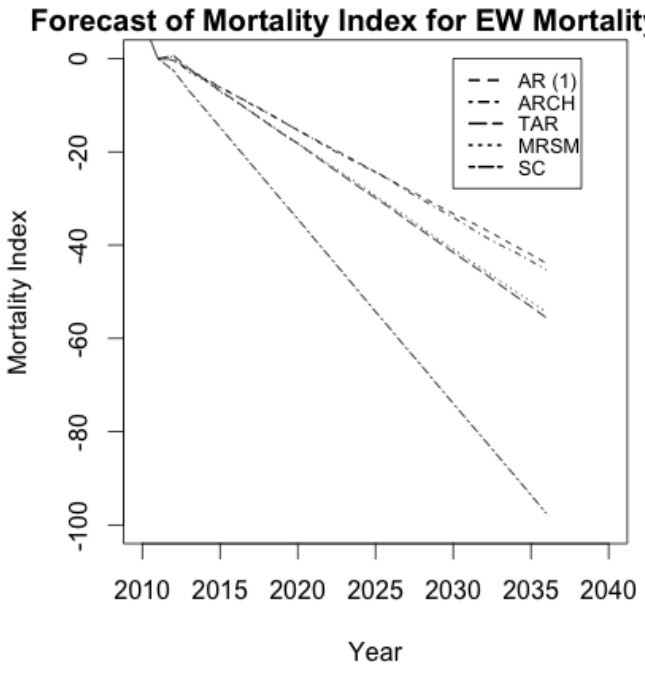


Figure 12: Mean Forecast Comparison of Mortality Index for EW Population

We also find that the mean forecasts of Δk_t by TAR model and Markov regime switching model are close, because these models have similar regime switching behaviours. In both models, Δk_t jumps between different regimes based on past observations. However, TAR model has wider confidence intervals. In TAR model, Δk_t jumps to other regimes more frequently, since the jump probability is either 0 or 1 and it is solely determined by the value of Δk_{t-1}. In Markov regime switching model, Δk_t tends to stay in the second regime, and it has a higher jump probability only when there is a shock to mortality. Therefore, the switching between different regimes is less active than in TAR model, and the simulated Δk_t has a narrower confidence interval.

Another finding evident in Figure 11 is that the mean forecast of structural change model is much lower than other models. This finding indicates the future mortality improvement projected by structural change model is more significant than other nonlinear models. In particular, the forecast of structural change model is solely determined by the third regime, which is the time series of Δk_t after year 1999. Due to the medical advances and strengthening health system in recent years, the mortality improvement in this period is faster. Therefore, the predicted mortality rates have a more significant downward trend.

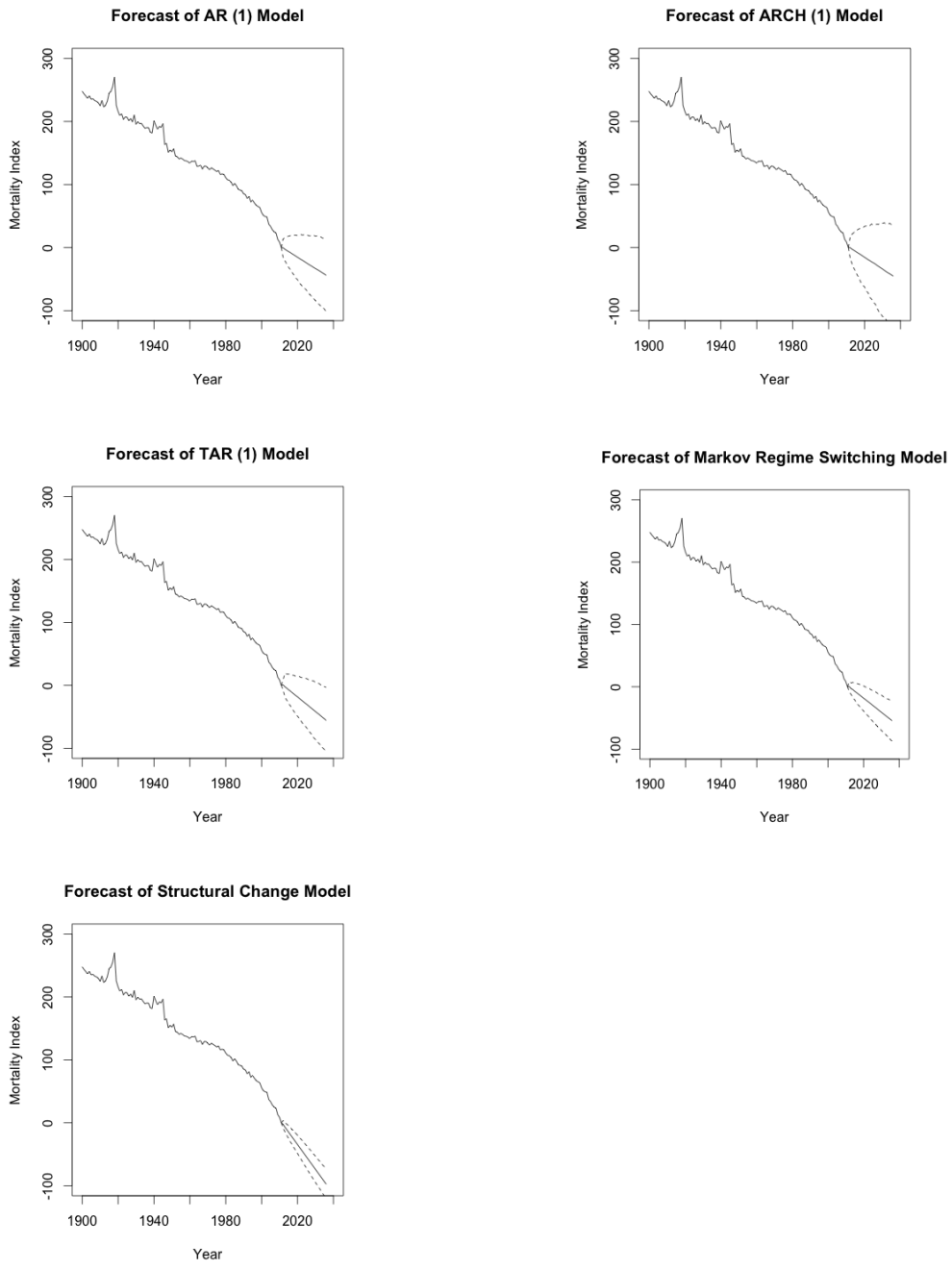


Figure 13: Forecasts and 95% Confidence Intervals of Mortality Index for Different Models

To better examine the forecasting performance, a 25 year in-sample forecast is obtained for four nonlinear models. 5000 mortality paths are simulated for time $t = t_1 - 24, \cdots, t_1$. Figure 14 shows the comparison of the in-sample forecast of mortality index and the realized mortality index for different models.

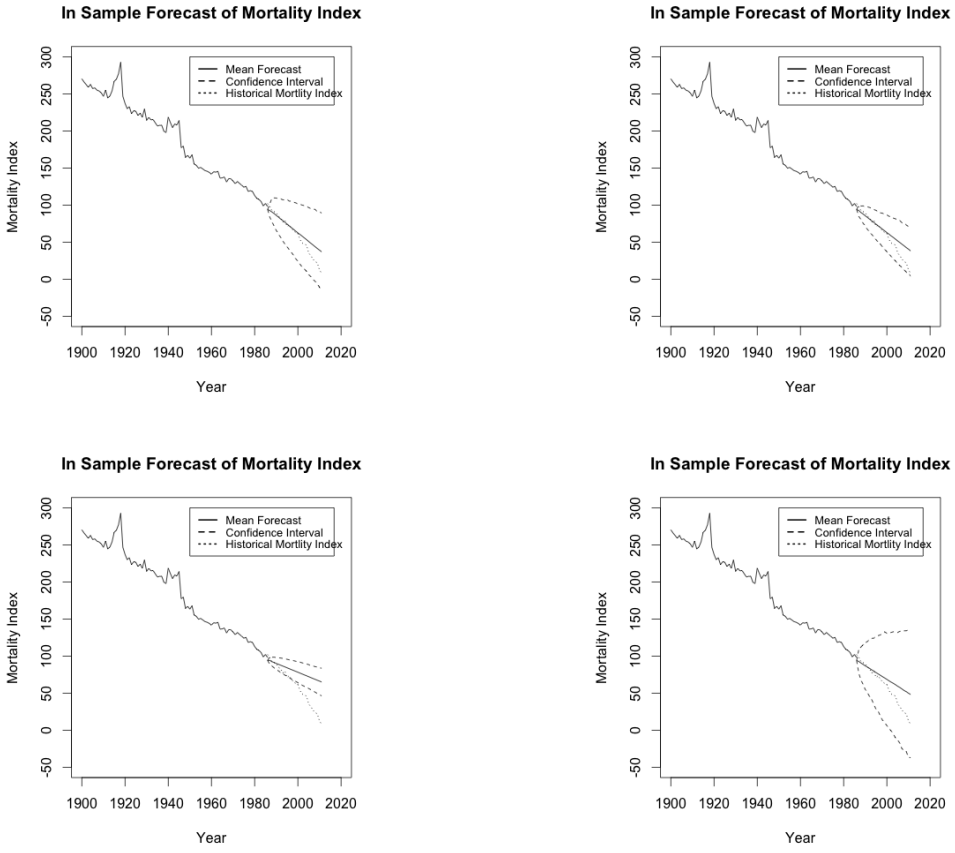


Figure 14: In-sample Forecasts of Mortality Index for Different Models

From Figure 14, we clearly observe that the predicted mortality index for TAR, Markov regime switching model and ARCH model are close to the realized mortality index. They diverge after 1999 is because of accelerating mortality improvement in past decade. Figure 14 provides evidence that the forecasts based on nonlinear models are significant. The in-sample forecast of structural change model shows the least

accuracy. The reason is that the in-sample forecast of mortality index projected by structural change model is solely determined by the second regime, which experienced a stable period with a less significant downward trend. However, the realized mortality index after 1999 are fitted to the third regime, which experienced a faster mortality improvement. Therefore the future mortality index projected by the third regime has a more significant downward trend.

4 Longevity Bond Pricing

4.1 Economic Pricing Approach

Economic pricing approach was first applied to mortality-linked securities by Zhou et al. (2015). The authors proposed considering the pricing problem from fundamental economic concepts of supply and demand. Buyers and suppliers are assumed to maximize their expected terminal utility. The market equilibrium is achieved at the price such that supply equals demand.

The setting for the longevity bond trading in this thesis is under hypothesis. Suppose we have two counterparties involved in the longevity bond trading. Counterparty A has life contingent liabilities, and counterparty B is an investor interested in the longevity bond for risk premiums. We assume that counterparty A sponsors a pension plan which contains 100 65-year-old pensioners and 100 66-year-old pensioners from EW population. This pension plan pays \$1 at the end of the year if a pensioner is still alive. The payment ends when the pensioner dies or reaches age 90. The pension liability for counterparty A at time t can be written as

$$f_t = 100 \prod_{i=0}^{t-1} (1 - q_{65+i,t_1+i}) + 100 \prod_{i=0}^{t-1} (1 - q_{66+i,t_1+i})$$

Counterparty B sells longevity bonds to counterparty A to earn risk premiums. The annual coupon of the bond payable at time t is the approximate survival rate of a 65-year-old cohort. The coupon payment at time t is

$$s_t = \prod_{i=0}^{t-1} (1 - m_{65+i,t_1+i})$$

Assume that P and Q are the price and quantity of the longevity bonds that A

and B both agree to trade. Figure 15 shows the cash flow of the transaction. At time t_1, counterparty A pays the price P for one unit of longevity bond purchased. At time t, counterparty A pays f_t to pensioners and receives the bond coupon s_t from counterparty B. When mortality of EW population improves, pensioners live longer and counterparty A is obligated to pay pension benefits for a longer period of time. At the same time, the bond payment becomes higher and offsets the increase of pension liability payments for counterparty A.

Figure 15: Cash Flow of a Longevity Bond Transaction

Let W_0^A and W_0^B be initial wealth of players A and B, and W_T^A and W_T^B be the wealth at the expiration of the bond. Zhou et al. (2015) assume that the only alternative investment is to deposit in a bank which earns a continuously compounding risk-free interest rate of r annually. The risk-free rate r is assumed to be 4.5 %. Given a longevity bond price P, Q^A and Q^B are quantities that agent A is willing to purchase and agent B has agreed to sell at time 0. We have the following equations:

$$W_T^A(P, Q^A) = W_0^A e^{rT} + Q^A e^{rT}(\sum_{t=1}^{T} s_t e^{-rt} - P) - \sum_{t=1}^{T} f_t e^{r(T-t)}$$

$$W_T^B(P, Q^B) = W_0^B e^{rT} + Q^B e^{rT}(P - \sum_{t=1}^{T} s_t e^{-rt})$$

Denote U^A and U^B as the utility functions of counterparties A and B, respectively.

Assuming exponential utility function, the terminal utility can be expressed as:

$$U^A(W_T^A) = 1 - e^{-c^A W_T^A}$$
$$U^B(W_T^B) = 1 - e^{-c^B W_T^B}$$

where c^A and c^B are the absolute risk aversion parameters for A and B, respectively. Since counterparty A purchases longevity bonds to hedge its longevity risk, and counterparty B takes on the longevity risk to earn premiums, it is reasonable to assume that counterparty A is more risk averse. We assume that c^A and c^B are 8×10^{-5} and 2×10^{-5}, respectively.

Let F_T represent the accumulated pension benefit payment at time T, and S_T represent the accumulated bond coupon payments at time T. We then have:

$$F_T = \sum_{t=1}^{T} f_t e^{r(T-t)}, \quad \text{and} \quad S_T = \sum_{t=1}^{T} s_t e^{r(T-t)}$$

Both counterparties maximize their terminal utilities. Therefore, given a price P, counterparties A and B maximize their expected terminal utility, and we have:

$$\begin{aligned} Q^A &= \text{argmax}_{Q^A} E[U^A(W_T^A(P, Q^A))] \\ Q^B &= \text{argmax}_{Q^B} E[U^B(W_T^B(P, Q^B))] \end{aligned} \tag{13}$$

First, we maximize the expected terminal utility for A. Recall that $W_T^A = W_0^A e^{rT} + Q^A e^{rT}(S_T - P) - F_T$. The conditions for maximizing the expected utility function of A are

$$\frac{\partial}{\partial Q^A} E(U^A(W_T^A(P, Q^A)) = 0, \quad \text{and} \quad \frac{\partial^2}{\partial^2 Q^A} E(U^A(W_T^A(P, Q^A)) < 0$$

The first condition can be rewritten as

$$E[c^A(S_T - e^{rT}P)e^{-c^A Q^A(S_T - e^{rT}P)+c^A F_T}] = 0$$

which implies

$$P = \frac{E[e^{-c^A Q^A S_T + c^A F_T} S_T]}{e^{rT} E[e^{-c^A Q^A S_T + c^A F_T}]}$$

Next, we maximize the expected utility for counterparty B. Recall that the terminal wealth of counterparty B is $W_T^B = W_0^B e^{rT} + Q^B e^{rT}(P - S_T)$. The conditions for maximizing the expected utility of B are

$$\frac{\partial}{\partial Q^B} E(U^B(W_T^B(P, Q^B)) = 0, \quad \text{and} \quad \frac{\partial^2}{\partial^2 Q^B} E(U^B(W_T^B(P, Q^B)) < 0$$

The first condition can be rewritten as

$$E[c^B(S_T - e^{rT}P)e^{c^B Q^B(S_T - e^{rT}P)}] = 0$$

which implies

$$P = \frac{E[e^{c^B Q^B S_T} S_T]}{e^{rT} E[e^{c^B Q^B S_T}]}$$

The market equilibrium is achieved when $Q^A = Q^B$. Therefore, we have

$$P = \frac{E[e^{-c^A Q S_T + c^A F_T} S_T]}{e^{r_f T} E[e^{-c^A Q S_T + c^A F_T}]} = \frac{E[e^{c^B Q S_T} S_T]}{e^{rT} E[e^{c^B Q S_T}]} \tag{14}$$

where P and Q are the equilibrium trading price and quantity

To calculate the expectations in Equation 14, we simulate 5000 mortality paths for each population and calculate function values based on these paths. The expectation is then obtained by taking the average of these simulated values.

4.2 Numerical Results

Taking Markov regime switching model as an example, we draw the supply and demand curves after maximizing expected utility. Figure 16 shows the supply and demand curves of the longevity bond. The intersection of the supply and demand curves is the market equilibrium. Counterparty A demands less longevity bonds when the price increases, and at the same time counterparty B is willing to sell more bonds.

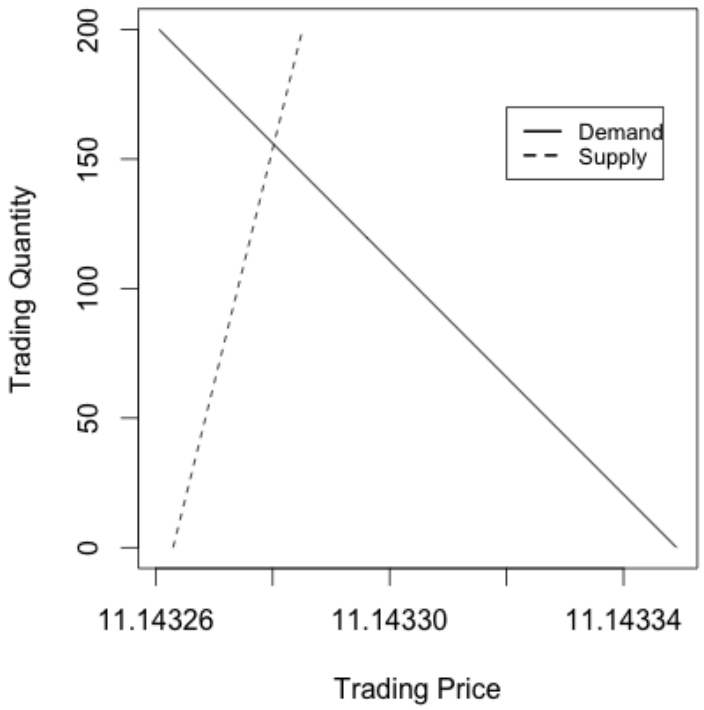


Figure 16: Supply and Demand Curves drawn Based on Markov Regime Switching Model

Table 14 summarizes the trading price, the quantity at market equilibrium, and the Expected Present Value (EPV) of total bond payments based on different mortality models. We have:

$$EPV = E[S_T e^{-rT}]$$

	AR	ARCH	TAR	MRSM	SC
Q	154.0294	153.5927	153.8952	154.2045	154.6909
P	11.0927	11.0893	11.1325	11.1209	11.3188
EPV	11.6030	11.5992	11.6447	11.6321	11.8397

*EPV indicates the expected present value of total bond payments

Table 14: Longevity Bond Price and Quantity at Market Equilibrium Obtained from Different Models

Trading prices obtained from linear AR (1) model and ARCH model are close and they are lower than the prices from other models. Recall that the mean forecasts of mortality index from the two models are very close in Figure 11. Therefore, their bond prices are close. In addition, their mean forecasts are significantly higher than other models, which means that they project slower mortality improvement and thus higher mortality rates. This results in lower survival rate, lower bond payments, and hence lower bond price.

TAR model and Markov regime switching model have similar bond prices because these models generate similar mean forecast for mortality index.

The price of longevity bond obtained through structural change model is much higher compared to the prices obtained from other models because the mortality improvement projected by structural change model is entirely based on the most recent segment, which has experienced a faster mortality improvement than that observed in history. The predicted cohort survival rate is much higher, and thus bond payment increases. The pension plan sponsor faces a longer period of pension benefit payout. Therefore, he/she is willing to pay a higher price to hedge the increased longevity risk.

EPV from different models shows that AR (1) and ARCH models generate lowest bond payments. In addition, TAR model and Markov regime switching model have similar bond payments, and structural change model has the highest bond payment. This result is consistent with the result obtained from the economic pricing method.

5 Conclusion

In this thesis, we examined the fitting and forecasting performance of several nonlinear time series models for modeling mortality improvement and studied their impacts on longevity bond pricing. Specifically, we first estimated the mortality index by fitting the Lee-Carter model to EW population mortality data using MLE method. The mortality data we used covered a sample age of 0-110+ and a sample period of 1900-2011. We then applied nonlinear models to the time series of the mortality index.

Several authors have considered nonlinear models in mortality index modeling (Li et al., 2011; Milidonis et al., 2010; Berkum et al., 2013). These authors have all shown that nonlinear models significantly improve model fitting. However, there has been no comparison between these models and it has been difficult to identify which model is more appropriate. This thesis fills the gap by examining the fitting and forecasting performance of four different nonlinear models: TAR model, Markov regime switching model, structural change model, and ARCH/GARCH model.

As expected, the fitting result shows that all four nonlinear models have lower BIC than AR (1) model, and hence nonlinear models significantly improve the mortality fitting. Interestingly, we find that Markov regime switching model with two regimes fits better than other nonlinear models, and ARCH/GARCH model offers the least improvement. Through simulation-based forecasting method, we obtained the mean forecast and prediction interval of future mortality index. We have three important findings from comparing the forecasts of different models. First, the mean forecast of ARCH model is much higher than other models because ARCH model can only capture the change in variance, but not the change of mean. This is also the reason why the goodness-of-fit for ARCH model is not desirable. Furthermore, part of the

mean change is viewed as random fluctuation. This enlarges the variance of residuals, and hence results in the widest prediction interval. Second, the mean forecasts of TAR model and Markov switching model are close, because mortality improvement switches between regimes are based on the immediate past observations in both two models. Third, the mean forecast of structural change model is much lower than other models. This is because the forecast of structural change model is solely determined by the last segment, and the mortality improvement has accelerated in recent years due to medical advances and a strengthening heath system.

We apply economic pricing method to explore the impact of different nonlinear models on longevity bond price. This pricing approach uses the fundamental economic concepts of supply and demand. To illustrate, consider that a longevity bond is traded between two counterparties. One counterparty is a pension plan sponsor who suffers from longevity risk and buys longevity bonds to hedge the risk. The other counterparty is an investor who takes on the longevity risk to earn risk premiums. ARCH model yields a similar bond price with linear AR (1) model. These are lower prices than prices obtained from other models, because the mean forecasts of AR (1) model and ARCH model are close. Both models cannot capture the change of mean mortality improvement and hence generate significantly higher mortality forecasts than other models. In contrast, structural change model gives a much higher bond price since its mortality forecast is based on the last regime which has experienced the fastest mortality improvement. TAR model and Markov regime switching model generate similar bond prices, since their mean forecast for mortality index are close. The variation of longevity bond pricing results shows that mortality modeling with non-linearity has a significant impact in longevity bond pricing.

This thesis verifies that nonlinear models lead to a significant improvement in mortality fitting. It further compares the fitting and forecasting of four different

nonlinear models. Although Markov regime switching model is the clear leader in goodness-of-fit, we should be cautious in applying it, as this model assumes that mortality switches between regimes. This highly volatile period experienced in the first half of the 20th century has a good chance of occurring again. In contrast, structural change model assumes that the mortality dynamic will follow the current trend until the next change point. However, this model provides no information regarding where the next change point will be. Therefore, the choice of nonlinear model is subjective and suffers from risk. Practitioners shall use expert judgment to determine which model is more suitable for their applications.

There are two venues of future work which can be explored.

1. In order to make the process of selecting a model easier, we can use the Bayesian method to determine the probability that the data follows a particular model. Presenting the probability of each model makes the decision more straightforward.

2. We considered only Lee-Carter model in this thesis. In future research, we may consider mortality models with cohort effects and examine if there is any non-linearity. We could investigate if the introduction of non-linearity improves fitting for these models.

Appendix A Maximizing the Log-likelihood of Lee-Carter Model

Let $D_{x,t}$ be the random variable that represents the number of deaths at age x and year t. Assume that

$$D_{x,t} \sim Poisson(\hat{m}_{x,t}E_{x,t})$$

The probability density function (pdf) of $D_{x,t}$ is

$$\frac{(\hat{m}_{x,t}E_{x,t})^{D_{x,t}}e^{(-\hat{m}_{x,t}E_{x,t})}}{D_{x,t}!}$$

Therefore, the log-likelihood of $D_{x,t}$ is

$$l = \sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1}[D_{x,t}\ln(\hat{m}_{x,t}E_{x,t}) - \hat{m}_{x,t}E_{x,t} - \ln(D_{x,t}!)]$$

Since $\ln(D_{x,t}!)$ and $D_{x,t}\ln(E_{x,t})$ are constants, and they do not depend on any unknown parameters, We can rewrite the log-likelihood as follows.

$$l = \sum_{x=x_0}^{x_1}\sum_{t=t_0}^{t_1}[D_{x,t}(a_x + b_x k_t) - e^{a_x + b_x k_t}E_{x,t}] + c$$

where c is a constant and $c = \ln(D_{x,t}!) + D_{x,t}\ln(E_{x,t})$.

We use the Newton-Raphson iterative process to maximize the log-likelihood function.

1. Set initial values: The initial value of a_x is set to the average log mortality rate for age x over the sample period. $a_x = \frac{1}{t_1-t_0+1}\sum_{t=t_0}^{t_1}\ln(m_{x,t})$. The initial value of b_x is set to $\frac{1}{x_1-x_0+1}$ to satisfy the constraint $\sum_{x=x_0}^{x_1} b_x = 1$. The initial value of k_t is set to random numbers generated from uniform distribution, and then $k_t = k_t - k_{t_1}$ to satisfy the constraint $k_{t_1} = 0$.

2. Update a_x: for $x = x_0, x_0+1, \cdots, x_1$, $\hat{a}_x = a_x - \frac{\partial l/\partial a_x}{\partial^2 l/\partial a_x^2}$, where $\frac{\partial l}{\partial a_x} = \sum_{t=t_0}^{t_1}(D_{x,t} - e^{a_x+b_xk_t}E_{x,t})$, and $\frac{\partial^2 l}{\partial^2 a_x} = \sum_{t=t_0}^{t_1}(-e^{a_x+b_xk_t}E_{x,t})$.

3. Update b_x: for $x = x_0, x_0+1, \cdots, x_1$, $\hat{b}_x = b_x - \frac{\partial l/\partial b_x}{\partial^2 l/\partial b_x^2}$, where $\frac{\partial l}{\partial b_x} = \sum_{t=t_0}^{t_1}(D_{x,t}k_t + -e^{a_x+b_xk_t}E_{x,t})$, and $\frac{\partial^2 l}{\partial^2 b_x} = \sum_{t=t_0}^{t_1}(-e^{a_x+b_xk_t}E_{x,t}k_t^2)$.

4. Update k_t: for $t = t_0, t_0+1, \cdots, t_1$, $\hat{k}_t = k_t - \frac{\partial l/\partial k_t}{\partial^2/\partial k_t^2}$, where $\frac{\partial l}{\partial k_t} = \sum_{x=x_0}^{x_1}(D_{x,t}b_x - e^{a_x+b_xk_t}E_{x,t}b_x)$, and $\frac{\partial^2 l}{\partial^2 l k_t} = \sum_{x=x_0}^{x_1}(-e^{a_x+b_xk_t}E_{x,t}b_x^2)$.

5. Apply the parameter identification constraints: we use $b_x = \frac{b_x}{\sum_{x=x_0}^{x_1}}$ and $k_t = k_t - k_{t_1}$.

6. Calculate the new log-likelihood $\hat{l}$, and the absolute log-likelihood difference $|\hat{l} - l|$.

7. Repeat step 3-6 until the absolute log-likelihood difference $|\hat{l} - l|$ is smaller than a prespecified tolerance value. We use 0.0001 as the tolerance value.

Appendix B Log-likelihood of Markov Regime Switching Model

To derive the log-likelihood of markov regime switching model, we first define the following notations:

$$\Phi_{j,t} = P(S_t = j \mid \Omega_t, \Theta)$$
$$p_{i,j} = P(S_t = j \mid S_{t-1} = i, \Theta)$$
$$\eta_{j,t} = f(\Delta k_t \mid S_t = j, \Omega_{t-1}, \Theta)$$

where f(.) is a genetic pdf. Since Δk_t follows a normal distribution, we have

$$\eta_{j,t} = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2\sigma_j^2}(\Delta k_t - \alpha^{(j)} - \beta^{(j)} \Delta k_{t-1})^2}$$

We start with a simple two-state regime switching model. The density function $f(\Delta k_t \mid \Omega_{t-1}, \Theta)$ can then be written as:

$$\begin{aligned} & f(\Delta k_t \mid \Omega_{t-1}, \Theta) \\ &= \sum_{j=1}^{2} \sum_{i=2}^{2} P(S_{t-1} = i \mid \Omega_{t-1}, \Theta) \times P(S_t = j \mid S_{t-1} = i, \Theta) \times f(\Delta k_t \mid S_t = j, \Omega_{t-1}, \Theta) \\ &= \sum_{j=1}^{2} \sum_{i=2}^{2} \Phi_{i,t-1} p_{i,j} \eta_{j,t} \end{aligned}$$

where $\Phi_{i,t-1}$ can be iteratively calculated using

$$\Phi_{j,t} = \frac{\sum_{i=1}^{2} p_{i,j} \Phi_{i,t-1} \eta_{j,t}}{f(\Delta k_t \mid \Omega, \Theta)}$$

Since we assume an AR (1) model in each regime, $f(\Delta k_t \mid \Omega_{t-1}, \Theta) = f(\Delta k_t \mid \Delta k_{t-1}, \Theta)$. The log-likelihood function can be written as

$$\ln f(\Delta k_{t_0+1} \mid \Theta) + \sum_{t=t_0+2}^{t_1} \ln[f(\Delta k_t \mid \Delta k_{t-1}, \Theta)$$

Let $\pi = (\pi_1, \pi_2)$ be the stationary distribution of the Markov Chain.

$$\pi_1 = p_{21}/(p_{12} + p_{21})$$

$$\pi_2 = p_{12}/(p_{12} + p_{21})$$

We then have

$$\begin{aligned}
& f(\Delta k_{t_0+1} \mid \Theta) \\
&= f(\Delta k_{t_0+1}, S_{t_0+1} = 1 \mid \Theta) + f(\Delta k_{t_0+1}, S_{t_0+1} = 2 \mid \Theta) \\
&= f(\Delta k_{t_0+1} \mid S_{t_0+1} = 1, \Theta) P(S_{t_0+1} = 1 \mid \Theta) + f(\Delta k_{t_0+1} \mid S_{t_0+1} = 2, \Theta) P(S_{t_0+1} = 2 \mid \Theta) \\
&= \pi_1 f(\Delta k_{t_0+1} \mid S_{t_0+1} = 1, \Theta) + \pi_2 f(\Delta k_{t_0+1} \mid S_{t_0+1} = 2, \Theta)
\end{aligned}$$

Similarly, for a n-regime switching model, the pdf of Δk_t is written as

$$\begin{aligned}
& f(\Delta k_t \mid \Omega_{t-1}, \Theta) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} P(S_{t-1} = i \mid \Omega_{t-1}, \Theta) \times P(S_t = j \mid S_{t-1} = i, \Theta) \times f(\Delta k_t \mid S_t = j, \Omega_{t-1}, \Theta) \\
&= \sum_{j=1}^{n} \sum_{i=1}^{n} \eta_{j,t} p_{i,j} \Phi_{i,t-1}
\end{aligned}$$

let $\pi = (\pi_1, \pi_2, \cdots, \pi_n)$ be the stationary distribution of the Markov Chain. We have

$$\begin{aligned} f(\Delta k_{t_0+1} \mid \Theta) &= \sum_{j=1}^{n} f(\Delta k_{t_0+1}, S_{t_0+1} = j \mid \Theta) \\ &= \sum_{j=1}^{n} f(\Delta k_{t_0+1} \mid S_{t_0+1} = j, \Theta) P(S_{t_0+1} = j \mid \Theta) \\ &= \sum_{j=1}^{n} \pi_j f(\Delta k_{t_0+1} \mid S_{t_0+1} = j, \Theta) \end{aligned}$$

The initial value of Φ will be

$$\Phi_{j,1} = \pi_j P(\Delta k_{t_0+1}, S_{t_0+1} = j \mid \Theta) / f(\Delta k_{t_0+1} \mid \Theta)$$

References

[1] Bera, A. K., & Higgins, M. L. (1993). ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys, 7*, 305-366.

[2] Berkum, F. V., Antonio, K., & Vellekoop, M. (2013). The Impact of Multiple Structural Changes on Mortality Predictions. *Scandinavian Actuarial Journal*, doi: 10.1080/03461238.2014.987807.

[3] Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics, 31*, 307-327.

[4] Mitchell, D., Brockett, P., Mendoza-Arriaga, R., & Muthuraman, K. (2013). Modeling and forecasting mortality rate. *Insurance: Mathematics and Economics, 52*, 275-285.

[5] Cairns, A. J. G., Blake, D., & Dowd, K. (2005). Pricing the Risk on Longevity Bonds. *Life and Pensions*, October, 41-44.

[6] Chen, H., MacMinn, R. D., & Sun, T. (2014). Multi-Population Mortality Models: A Factor Copula Approach. *Insurance: Mathematics and Economics*, Working Paper, Temple University.

[7] Engle, R. F (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50*, 987-1007.

[8] Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometria, 57*, 357-384.

[9] Hainaut, D. (2012). Multi-Dimensional Lee-Carter Model with Switching Mortality Processes. *Insurance: Mathematics and Economics, 50*, 236-246.

[10] Hollmann, F. W., Mulder, T. J., & Kallan, J. E. (2000). Methodology and Assumptions for the Population Projections of the United States 1999-2100. *Population Division Working Paper (Washington, D.C.), 38*, 1-24.

[11] Hong, T. S. (2011). *Essays on Forecasting Life Expectancy and Fiscal Sustainability* (Unpublished Doctoral Thesis). National University of Singapore, Singapore.

[12] Human Mortality Database (2014). Technical report, University of California, Berkeley (USA), and Max Planck Institute of Demographic Research (Germany). Retrieved from http://www.mortality.org or http://www.humanmortality.de.

[13] Lee, R., & Carter, L. (1992). Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association, 87*, 659-671.

[14] Levantesi, S., Menzietti, M., & Torri, T. (2008). Longevity Bond Pricing Models: an Application to the Italian Annuity Market and Pension Scheme. *Proceeding of the 18th International AFIR Colloquium*, Rome, Italy.

[15] Li, S. H., Chan, W. S., & Cheung, S. H. (2011). Structural Changes in the Lee-Carter Mortality Indexes: Detection and Implications. *North American Actuarial Journal, 15*, 13-31.

[16] Lewis, W. A. (1955). *The Theory of Economic Growth*. London: George Allen & Unwin.

[17] Lin, Y., & Cox, S. H. (2005). Securitization of Mortality Risks in Life Annuities. *Journal of Risk and Insurance, 72*, 227-252.

[18] Milidonis, A., Lin, Y., & Cox, S. H. (2011). Mortality Regimes and Pricing. *North American Actuarial Journal, 15*, 266-289.

[19] Mitchell, D., Brockett, P. L., Mendoza-Arriaga, R., & Muthuraman, K. (2013). Modeling and Forecasting Mortality Rates. *Insurance: Mathematics and Economics, 52*, 275–285.

[20] Renshaw, A. E., Haberman, S. (2003). Lee-Carter Mortality Forecasting with Age-specific Enhancement. *Insurance: Mathematics and Economics, 33* 255-272.

[21] Antolin, P. (2010). Longevity Risk and Private Pensions. In M. Micocci, G. N. Gregoriou, & G. B. Masala (Eds.), *Pension Fund Risk Management: Financial and Actuarial Modeling* (237-266). London: Taylor & Francis Group.

[22] Sweeting, P. J. (2011). A Trend-change Extension of the Cairns-Blake-Dowd Model. *Annals of Actuarial Science, 5*, 143-162.

[23] Tong, H. (1978). *On a Threshold Model in Pattern Recognition and Signal Processing*. Amsterdam: Sijhoff and Noordhoff.

[24] Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis*. New York: Springer-Verlag.

[25] Tong, H., & Lim, K. S. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society, Ser. B, 42,* 245-292.

[26] Wilmoth, J. R. (1993). Computational Methods for Fitting and Extrapolating the Lee-Carter Model of Mortality Change. Technical Report, University of California, Berkeley.

[27] Zhou, R., Li, J. S., & Tan, K. S. (2015). Economic Pricing of Mortality-Linked Securities: A Tâtonnement Approach. *Journal of Risk and Insurance, 82*, 65-96.

[28] Zivot, E. (2009). Practical Issues in the Analysis of Univariate GARCH Models. In T. G. Andersen, J. Kreiss, R. A. Davis, & T. V. Mikosch (Eds.), *Handbook of Financial Time Series* (113-155). Berlin Heidelberg: Springer-Verlag.