ARC Centre of Excellence in Population Ageing Research

Working Paper 2019/17

Bayesian hierarchical multi-population mortality modelling for China's provinces

Qian Lu, Katja Hanewald and Xiaojun Wang

Bayesian hierarchical multi-population mortality modelling for China's provinces

Qian Lu[1], Katja Hanewald[2], and Xiaojun Wang[1,*]

[1]*School of Statistics, Renmin University of China*

[2]*School of Risk and Actuarial Studies and CEPAR, UNSW Sydney*

20 November 2019

ABSTRACT

China has experienced large improvements in mortality rates, but there remain substantial variations at the provincial level. This paper develops new models to project mortality at both the national and provincial levels in China. We propose two models in a Bayesian hierarchical framework based on principal components and a random walk process, and compile a new comprehensive database containing mortality data for 31 provinces over the period 1982–2010. The baseline two-level model with a national–province hierarchy allows for information pooling across provinces, common national factors and consistency conditions. The extended three-level model with a national–region–province hierarchy pools information in the region and also allows for common factors within the region. Both models provide good estimates and reasonable forecasts for China and its provinces. The baseline two-level model provides good fit and reasonable forecasts with equal width intervals for the provinces. The three-level model has a better fit with a lower deviance information criterion and provides forecast intervals reflecting regional uncertainty. The sensitivity analyses show that the forecasts are robust when changing the trend assumptions and regional groups.

JEL classification: C11; C13; C53; G22; J11

Keywords: Mortality modelling, Bayesian framework, Hierarchical models, Coherent mortality projection, China

---

[*]Corresponding author. Email address: xiaojun_wang@ruc.edu.cn. Postal address: School of Statistics, Renmin University of China, No. 59 Zhongguancun Street, Haidian District Beijing, 100872, P.R. China.

1. Introduction

China has experienced a steady decline in mortality rates and great improvements in life expectancy, with life expectancy at birth increasing from 43.8 years in 1950–1955 to 76.6 years in 2015–2020 (United Nations, 2019). But there are still large variations at the provincial level, and more developed provinces have much longer life expectancies than less developed provinces. For example, life expectancy in Beijing was 83.5 years in 2015, while it was 69.4 years in Qinghai (Zhou et al., 2016). Thus, statistics at the national level do not reflect actual mortality and life expectancy at the provincial level. Therefore, the provincial mortality variations need to be modelled. This paper proposes new methods to model province-level variations in mortality in China in a Bayesian framework, and provides good estimates and reasonable forecasts for China and its provinces based on a new large mortality database.

A growing literature focuses on mortality models for multiple populations. Li and Lee (2005) propose a "coherent multi-population mortality model" to model the mortality rates of several populations so that they do not cross over or diverge. Li and Lee's (2005) model is an extension of the classic age-period-cohort (APC) model, which has seen several extensions (e.g., Dowd et al., 2011; Kleinow, 2015; Danesi et al., 2015). New stochastic methods are also used in the modelling of several related populations. For example, De Beer (2012) use the TOPALS (tool for projecting age-specific rates using linear splines) model; Hyndman et al. (2013) apply the product-ratio functional method; and Bergeron-Boucher et al. (2017) employ the compositional data analysis method in multi-population mortality modelling.

The development of coherent mortality models has contributed to the emergence of subnational mortality modelling. Some researchers, such as Hyndman et al. (2013) and Gonzaga and Schmertmann (2016) use multi-population models to model mortality in subnational areas. Alternatively, Giannakouris (2010), Smith et al. (2013) and the Office for National Statistics (ONS, 2016) compute subnational mortality via simple relationships to national mortality forecasts. Other recent studies use Bayesian methods to model subnational mortality. For example, Alexander et al. (2017) and Schmertmann and Gonzaga (2018) employ the Bayesian method to correct for missing values and underreported deaths in subnational mortality modelling.

There is a growing interest in China's mortality rates in the literature. Banister and Hill (2004) study long-term mortality trends in China using Chinese census data for 1982, 1990 and 2000. Zhao (2012) and Zhao et al. (2013) apply a modified Lee–Carter model based on census and sample survey data for China. Huang and Browne (2017) and Huang (2017) modify the continuous mortality investigation method of the UK's Institute and Faculty of Actuaries and apply it to China's mortality using data from 1997–2011. Apart from stochastic modelling, Bayesian methods have also been introduced into mortality modelling in China. Li (2014) uses a Bayesian framework to estimate the Lee–Carter model with limited data, and Li et al. (2019) use a Bayesian framework and sequential Kalman filter to model Chinese mortality. But none of the studies model mortality at the subnational or province level in China. One reason for this is the shortage of data on provincial mortality.

This paper focuses on provincial mortality modelling in China and proposes a baseline two-level model and an extended three-level model in a Bayesian framework. Our study is based on a new comprehensive database for 31 Chinese provinces over the period 1982–2010, which we compiled using online and archive resources. The baseline two-level model uses China and 31 provinces as its two levels and considers information pooling, common national factors and consistency conditions between China and its provinces. The three-level model introduces a regional level between China and its provinces, pools information in regions and allows for both common regional factors and common national factors. Both models provide good fit and reasonable forecasts for China and its provinces. The baseline two-level model has equal width intervals for provinces, while the three-level model has intervals of different widths for provinces, thereby reflecting different regional uncertainties. Compared to the baseline model, the three-level model has a better fit with a lower deviance information criterion (DIC). The sensitivity analyses show that the forecasts are relatively robust to the trend assumptions and the methods for grouping provinces into regions.

We contribute to the literature in two ways. First, we propose a new three-level hierarchical mortality model in a Bayesian framework allowing for information pooling with common factors. Second, we study provincial mortality in China and model provincial and national mortality in China together. In so doing, we obtain coherent and consistent forecasts for China and its provinces. Our results inform the development of China's emerging market for private annuities and the ongoing reforms of China's complex pension system (see, e.g., Fang and Feng, 2018; Wang and Li, 2018).

The remainder of this paper is organized as follows. Section 2 describes the data we use. Section 3 introduces the two models we propose. Section 4 displays and compares the results of the models. Section 5 provides some conclusions and ideas for future research.

2. Data

China has conducted six censuses to date (1953, 1964, 1982, 1990, 2000 and 2010). The post-1982 censuses were carried out by the National Bureau of Statistics (NBS) and included populations and deaths by age and sex for every province in China.[1]

We obtained the 1982 and 1990 census data from hard copies contained in the NBS archives. The 2000 and 2010 censuses data are from the NBS website (NBS, 2002; NBS, 2012). In addition to the censuses, China conducted four 1% sample surveys in 1987, 1995, 2005 and 2015. The 2005 and 2015 surveys contain the populations and deaths for China and its provinces. These data were obtained from the 1% Sample Survey Materials in China (NBS, 2006; 2016). As the sample ratios are only around 1%, the sample sizes at provincial level are small, and the data are more volatile compared to the census data.

We use census data after 1982 for provinces in our model and sample survey data in 2005 and 2015 as a reference for the in-sample and out-of-sample forecasts.[2] Table 1 provides an overview of the data employed.

Table 1. Data employed.

Year	1982	1990	2000	2005	2010	2015
No. of provinces	29	30	31	31	31	31
Survey type	Census	Census	Census	1% sample	Census	1% sample

Notes: No. of provinces reports the number of provinces for which data were available. The census data were used to estimate the model. Sample survey data for 2005 and 2015 are used as references for in-sample and out-of-sample forecasts.

The data in the census and sample survey materials are available in 5-year age groups for all ages until 90+ years old.[3] We gathered data in the 0, 1–4, 5–9, 10–14, …, 85–89 and 90+ age groups. We set 90+ as the highest age group because it is the highest age group of deaths in 1990, and the data are sparse in other census years beyond that age group.

As the census enumeration is the end-year population data surviving from the risk of death in the year prior to the census (Cai, 2005), and the central age-specific mortality rate is calculated based on mid-year population, we approximate the mid-year population and account

[1] The NBS sums all the provincial data to obtain the national data for China. The 1982 census materials have populations and deaths by age and sex for China and every province except Hainan (not established then), Chongqing (not established then) and Tibet (only population data, no deaths data). The 1990 census materials have populations and deaths by age and sex for China and every province except Chongqing province (not established then).

[2] We use the raw data in the census and sample survey materials to retain the characteristics and information in the data.

[3] The 1982 census data were collected for each age (0, 1, 2, 3, 4, 5, …) and include highest ages ranging from 105 years old to 140 years old for different provinces. Census data after 1990 were collected in 5-year age groups with data for every age and the highest age being 100+ years old (0–4, 0, 1, 2 3, 4, 5-9, 5, 6, 7, 8, 9, …) except in 1990, where the highest ages were 90+ years old.

for deaths reported for the 12 months prior to the census:

$$ {}_nP_x^m = {}_nP_x + \frac{{}_na_x}{n} \cdot {}_nD_x \ , \tag{1} $$

where ${}_nP_x$ and ${}_nP_x^m$ are the census population and mid-year population between age x and $x+n$ in the census year; ${}_nD_x$ are the observed deaths between age x and $x+n$ in the census year; and $\frac{{}_na_x}{n}$ is the average number of years lived in the year prior to the census for those who died in the year prior to the census.

As in Cai (2005), ${}_na_x$ equals $\frac{n}{2}$ for all ages except for the age group 0 and 1–4. ${}_1a_0$ and ${}_4a_1$ are chosen according to the adapted Coale and Demeny formula (Coale et al., 1983; Preston et al., 2000), as shown in Table 2. ${}_1m_0$ can be obtained by solving the equation:

$$ {}_1m_0 = \frac{{}_1D_0}{{}_1P_0 + (\alpha + \beta \cdot {}_1m_0)\,{}_1D_0} \tag{2} $$

Table 2. Coale and Demeny formula for male.

${}_1a_0$	0.330	if ${}_1m_0 \geq 0.107$,
	$0.045+2.684\,{}_1m_0$	otherwise
${}_4a_1$	1.352	if ${}_1m_0 \geq 0.107$,
	$1.651-2.816\,{}_1m_0$	otherwise

3. Method

An emerging literature uses the Bayesian hierarchical framework for subnational mortality modelling (e.g., Alexander et al., 2017; Schmertmann and Gonzaga, 2018). Alexander et al. (2017) emphasize the advantages of the Bayesian framework in dealing with missing data and develop a model constructed by principal components at the subnational level in a Bayesian hierarchical framework for the United States and France. As data for some provinces are missing and our dataset is relatively limited, we use the Bayesian framework to correct for the missing and limited data. Furthermore, because the model constructed by principal components is parsimonious and easy to construct, we apply and extend Alexander et al.'s (2017) modelling approach. In a first step, we modify Alexander et al.'s (2017) model to a baseline two-level model for China and its provinces with common national factors. Then we propose a new extended three-level hierarchical model with additional common factors within the region. The baseline two-level model pools and shares information across all the provinces, and the three-level model pools and shares information in the same region.

3.1. Two-level model

The baseline two-level model uses China and its provinces for the two levels. Let $D_{x,t}^{i}$ denote deaths at age *x* in province *i* at time *t*, and let $D_{x,t}^{China}$ denote deaths at age *x* in China at time *t*. We assume that the deaths are from the Poisson distributions:

$$D_{x,t}^{i} \sim Poisson(m_{x,t}^{i} \cdot P_{x,t}^{i}),\ \ D_{x,t}^{China} \sim Poisson(m_{x,t}^{China} \cdot P_{x,t}^{China})\ \ , \tag{3}$$

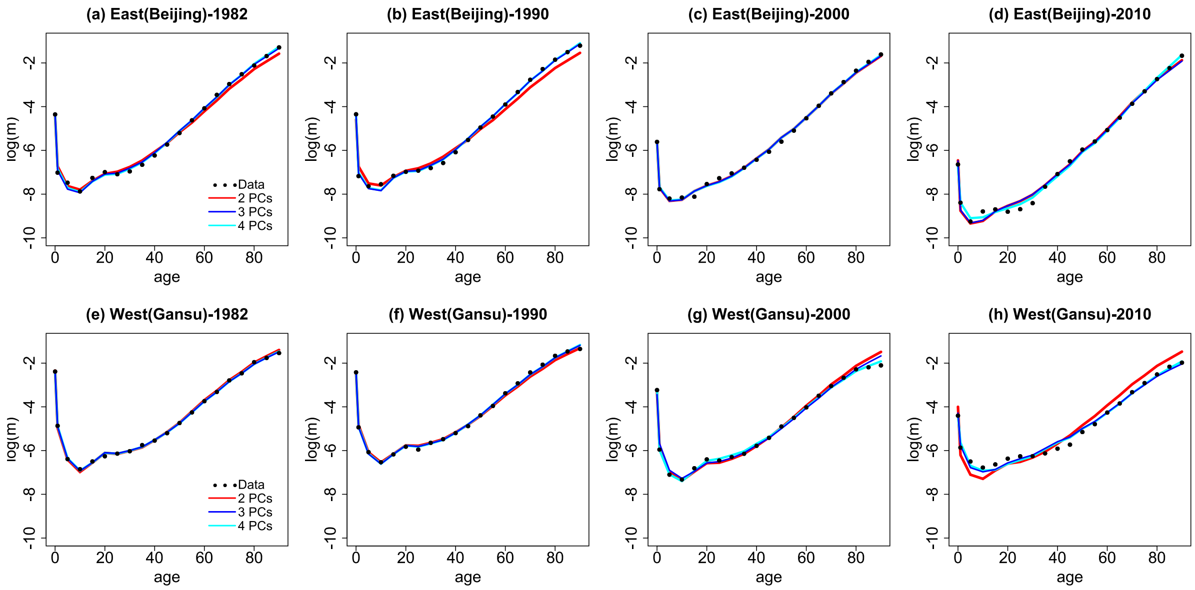
where $m_{x,t}^{i}$ is the mortality rate and $P_{x,t}^{i}$ is the population at risk at age *x* and time *t* in province *i*, while $m_{x,t}^{China}$ and $P_{x,t}^{China}$ are the same variables at the national level. As the NBS computes the data for China by summing the provincial data, we apply the following consistency conditions to obtain the data for China and to ensure that the mortality rates at the national and provincial levels are consistent:

$$\sum D_{x,t}^{i} = D_{x,t}^{China},\ \sum P_{x,t}^{i} = P_{x,t}^{China}\ \ . \tag{4}$$

Alexander et al. (2017) show that the model constructed by principal components is flexible and fits the data well. As the model with principal components is parsimonious and easy to construct, we also use a functional form based on principal components.

Alexander et al. (2017) suggest using the first three principal components because higher-order principal components would pick up the residual variance in the data. In our analysis, as shown in Figure 1, a model with two principal components (red lines) either underestimates or overestimates mortality rates, whereas three principal components (blue lines) reproduce mortality curves to a satisfactory level. To avoid underfitting and overfitting, we use the first three principal components in our models.

Figure 1. Fitted mortality for different numbers of principal components.



Notes: The dots indicate the historical mortality rates, the red lines represent fitted mortality based on two principal components, the blue lines represent fitted mortality based on three principal components, and the cyan lines represent fitted mortality based on four principal components.

In multi-population mortality modelling, common factors are usually used to maintain coherence, such as in Li and Lee (2005) and Kleinow (2015). Similar to the common factor models, we propose to model $m_{x,t}^{i}$ and $m_{x,t}^{China}$ on the log scale with province-level and common principal components:

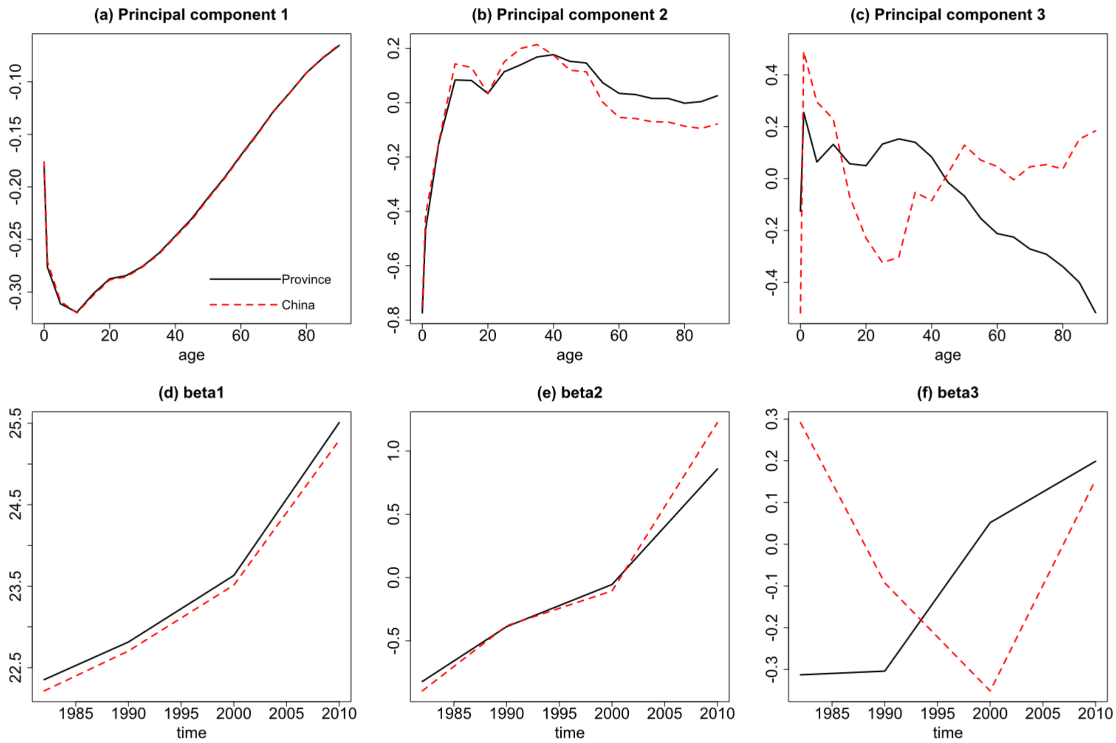
$$\log(m_{x,t}^{i}) = \beta_{1,t}^{i} Y_1^{i} + \beta_{2,t} Y_2 + \beta_{3,t} Y_3 + \varepsilon_{i,t}^{i}, \tag{5}$$

$$\log(m_{x,t}^{China}) = \beta_{1,t}^{China} Y_1^{China} + \beta_{2,t} Y_2 + \beta_{3,t} Y_3 + \varepsilon_{i,t}^{China}, \tag{6}$$

where Y_1^{i} is the first principal component for the individual province, Y_1^{China} is the first principal component for China, and Y_2 and Y_3 are the common principal components for all provinces. $\beta_{1,t}^{i}$, $\beta_{1,t}^{China}$ and $\beta_{p,t}$ ($p = 2, 3$) are the corresponding coefficients of each principal component at time t ($t = 1$–4). Our analysis shows that the first principal component and coefficient are crucial for good data fit. To account for the different mortality characteristics across provinces we propose using the first factors $\beta_{1,t}^{i}$ and Y_1^{i} as province-level factors to capture the provincial patterns, and use the second and third factors $\beta_{p,t}$ and Y_p ($p = 2, 3$) as common national factors to capture the common patterns of all the provinces and maintain coherence. By using the common national factors, all provinces share the same information.

Figure 2 shows the average principal components and coefficients for China (Y_p^{China}, $\beta_{p,t}^{China}$, $p=1, 2, 3$) and the provinces ($\bar{Y}_p$, $\bar{\beta}_{p,t}$, $p=1, 2, 3$), computed by singular value decomposition (SVD) of the mortality matrices for both. $\bar{Y}_p$ and Y_p^{China} describe the age patterns of mortality. Figure 2 shows that $\bar{Y}_p$ and Y_p^{China} have similar shapes and values, which means that China and its provinces have similar age patterns. $\bar{\beta}_{p,t}$ and $\beta_{p,t}^{China}$ describe the average time trend of mortality. $\bar{\beta}_{p,t}$ and $\beta_{p,t}^{China}$ ($p=1, 2$) have similar trends but slightly different levels, which shows that the mortality of China and its provinces have similar time trends with slightly different slopes. $\beta_{3,t}^{China}$ is more volatile because Y_3^{China} fluctuates more than the first two principal components, as shown in Figure 2(c).

Figure 2. Principal components and coefficients for China and its provinces.



Notes: The black lines show the principal components and average coefficients of the provinces. The red dashed lines are the principal components and coefficients of China. The principal components shown are $\bar{Y}_p$ and Y_p^{China}, and the coefficients are $\bar{\beta}_{p,t}$ and $\beta_{p,t}^{China}$.

We use informative priors to estimate the parameters in the models. We estimate the Y_1^{China} by drawing from the following informative normal distribution:

$$Y_1^{China} \sim N(Y_1^{China,I}, \sigma_{y1}^2), \tag{7}$$

where $Y_1^{China,I}$ is the expectation of the informative prior, which is the first principal component computed by SVD of the mortality matrix of China.

The parameters of China and its provinces in Figure 2 have similar shapes and values. Therefore, it is reasonable for us to model the province-level Y_1^i by borrowing information from China as follows:

$$Y_1^i \sim N(Y_1^{China}, \sigma_{y1}^2), \tag{8}$$

where Y_1^i is drawn from a normal distribution with a mean Y_1^{China}. The common Y_2 and Y_3 for China and its provinces are drawn from the following distribution:

$$Y_p \sim N(Y_p^{China,I}, \sigma_{yp}^2),\ \ p = 2,3. \tag{9}$$

where $Y_p^{China,I}$ is the pth (p = 2, 3) principal component computed by SVD of the mortality matrix of China.

$\beta_{1,t}^{China}$ is estimated similarly by drawing from a normal distribution; Alexander et al. (2017) assume that the coefficients share information across geographic areas. Similarly, we allow $\beta_{1,t}^i$ to pool information across all provinces:

$$\beta_{1,t}^i \sim N(\bar{\beta}_{1,t}, \sigma_t^2), \tag{10}$$

where $\bar{\beta}_{1,t}$ is the population-weighted mean of the first coefficient computed by SVD of the mortality matrix of provinces. $\bar{\beta}_{1,t}$ contains the overall information of all 31 provinces, which implies that the $\beta_{1,t}^i$ pools information from other provinces. Owing to population weighting, provinces with larger populations have a bigger effect on $\bar{\beta}_{1,t}$.

The common $\beta_{2,t}$ and $\beta_{3,t}$ for China and its provinces are estimated similarly as follows:

$$\beta_{p,t} \sim N(\bar{\beta}_{p,t}, \sigma_{p,t}^2),\ p = 2,\ 3. \tag{11}$$

where $\bar{\beta}_{p,t}$ is the population-weighted mean of the pth (p = 2, 3) coefficient computed by SVD of the mortality matrix of provinces.

The variances are assigned conjugated inverse gamma (IG) priors, which are commonly used as, for example, in Kogure et al. (2009), Li (2014) and Khana et al. (2018):

$$\sigma_{yp}^2 \sim IG(1, 0.001),\ \ \sigma_t^2 \sim IG(1, 0.001),\ \ \sigma_{c,t}^2 \sim IG(1, 0.001),\ \ \sigma_{p,t}^2 \sim IG(1, 0.001).$$

$\varepsilon_{i,t}^i$ and $\varepsilon_{i,t}^{China}$ are normally distributed random error terms:

$$\varepsilon_{i,t}^i \sim N(0, \sigma_{\varepsilon,i}^2),\ \ \varepsilon_{i,t}^{China} \sim N(0, \sigma_{\varepsilon,China}^2). \tag{12}$$

As the provincial data are limited, we use a simple random walk process to forecast future values of $\hat{\beta}^i_{p,t+n}$. The random walk process is commonly used to describe the dynamics of the period effects in multi-population and Bayesian mortality models, such as in Li and Lee (2005) and Czado et al. (2005). Here we consider the temporal and provincial uncertainty and draw $\hat{\beta}^i_{1,t+n}$ from a normal distribution:

$$\hat{\beta}^i_{1,t+n} \sim N(\beta^i_{1,t+n}, \sigma^2_{\beta_{t+n}}),\ \ \beta^i_{1,t+n} = \beta^i_{1,t} + n \cdot d^i_1 , \tag{13}$$

$$\hat{\beta}^{China}_{1,t+n} \sim N(\beta^{China}_{1,t+n}, \sigma^2_{\beta c_{t+n}}),\ \ \beta^{China}_{1,t+n} = \beta^{China}_{1,t} + n \cdot d^c_1 , \tag{14}$$

where n is the number of future years to forecast; and d^i_1 and d^c_1 are the drifts of the random walk processes, which are computed by the historical change:

$$d^i_1 = (\beta^i_{1,4} - \beta^i_{1,1}) / T\ , d^c_1 = (\beta^{China}_{1,4} - \beta^{China}_{1,1}) / T\ , \tag{15}$$

where T is the number of years covered in the data, with $T = 28$ from 1982–2010. We use the same uncertainty to make forecasts for $\hat{\beta}^i_{1,t+n}$ and $\hat{\beta}^{China}_{1,t+n}$. The variances $\sigma^2_{\beta_{t+n}}$ and $\sigma^2_{\beta c_{t+n}}$ are calculated by:

$$\sigma^2_{\beta_{t+n}} = Var(\beta^i_{1,4}) + \frac{n^2(Var(\beta^i_{1,4}) + Var(\beta^i_{1,1}))}{T^2},\ \ \sigma^2_{\beta c_{t+n}} = \sigma^2_{\beta_{t+n}} , \tag{16}$$

where $Var(\beta^i_{1,1}) = Var(\beta^{1:m}_{1,1})$ and $Var(\beta^i_{1,4}) = Var(\beta^{1:m}_{1,4})$ are posterior variances of all provincial $\beta^i_{1,1}$ and $\beta^i_{1,4}$, and m is the number of provinces (m = 31).

As $\beta_{2,t}$ and $\beta_{3,t}$ are common, there will be no provincial uncertainty in them. $\hat{\beta}_{2,t+n}$ and $\hat{\beta}_{3,t+n}$ are forecasted by:

$$\hat{\beta}_{p,t+n} = \beta_{p,t} + n \cdot d_p,\ \ d_p = (\beta_{p,4} - \beta_{p,1}) / T,\ \ p = 2,3 . \tag{17}$$

The future mortality rates of the provinces and China are forecasted by:

$$\log(\hat{m}^i_{x,t+n}) = \hat{\beta}^i_{1,t+n} Y^i_1 + \hat{\beta}_{2,t+n} Y_2 + \hat{\beta}_{3,t+n} Y_3\ , \tag{18}$$

$$\log(\hat{m}^{China}_{x,t+n}) = \hat{\beta}^{China}_{1,t+n} Y^{China}_1 + \hat{\beta}_{2,t+n} Y_2 + \hat{\beta}_{3,t+n} Y_3\ . \tag{19}$$

3.2. Three-level model

The two-level model above pools the information and uses common national factors across all the provinces. However, this common-factor structure may be questioned given that China's provinces vary greatly in their economic development and mortality. Therefore, we introduce a regional level and propose a three-level model to account for the regional mortality differences.

According to the Development Research Center of the State Council (DRC, 2005), China can be divided into four economic regions. NBS (2011) defines these four economic regions as follows: east (Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan); central (Shanxi, Anhui, Jiangxi, Henan, Hubei and Hunan); west (Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia and Xinjiang); and northeast (Liaoning, Jilin and Heilongjiang). Life expectancies are similar within each region but vary across regions (Zhou et al., 2016). For example, the provinces in the eastern region have longer life expectancies than those in central and west China, which makes the regions good representatives of mortality differences. We introduce a regional level in the hierarchy between China and the provinces according to the DRC and NBS definitions and use common factors within the same region.

In this three-level model, the mortality rates are modelled as follows:

$$\log(m_{x,t}^{i}) = \beta_{1,t}^{i} Y_1^{i} + \beta_{2,t}^{k} Y_2^{k} + \beta_{3,t} Y_3 + \varepsilon_{i,t}^{i}, \tag{20}$$

$$\log(m_{x,t}^{China}) = \beta_{1,t}^{China} Y_1^{China} + \beta_{2,t}^{China} Y_2^{China} + \beta_{3,t} Y_3 + \varepsilon_{i,t}^{China}, \tag{21}$$

where $\beta_{2,t}^{k}$ and Y_2^{k} are common factors within the region to capture the regional patterns. We allow the province level $\beta_{1,t}^{i}$ to pool information within region *k*:

$$\beta_{1,t}^{i} \sim N(\mu_{1,t}^{region\ k}, \sigma_{k1,t}^{2}), \text{when } i \text{ in region } k, \tag{22}$$

where $\mu_{1,t}^{region\ k}$ is the population-weighted mean of the first coefficient in the *k*th region (*k* = 1, 2, …, K; K = 4) computed by SVD. Provinces in the same region share the same regional common factors:

$$\beta_{2,t}^{k} \sim N(\mu_{2,t}^{region\ k}, \sigma_{k2,t}^{2}), Y_2^{k} \sim N(Y_2^{China,I}, \sigma_{y2,t}^{2}). \tag{23}$$

where $\mu_{2,t}^{region\ k}$ is the population-weighted mean of the second coefficient in the *k*th region computed by SVD.

The second factors for China are estimated by:

$$\beta_{2,t}^{China} \sim N(\beta_{2,t}^{China,I}, \sigma_{c2,t}^{2}),\ Y_2^{China} \sim N(Y_2^{China,I}, \sigma_{y2,t}^{2}), \tag{24}$$

where the variances have IG priors: $\sigma_{kp,t}^{2} \sim IG(1, 0.001)$, $p = 1,\ 2$; $\sigma_{c2,t}^{2} \sim IG(1, 0.001)$, $\sigma_{y2,t}^{2} \sim IG(1, 0.001)$.

The province-level Y_1^{i}, and common national factors $\beta_{3,t}$ and Y_3 are estimated in the same way as the two-level model.

The future $\hat{\beta}^{i}_{1,t+n}$ are forecasted by drawing from a normal distribution allowing for temporal and regional uncertainty:

$$\hat{\beta}^{i}_{1,t+n} \sim N(\beta^{i}_{1,t+n}, \sigma^{2}_{\beta k_{t+n}}), \tag{25}$$

where the parameters $\beta^{i}_{1,t+n}$ are forecasted by a random walk process:

$$\beta^{i}_{1,t+n} = \beta^{i}_{1,t} + n \cdot d^{i}_{1},\ d^{i}_{1} = (\beta^{i}_{1,4} - \beta^{i}_{1,1}) / T . \tag{26}$$

The regional uncertainty $\sigma^{2}_{\beta k_{t+n}}$ has the form:

$$\sigma^{2}_{\beta k_{t+n}} = Var(\beta^{k}_{1,4}) + \frac{n^{2}(Var(\beta^{k}_{1,4}) + Var(\beta^{k}_{1,1}))}{T^{2}}, \tag{27}$$

where $Var(\beta^{k}_{1,1})$=$Var(\beta^{1:m_k}_{1,1})$ and $Var(\beta^{k}_{1,4})$=$Var(\beta^{1:m_k}_{1,4})$ are posterior variances of $\beta^{i}_{1,1}$ and $\beta^{i}_{1,4}$ in region *k*, and m_k is the number of provinces in region *k*. Similarly, the $\hat{\beta}^{China}_{1,t+n}$ are forecasted by:

$$\hat{\beta}^{China}_{1,t+n} \sim N(\beta^{China}_{1,t+n}, \sigma^{2}_{\beta c^{*}_{t+n}}),\ \beta^{China}_{1,t+n} = \beta^{China}_{1,t} + n \cdot d^{c}_{1},\ d^{c}_{1} = (\beta^{China}_{1,4} - \beta^{China}_{1,1}) / T . \tag{28}$$

where the uncertainty for China is:

$$\sigma^{2}_{\beta c^{*}_{t+n}} = Var(\bar{\beta}^{k}_{1,4}) + \frac{n^{2}(Var(\bar{\beta}^{k}_{1,4}) + Var(\bar{\beta}^{k}_{1,1}))}{T^{2}}, \tag{29}$$

where $Var(\bar{\beta}^{k}_{1,1})$ and $Var(\bar{\beta}^{k}_{1,4})$ are the average regional posterior variances.

There is no regional uncertainty in $\beta^{k}_{2,t}$ as provinces have the same value of $\beta^{k}_{2,t}$ in the same region. The $\hat{\beta}^{k}_{2,t+n}$ are forecasted by:

$$\hat{\beta}^{k}_{2,t+n} = \beta^{k}_{2,t} + n \cdot d^{k}_{2},\ d^{k}_{2} = (\beta^{k}_{2,4} - \beta^{k}_{2,1}) / T . \tag{30}$$

$\hat{\beta}^{China}_{2,t+n}$ is forecasted by:

$$\hat{\beta}^{China}_{2,t+n} = \beta^{China}_{2,t} + n \cdot d^{c}_{2},\ d^{c}_{2} = (\beta^{c}_{2,4} - \beta^{c}_{2,1}) / T . \tag{31}$$

The forecasting process for $\hat{\beta}_{3,t+n}$ is the same as in the two-level model. The $\log(\hat{m}^{i}_{x,t+n})$ and $\log(\hat{m}^{China}_{x,t+n})$ are forecasted by:

$$\log(\hat{m}^{i}_{x,t+n}) = \hat{\beta}^{i}_{1,t+n} Y^{i}_{1} + \hat{\beta}^{k}_{2,t+n} Y^{k}_{2} + \hat{\beta}_{3,t+n} Y_{3}, \tag{32}$$

$$\log(\hat{m}^{China}_{x,t+n}) = \hat{\beta}^{China}_{1,t+n} Y^{China}_{1} + \hat{\beta}^{China}_{2,t+n} Y^{China}_{2} + \hat{\beta}_{3,t+n} Y_{3} . \tag{33}$$

The estimation and forecasts were completed using the software R and JAGS through the R package “rjags”. We generated samples from the posterior distributions via the Markov Chain

Monte Carlo (MCMC) algorithm using Gibbs sampling. For each model, we generated the samples with two chains and thinned the chains by sampling every 10^{th} observation to reduce sample autocorrelation. After a burn-in of 10,000 iterations and convergence tests, posterior distributions were estimated based on the last 10,000 recorded samples.

4. Results

In this section, we evaluate the performance of the baseline two-level model and the three-level model and show the estimation and forecast results of the models.

4.1. Model performance evaluation

We evaluate the performance of the baseline two-level model and the three-level model based on the DIC. The DIC is a Bayesian extension of the Akaike information criterion (AIC) and a popular model selection criterion to deal with informative priors in Bayesian hierarchical models (Yan et al., 2018). According to Spiegelhalter et al. (2002), the DIC is calculated as:

$$DIC = \bar{D} + p_D , \tag{34}$$

where $\bar{D}$ is the expected deviance, which measures the model fit, and p_D is the effective number of parameters measuring the model complexity.

The DIC of the two-level model and the three-level model are shown in Table 3.[4] A lower DIC value indicates a better model fit and a DIC difference of 3–5 is considered significant (Lawson, 2013; Khana et al., 2018). Although the three-level model has more parameters, it has a lower DIC compared to the two-level model, which indicates that adding a regional hierarchy improves the model performance.

Table 3. DIC of the models.

Model	Two-level	Three-level
DIC	29631.5	29615.6

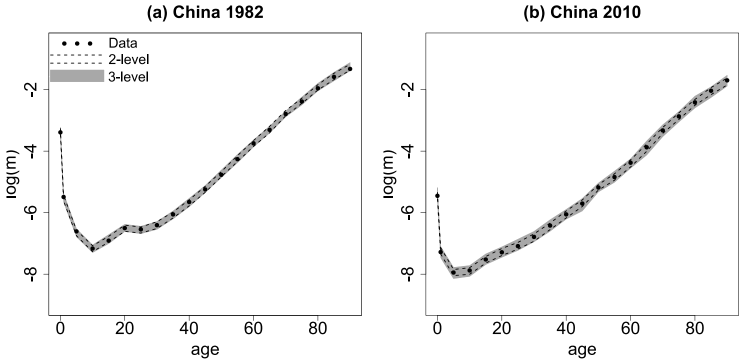
4.2. Estimation results

The fan charts in Figure 3 show the estimated mortality rates from the two-level model and the three-level model for China in the census years 1982 and 2010. As the fitting results are similar for all four census years, we use the first and last census years as representatives. The black dots are the historical data, the intervals between the black dashed lines are the 95% fitting intervals of the two-level model, and the grey intervals are the 95% fitting intervals of the three-level model. Figure 3 shows that both models reproduce the historical mortality rates

[4] We run 30 iterations for the DIC computation and take the average.

of China well. The three-level model has a wider interval than the two-level model in 2010.

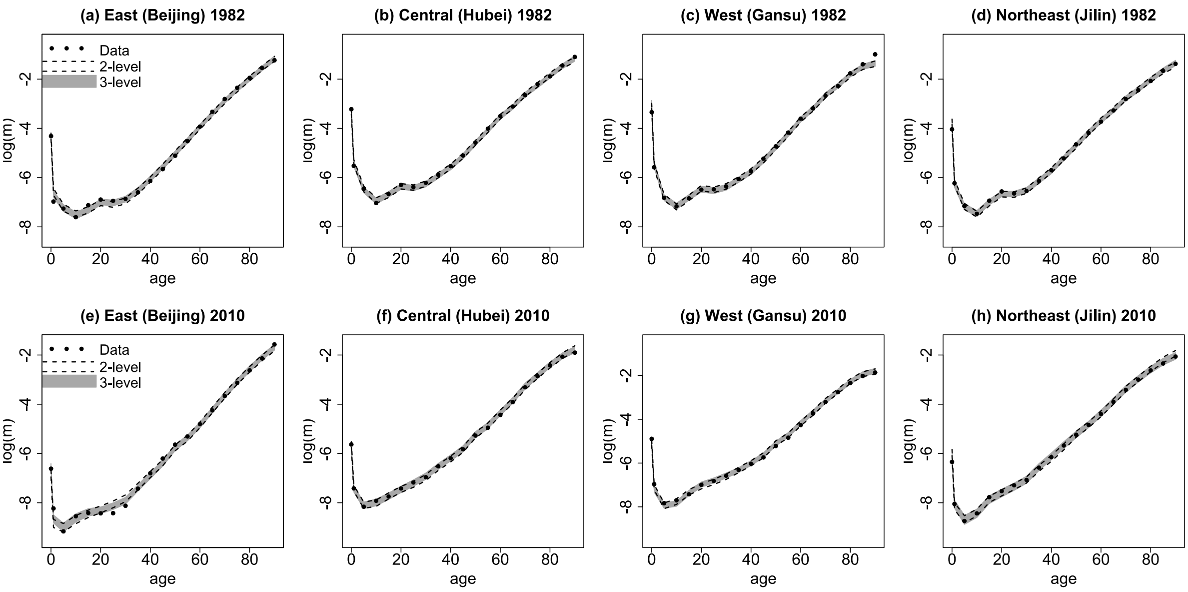
Figure 3. Fitting intervals of two-level and three-level models for China.



Notes: The black points represent historical data, the black dashed lines are the fitting intervals of the two-level model, and the grey shading represents the fitting intervals of the three-level model.

The charts in Figure 4 show the estimated mortality rates for four provinces[5] in different regions. The estimation results for the provinces show that both models perform well at reproducing historical data.

Figure 4. Estimation intervals of two-level and three-level models for four provinces.



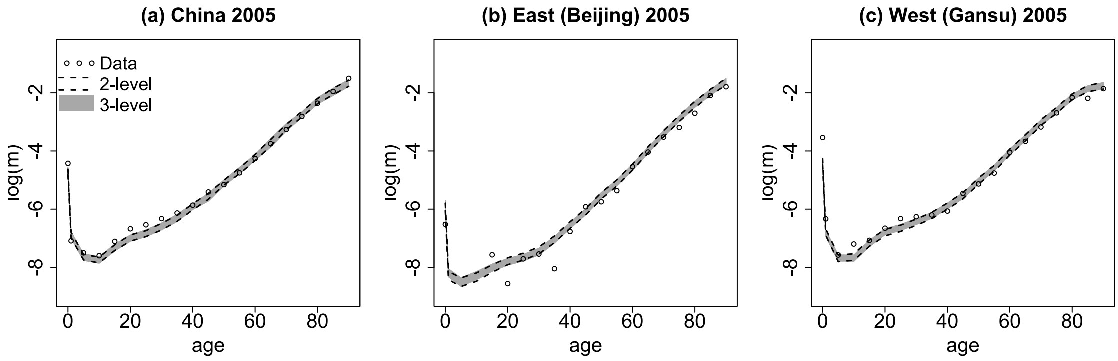
Notes: The black points represent historical data, the black dashed lines are the fitting intervals of the two-level model, and the grey shading represents the fitting intervals of the three-level model.

Apart from the census, China and its provinces also have mortality data from the 1% sample survey data, as mentioned in Section 2. Figure 5 shows the estimation results for China and

[5] Beijing represents the most developed east region; Hubei represents the central region; Gansu represents the west region; and Jilin represents the northeast region. All four provinces have moderate life expectancies (Zhou et al., 2016).

two provinces in 2005, where the sample survey data are shown as hollow dots.[6] As noted before, the data quality of the 1% sample survey at the province level is relatively poor due to small sample sizes. Although the sample survey data are more volatile than the census data, our two models produce reasonable estimates both for China and its provinces.

Figure 5. In-sample forecast of the two-level and three-level models for China in 2005.



Notes: The hollow dots are the sample survey data, the intervals between the black dashed lines are the fitting intervals of the two-level model, and the grey shading represents the fitting intervals of the three-level model.

We note that our models handle missing data well. Although provincial mortality data for Tibet in 1982 are missing, the models can still estimate mortality by using information from other provinces.

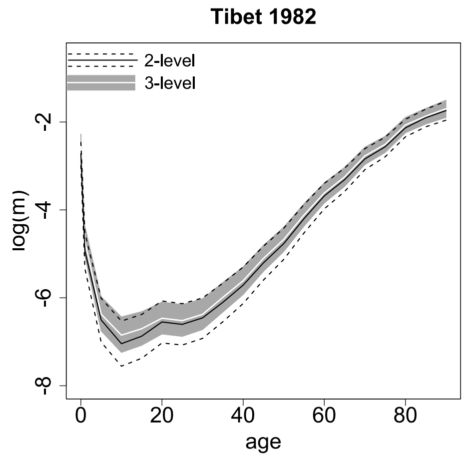
Figure 6 shows the estimation results for Tibet in 1982. The black lines are the estimated mortality rates of the two-level model and the interval between the black dashed lines is the corresponding estimation interval. The white lines and the grey interval are for the three-level model.

Figure 6 shows that the estimated mortality rates of the three-level model are higher than those of the two-level model, which is due to the different information-sharing structures of the two models; the three-level model additionally shares information in the west region, where the mortality rates are higher than they are in other regions, and contributes to a higher estimation for Tibet.

The intervals in Figure 6 are much wider than they are for the other provinces in Figure 4 due to greater uncertainty stemming from missing data. The interval of the two-level model is wider than that of the three-level model, indicating that nationwide uncertainty is greater than regional uncertainty.

[6] The estimation results in 2005 are similar for all provinces; results from the Beijing and Gansu provinces represent low mortality provinces and high mortality provinces, respectively.

Figure 6. Estimated mortality rates for Tibet in 1982.

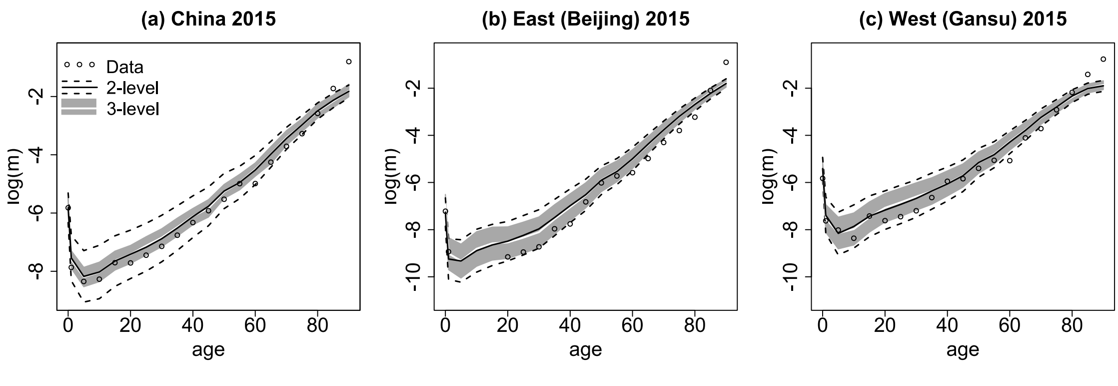


Notes: The black and white lines are estimation medians of the two-level and three-level models, respectively.

4.3. Model forecast

Figure 7 shows the forecast results of the baseline two-level and three-level models in 2015. The two models have similar forecasts but different interval widths. The forecast intervals of the two-level model are wider because nationwide uncertainty is larger than regional uncertainty.

Figure 7. Out-of-sample forecast results of the two-level and three-level models in 2015.



Notes: The results are for China, Beijing (in the east region) and Gansu (in the west region) respectively. The hollow dots are the sample survey data, the intervals between the black dashed lines are the fitting intervals of the two-level model, and the grey shaded areas are the fitting intervals of the three-level model.

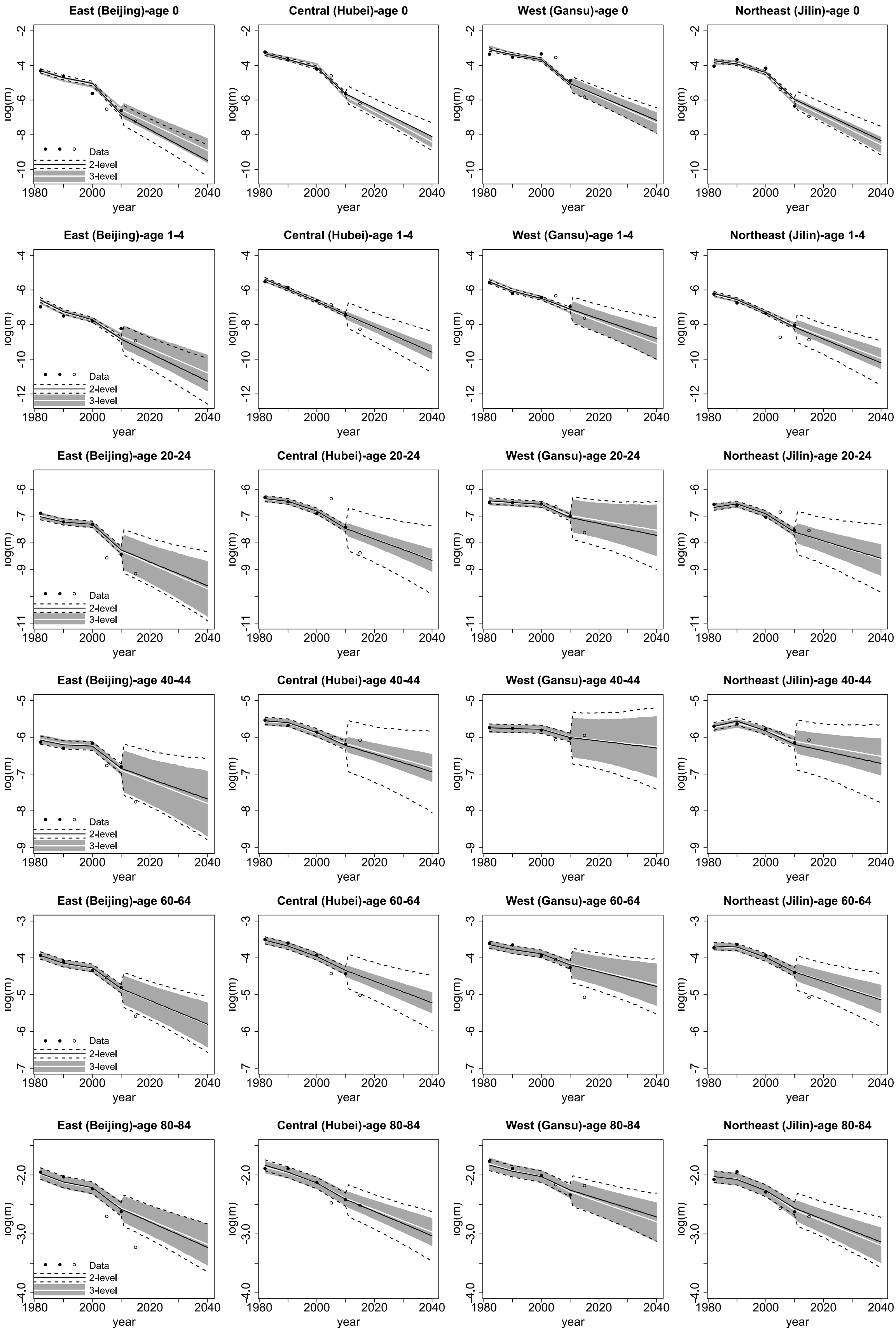
The charts in Figure 8 display the estimation and forecast results for different age groups for the four provinces.[7] The two-level and three-level models have similar estimates but different forecast trends and forecast intervals because of their different information-sharing structures.

[7] The provincial fitting and forecast intervals over time are similar for different age groups. Here we show the 0 and 1–4 age groups as representative of early childhood; the 20–24 and 40–44 age groups as representative of middle age; and the 60–64 and 80–84 age groups as representative of old age.

The two-level model shares information across all provinces and uses the same uncertainty for all provinces in the forecast. As a result, within one age group, the forecast trend and future intervals are the same for different provinces, as shown across the rows of Figure 8. The three-level model shares information both within the same region and across the provinces and uses uncertainty within the region in forecasts. Therefore, the grey intervals in the rows of Figure 8 show that the provincial forecasts of the three-level model have different levels of uncertainty for the same age groups. The forecasts of the three-level model in Figure 8 also show that the uncertainty of younger age groups (under 4 years old) is smaller than that in the older age groups within regions, indicating that infant and childhood mortality converge to a similar level within regions. When comparing different regions, the east region (the first column) and the west region (the third column) have larger uncertainty, indicating that the provincial variations are bigger in these two regions, while the forecasts of the central region (the second column) and the northeast region (the last column) have smaller uncertainty, indicating smaller provincial variations in these regions.[8] From the point of view of forecast uncertainty, the three-level model captures different levels of uncertainty for different regions. Although sample survey data are quite volatile, the forecast intervals of the three-level model cover 57% of the sample survey data for all provinces in 2015.

[8] In further analysis, we find that provincial or regional uncertainty is much bigger than temporal uncertainty in China. When provincial or regional uncertainty is ignored, the forecast intervals are much narrower. The results of this analysis are available upon request from the authors.

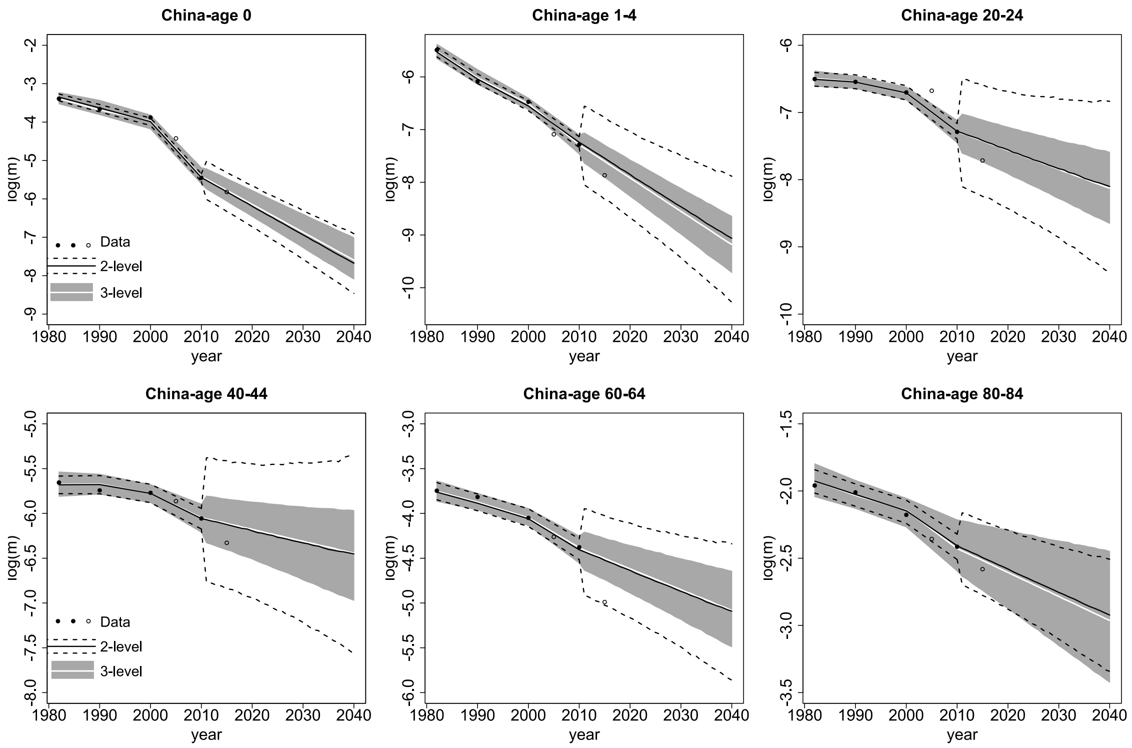
Figure 8. Fitting and forecast results of the two-level and three-level models.



Notes: The black dots are census data from 1982, 1990, 2000 and 2010, and the hollow dots are sample survey data from 2005 and 2015. The black and white lines show the forecast medians of the two-level and three-level models respectively. The intervals between the black dashed lines are the forecast intervals of the two-level model, and the grey shaded areas represent the forecast intervals of the three-level model.

The estimates for 1982–2010 and the forecast results for 2011–2040 for China are shown in Figure 9. Both models produce good fits of the historical trends as well as reasonable forecasts. While the two-level model uses the uncertainty of all provinces to make forecasts, and the three-level model uses uncertainty of the four regions, the two models have different width of forecast intervals. Below age 80, the forecast intervals of the two-level model are wider than those of the three-level model. The forecast intervals are more similar at higher ages (above age 80) because regional uncertainty is larger in older age groups.

Figure 9. Estimates and forecasts for China.



Notes: The black dots represent census data from 1982, 1990, 2000 and 2010, and the hollow dots are sample survey data from 2005 and 2015. The black and white lines show the forecast medians of the two-level and three-level models respectively. The intervals between the black dashed lines are the forecast intervals of the two-level model, and the grey shaded areas represent the forecast intervals of the three-level model.

4.4. Sensitivity analysis

As we only have four sets of census data for every province, we assume that the mortality rates change linearly and use the simple random walk process in our models to make forecasts. But it is difficult to determine whether the random walk process is suitable for our models considering the limited number of data points.

Furthermore, in our main analysis, we use the official classification developed by the NBS to group the provinces into different regions based on their geographic and economic characteristics. However, this regional grouping assumption may be questioned and different grouping assumptions may lead to different forecasts.

In this section, we conduct sensitivity analyses on different trend assumptions and alternative regional grouping assumptions.

4.4.1. Trend assumptions

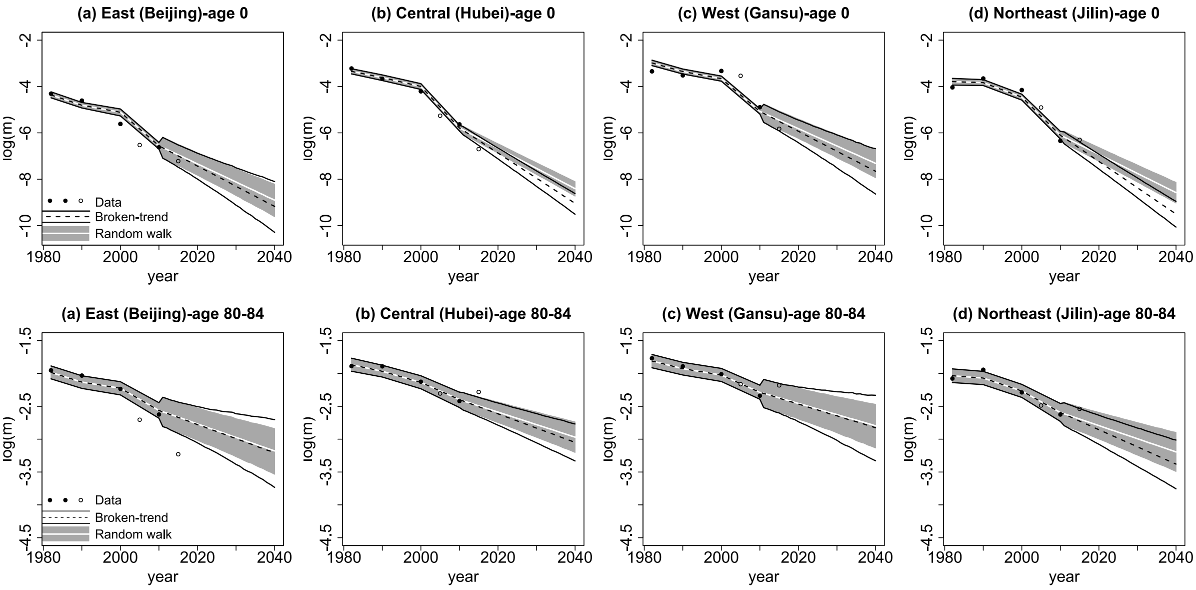
The modelling of the random walk process is based on the linear trend between the first and last historical data points. However, there is evidence of nonlinearity of the mortality trend in the literature. For example, Li et al. (2011) and Sweeting (2011) find structural changes in the time index of both the Lee-Carter model and the Cairns-Blake-Dowd model and use broken-trend models to forecast mortality. Based on this evidence, Van Berkum et al. (2014) allow for multiple structural changes in 14 different mortality models and obtain more robust projections.

In this trend assumption analysis, we use as an alternative the broken-trend model (Li et al., 2011) to derive forecasts and compare the results for the three-level model with the random walk process. We focus on the three-level model because the above analyses show that it performs better than the two-level model based on our data set.

The broken-trend model examines the breakpoints and makes use of all the data points after the breakpoints to make forecasts, which incorporates more information on the historical data. Li et al. (2011) model the time index using a linear regression with five parameters and employ the standard student t statistic to test whether time t is a breakpoint. However, we only have four available data points, which are not enough to estimate five parameters. In order to test the trend assumptions and incorporate more historical information, we assume that there was a structural change around 1990 and set 1990 as the breakpoint.

The results of the broken-trend model are similar for different age groups, therefore we only show the 0 and 80–84 age groups as representatives in Figure 10. The dashed black lines are the fitted and forecasted trends of the broken-trend model, and the intervals between the black lines are the 95% fit and forecast intervals of the broken-trend model. Instead of only using data from 1982 and 2010, the broken-trend model uses three data points (1990, 2000 and 2010), and therefore has wider intervals reflecting larger uncertainty. Because the broken-trend model leaves out the data from 1982 and incorporates more recent information, it produces steeper trends compared to the random walk process, as shown in Figure 10, indicating an acceleration in mortality decline in recent years. As we have limited provincial data and the sample survey data are of relatively low quality, actuarial judgement is needed to choose the appropriate trend assumption. We suggest that researchers and practitioners choose carefully when modelling mortality rates in China.

Figure 10. Forecasts of the broken-trend model and the random walk process in the three-level model.



Notes: The black dots show the census data, and the hollow dots show the sample survey data. The grey intervals and white lines are the 95% fitting and forecast intervals and medians of random walk in the three-level model respectively. The black line intervals and the dashed black lines are the 95% fitting and forecast intervals and medians of the three-level model with broken trend respectively.

4.4.2. Alternative regional grouping assumptions

In the main analysis, Hebei province is grouped within the most developed region (i.e., the east region) in our three-level model. But Hebei's economic strength (e.g., GDP per capita; NBS, 2019) and life expectancy (Zhou et al., 2016) are similar to other less developed regions. Because Hebei lies adjacent to the central, west and northeast regions, it can be grouped with any of these regions. In the following, we use different grouping assumptions to derive forecasts in the three-level model with a random walk process. We compare the results based on the original grouping where Hebei is part of the east region, with three alternative grouping assumptions: (1) Hebei grouped into the central region; (2) Hebei grouped into the west region; and (3) Hebei grouped into the northeast region.

Figure 11 shows the forecast results based on the original grouping and the other three grouping assumptions. The forecast results are quite similar for different age groups, so we show the results at age 70–74 for Guangdong as representative of the east region. We also show the results from Heilongjiang (which is in the northeast region) in Figure 11 because the northeast region is the smallest region and it is interesting to observe the changes when Hebei is included.

The first column in Figure 11 shows that the forecast trends are slightly steeper and the forecast intervals are narrower when Hebei is excluded from the east region. The second column in Figure 11 shows that when Hebei is included in the northeast region, the forecast trend of Heilongjiang is slightly less steep and the forecast interval is slightly wider than in the original grouping. Overall, the differences are limited when changing the grouping assumption. The third column in Figure 11 shows that the fits and forecasts of Hebei change according to which region it is included in. A comparison of Figure 11 (c) with (f) and (i) shows that when Hebei is included in the central region, the slope of the forecast trend is less steep compared to the original grouping, while the forecast interval is much narrower compared to the original grouping assumption and other new grouping assumptions, indicating that the Hebei mortality rates are more similar to those in the central region.

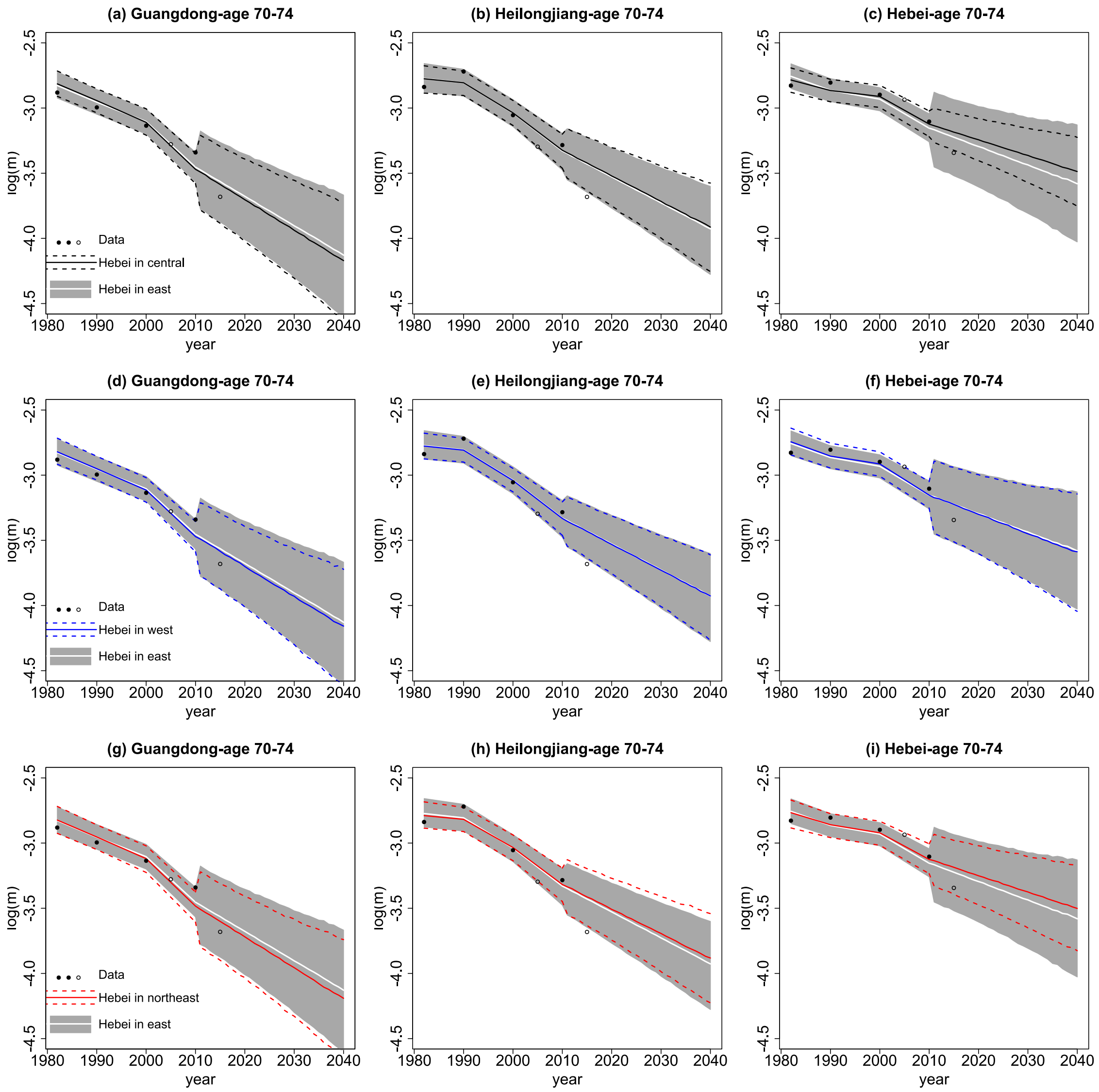
We define the coverage rate as the percentage of the census and sample survey data from China and all provinces that are covered by the fit and forecast intervals. Table 4 reports the coverage rates of the different grouping assumptions. Compared to the new grouping assumptions, the original grouping covers the most historical data, indicating that the official classification is quite reasonable.

In summary, we conducted sensitivity analyses on different trend assumptions and compared the results based on a broken-trend model and the random walk process. The broken-trend model gives steeper trends and wider forecast intervals than the random walk process. We also conducted sensitivity analyses on alternative grouping assumptions. We find that the grouping assumptions have only a limited impact on the forecast trends and intervals. The coverage rates of historical data by different grouping methods show that the official classification by the NBS (which we used in the main analysis) is reasonable.

Table 4. Coverage rates of the different grouping assumptions.

Grouping assumptions	Original	Hebei in central	Hebei in west	Hebei in northeast
Coverage (%)	71.3	58.7	53.2	61.4

Figure 11. Forecasts of different grouping assumptions in the three-level model.



Notes: The black dots show the census data, and the hollow dots show the sample survey data. The grey intervals and white lines are the 95% fitting and forecast intervals and medians of the original three-level model with Hebei in the east region respectively. The black dashed intervals and the black lines are the 95% fitting and forecast intervals and the medians of grouping assumption (1) respectively. The blue dashed intervals and the blue lines are the 95% fitting and forecast intervals and the medians of grouping assumption (2) respectively. The red dashed intervals and the red lines are the 95% fitting and forecast intervals and the medians of grouping assumption (3) respectively.

5. Conclusion

Over the past 60 years, life expectancy in China has increased by more than 30 years at the national level, but provincial variations in mortality and life expectancy are large. Although there is a growing interest in China's mortality rate in the literature, few previous studies focus on modelling provincial mortality. This paper provides new models to estimate and forecast mortality in China at both the provincial and national levels.

We propose two hierarchical models: a baseline two-level model and an extended three-level model in a Bayesian framework based on principal components and a random walk process. The two-level model uses China and its provinces as its two hierarchical levels and uses province-level factors and common national factors, which allows for information pooling and sharing across provinces and consistency conditions between China and its provinces. The extended three-level model introduces a regional level in the hierarchy between China and its provinces. Apart from pooling and sharing information across provinces, the three-level model also pools and shares information in four major regions using common factors within the region. The models are estimated based on a new database that contains census data for the 31 provinces and China over the period 1982–2010. We use sample survey data in 2005 and 2015 as a reference for in-sample and out-of-sample forecasts.

The baseline two-level model provides a good fit for the census years and sample survey years and provides reasonable forecasts for China and its provinces. The forecast intervals are of equal width for all provinces within one age group as a result of the model's information-sharing structure. Compared with the baseline model, the extended three-level model has a better fit with a lower DIC and provides coherent forecasts for the provinces with different widths of intervals reflecting regional uncertainty. Overall, both models provide good estimates and reasonable forecasts for China and its provinces, and we recommend using the models in both national mortality modelling and subnational mortality modelling in China.

We conducted sensitivity analyses on different trend assumptions and alternative grouping assumptions. The results show that the forecasts are relatively robust when changing the trend and grouping assumptions. The original grouping based on the NBS' official classification is reasonable. As the census data are limited and the sample survey data are quite volatile at the provincial level, actuarial judgement is still needed to choose the appropriate trend assumption. We recommend that researchers choose the trend assumption carefully when modelling mortality rates in China.

We note again that the available provincial mortality data in China are very limited, with missing data for some provinces. However, both of our models can deal with missing data and can provide reasonable estimates for provinces where data are missing.

The models we have developed are based on principal components considering the parsimony and flexibility of this approach. The Bayesian framework can incorporate other functional forms, for example, the Cairns-Blake-Dowd model and models that include cohort effects. Future work can also consider other estimation methods like the Kalman filter to improve the models. It would also be interesting to study regional annuity pricing and pension liabilities based on the models we have developed here.

Acknowledgements

The authors acknowledge the financial support from the Australian Research Council Centre of Excellence in Population Ageing Research (CEPAR) [grant number CE17010005] and the China Scholarship Council. This research includes computations using the computational cluster Katana supported by Research Technology Services at UNSW Sydney. We are grateful for comments received from Andrés Villegas and Cheng Wan.

References

Alexander, M., Zagheni, E., Barbieri, M., 2017. A flexible Bayesian model for estimating subnational mortality. *Demography,* 54 (2), 1–17.

Banister, J., Hill, K., 2004. Mortality in China 1964–2000. *Population Studies,* 58, 55–75.

Bergeron-Boucher, M.P., Canudas-Romo, V., Oeppen, J., Vaupel, J.W., 2017. Coherent forecasts of mortality with compositional data analysis. *Demographic Research,* 37, 527–566.

Cai, Y., 2005. National, provincial, prefectural and county life tables for China based on the 2000 census. Center for Studies in Demography and Ecology (CSDE) Working Paper, 05-03.

Coale, A. J., Demeny, P., Vaughan, B., 1983. *Regional Model Life Tables and Stable Populations: Studies in Population*. 2nd edition. New York, NY: Academic Press.

Czado, C., Delwarde, A., Denuit, M., 2004. Bayesian Poisson log-bilinear mortality projections. *Insurance: Mathematics & Economics,* 36 (3), 260–284.

Danesi, I.L., Haberman, S., Millossovich, P., 2015. Forecasting mortality in subpopulations using Lee-Carter type models: A comparison. *Insurance: Mathematics and Economics,* 62, 151–161.

De Beer, J., 2012. Smoothing and projecting age-specific probabilities of death by TOPALS. *Demographic Research,* 27 (6), 543–592.

Development Research Center of the State Council (DRC), 2005. Outline of the Eleventh Five-Year National Economic and Social Development Planning of the People's Republic of China. http://ghs.ndrc.gov.cn/zttp/ghjd/quanwen/.

Dowd, K., Cairns, A., Blake, D., Coughlan, G., Khalaf-Allah, M., 2011. A gravity model of mortality rates for two related populations. *North American Actuarial Journal,* 15 (2), 334–356.

Fang, H., Feng, J., 2018. The Chinese pension system. NBER Working Paper No. w25088. National Bureau of Economic Research.

Giannakouris K., 2010. Regional Population Projections EUROPOP2008: Most EU Regions Face Older Population Profile in 2030. Eurostat Statistics in Focus 1/2010. Luxembourg: European Commission.

Gonzaga, M. R., Schmertmann, C. P., 2016. Estimating age-and sex-specific mortality rates for small areas with TOPALS regression: an application to Brazil in 2010. *Revista Brasileira de Estudos de População*, 33(3), 629-652.

Huang, F., Browne, B., 2017. Mortality forecasting using a modified Continuous Mortality Investigation Mortality Projections Model for China I: Methodology and country-level results. *Annals of Actuarial Science*, 11 (1), 20–45.

Huang, F. (2017). Mortality forecasting using a modified CMI Mortality Projections Model for China II: Cities, towns and counties. *Annals of Actuarial Science,* 11 (1), 46–66.

Hyndman, R.J., Booth, H., Yasmeen, F., 2013. Coherent mortality forecasting: The product-ratio method with functional time series models. *Demography,* 50 (1), 261–283.

Khana, D., Rossen, L.M., Hedegaard, H., Warner, M., 2018. A Bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with R-INLA. *Journal of Data Science,* 16 (1), 147.

Kleinow, T., 2015. A common age effect model for the mortality of multiple populations. *Insurance: Mathematics and Economics* 63, 147–152.

Kogure, A., Kitsukawa, K., Kurachi, Y., 2009. A Bayesian comparison of models for changing mortalities toward evaluating longevity risk in Japan. *Asia-Pacific Journal of Risk and Insurance*, 3 (2).

Lawson, A.B., 2013. *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. London: Chapman and Hall/CRC.

Li, J., 2014. An application of MCMC simulation in mortality projection for populations with limited data. *Demographic Research*, 30 (1), 1–48.

Li, J.S., Zhou, K.Q., Zhu., X., et al., 2019. A Bayesian approach to developing a stochastic mortality model for China. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, forthcoming.

Li, N., Lee, R., 2005. Coherent mortality forecasts for a group of populations: An extension of the Lee-Carter method. *Demography,* 42 (3), 575–594.

Li, S.H., Chan, W.S., Cheung, S.H., 2011. Structural changes in the Lee-Carter mortality indexes. *North American Actuarial Journal,* 15 (1), 13–31.

National Bureau of Statistics of China (NBS), 2002. China's 2000 National Census Data. http://www.stats.gov.cn/tjsj/pcsj/rkpc/5rp/index.htm.

National Bureau of Statistics of China (NBS), 2006. National 1% Population Sample Survey Data for 2005. http://www.stats.gov.cn/tjsj/ndsj/renkou/2005/renkou.htm.

National Bureau of Statistics of China (NBS), 2011. Division Methods of Eastern, Western, Central and Northeastern Regions (in Chinese). http://www.stats.gov.cn/ztjc/zthd/sjtjr/dejtjkfr/tjkp/201106/t20110613_71947.htm.

National Bureau of Statistics of China (NBS), 2012. China's 2010 National Census Data. http://www.stats.gov.cn/tjsj/pcsj/rkpc/6rp/indexch.htm.

National Bureau of Statistics of China (NBS), 2016. National 1% Population Sample Survey Data for 2015. Beijing: China Statistics Press.

National Bureau of Statistics of China (NBS), 2019. National Data, Annual by Provinces. http://data.stats.gov.cn/english/easyquery.htm?cn=E0103.

Office for National Statistics (ONS), 2016. Methodology Used to Produce the 2014-Based

Subnational Population Projections for England. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/methodologies/methodologyusedtoproducethesubnationalpopulationprojectionsforengland.

Preston, S., Heuveline, P., Guillot, M., 2000. *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell, Malden, MA.

Schmertmann, C.P., Gonzaga, M.R., 2018. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography,* 55 (4), 1363–1388.

Smith, S.K., Tayman, J., Swanson, D.A., 2013. *A Practitioner's Guide to State and Local Population Projections*. Springer, Berlin.

Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2014. The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* 76 (3), 485–493.

Sweeting, P.J., 2011. A trend-change extension of the Cairns-Blake-Dowd model. *Annals of Actuarial Science*, 5 (2), 143–162.

United Nations, 2019. World Population Prospects: The 2019 Revision. United Nations Population Division. https://population.un.org/wpp/.

Van Berkum, F., Antonio, K., Vellekoop, M., 2014. The impact of multiple structural changes on mortality predictions. *Social Science Electronic Publishing,* 2016 (7), 1–22.

Wang, X., Li, Q., 2018. Trends in the Chinese Annuity Market and the Management of Longevity Risk. Report published by the Society of Actuaries, December.

Yan, H., Peters, G., Chan, J., 2018. Mortality models incorporating long memory improves life table estimation: a comprehensive analysis. Working paper available at SSRN: https://ssrn.com/abstract=3149914.

Zhao, B.B., 2012. A modified Lee-Carter model for analysing short-base-period data. *Population Studies,* 66, 39–52.

Zhao, B.B., Liang, X., Zhao, W., Hou, D., 2013. Modeling of group-specific mortality in China using a modified Lee-Carter model. *Scandinavian Actuarial Journal,* 5, 383–402.

Zhou., M., Li, Y., Wang H., et al., 2016. Analysis on life expectancy and healthy life expectancy in China, 1990-2015. *Chinese Journal of Epidemiology*, 2016, 37(11): 1439-1443. (in Chinese).