A Bayesian Approach in Projecting Future Mortality at the Provincial Level in China

Xiaobai Zhu
Joint work with Johnny Li, Kenneth Zhou, and Wai-Sum Chan

The Chinese University of Hong Kong
xiaobaizhu58@cuhk.edu.hk

Overview

1. Background
2. CAE Model
3. Model Selection and Estimation
4. Illustrations
5. Conclusion and Discussions

Background

Mortality Modeling for Chinese Data

- **National Population**
  Li (2014); Huang and Browne (2017); Wang and Huang (2011); Li et al. (2019), Wu and Wang (2014), etc.

Lee Carter Estimation - National Population

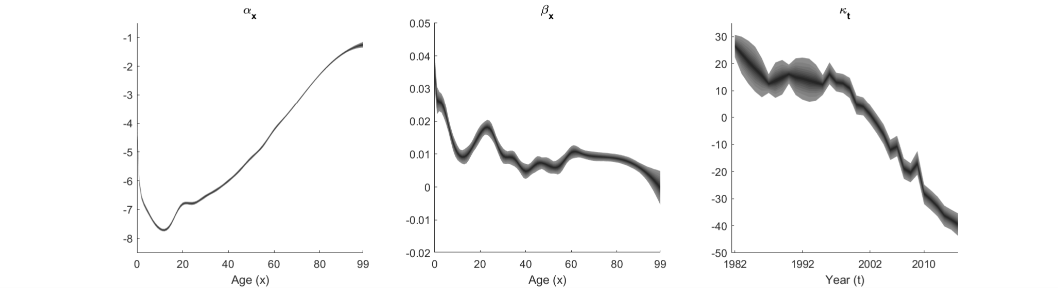


Figure 1: Lee Carter model $\ln m_{x,t} = \alpha_x + \beta_x \cdot \kappa_t$ estimation, α_x (left), β_x (middle) and κ_t (right)

Mortality Modeling for Chinese Data

- **National Population**
  Li (2014); Huang and Browne (2017); Wang and Huang (2011); Li et al. (2019), Wu and Wang (2014), etc.
- **Multi-Population Modeling**
  Lu et al. (2020, 2021); Bai et al. (2023); Huang and Browne (2017); Zhao and Wang (2020); Wang et al. (2021); Zhao (2023), etc.

Data Limitations

Region	1981	1986	1989	1995	2000	2005	2010	2015
GuangDong								
JiangSu								
Hebei								
Hainan								
Yunan								
Jilin								
ShanDong								

Table 1: Data availability for selected provinces, represents the census data, represents the 1% sampling data, represents no data.

Modeling Criterion

For **national mortality modeling**:

1. Exploit as much information as possible from the data.
2. Provide appropriate provisions for uncertainty.
3. Parsimonious and yield biologically reasonable projections.

in addition, for **provincial mortality modeling**:

1. Relative modeling where the relation between national and provincial mortality can be reflected.
2. The provincial data should not affect the modeling of the national population.

Our Approach

- Common Age Effect (CAE) model with age smoothing, where the mortality of the national population is modeled by the Lee-Carter model.
- Over-dispersed Poisson model to take into account sampling uncertainty and Bayesian Estimation to take into account parameter uncertainty.
- Two-step procedure that incorporates national data into the provincial modeling, and coherence testing for each province.

CAE Model

CAE Model

National Population

$$D_{x,t} \sim \text{Poisson}(E_{x,t} \cdot m_{x,t})$$

$$\ln m_{x,t} = \alpha_x + \beta_x \cdot \kappa_t + \epsilon_{x,t}$$

$$\kappa_t = \kappa_{t-1} + \mu + \omega_t$$

CAE Model

National Population

$$D_{x,t} \sim \text{Poisson}(E_{x,t} \cdot m_{x,t})$$
$$\ln m_{x,t} = \alpha_x + \beta_x \cdot \kappa_t + \epsilon_{x,t}$$
$$\kappa_t = \kappa_{t-1} + \mu + \omega_t$$

Provincial Population

$$D_{x,t}^{(i)} \sim \text{Poisson}(E_{x,t}^{(i)} \cdot m_{x,t}^{(i)})$$
$$\ln m_{x,t} - \ln m_{x,t}^{(i)} = \alpha_x^{(i)} + \beta_x \cdot \kappa_t^{(i)} + \epsilon_{x,t}^{(i)}$$

See Cairns et al. (2011) for a more detailed discussion.

Overdispersion

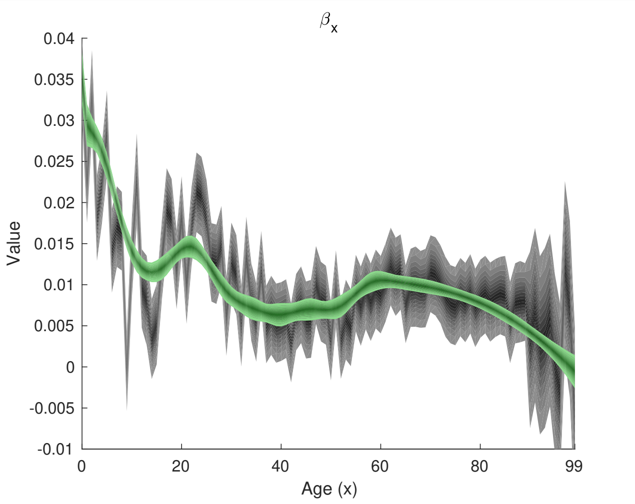
The variance of $\epsilon_{x,t}$ and $\epsilon_{x,t}^{(i)}$ can be interpreted as overdispersion parameters. Based on the analysis of the national population, and preliminary analysis of the provincial population, we choose

- Population-specific, and age-specific overdispersion parameter
- Time-independent overdispersion parameter

such that

$$\text{Var}(\epsilon_{x,t}) = s_x^2, \quad \text{and} \quad \text{Var}(\epsilon_{x,t}^{(i)}) = \left(s_x^{(i)}\right)^2$$

Smoothing of Age Parameters



Smoothing of Age Parameters

Cubic splines are applied to β_x, α_x and $\alpha_x^{(i)}$ to avoid jagged pattern across ages, e.g.

$$\beta_x = c_0 + c_1 \cdot \ln(x+1) + c_2 \cdot \ln(x+1)^2 + c_3 \cdot \ln(x+1)^3$$
$$+ \sum_{j=1}^{r} c_{3+j} \cdot (\ln(x+1) - \ln(x_j+1))_+^3$$
$$\boldsymbol{\beta} = \boldsymbol{A}_\beta \times \boldsymbol{c}$$

The knots are selected based on Deviance Information Criteria (DIC).

Underlying Trend $\kappa_t^{(i)}$

1. An AR(1) process with intercept:
$$\kappa_t^{(i)} = \mu_\Delta^{(i)} + \phi^{(i)}\kappa_{t-1}^{(i)} + \omega_t^{(i)}$$
2. A random walk with drift:
$$\kappa_t^{(i)} = \mu_\Delta^{(i)} + \kappa_{t-1}^{(i)} + \omega_t^{(i)}$$
3. A white noise with non-zero mean:
$$\kappa_t^{(i)} = \mu_\Delta^{(i)} + \omega_t^{(i)}$$
4. An AR(1) process without intercept (i.e., mean-reverting to zero):
$$\kappa_t^{(i)} = \phi^{(i)}\kappa_{t-1}^{(i)} + \omega_t^{(i)}$$
5. A random walk without drift:
$$\kappa_t^{(i)} = \kappa_{t-1}^{(i)} + \omega_t^{(i)}$$
6. A white noise with zero mean:
$$\kappa_t^{(i)} = \omega_t^{(i)}$$

Identifiability Constraints

To stipulate parameter uniqueness, β_x is constrained by

$$\sum_x \beta_x = 1,$$

κ for the national population is constrained initially by

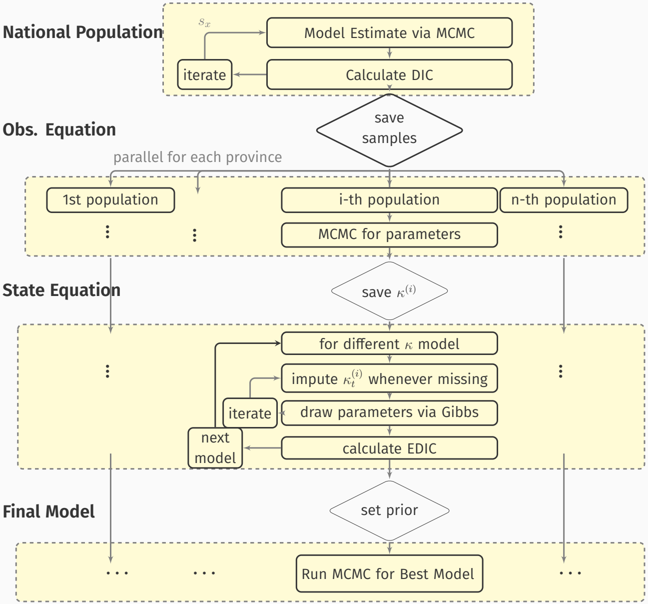
$$\kappa_{t_0} = 0$$

and for each province, we adopt the constraint

$$\alpha_{\hat{x}}^{(i)} = \ln m_{\hat{x},\hat{t}}^{(i)}$$

where $\hat{x} = x_0$ and $\hat{t} = 2010$ are arbitrarily chosen.

Model Selection and Estimation


National Population
s_x
Model Estimate via MCMC
iterate
Calculate DIC
save samples
Obs. Equation
parallel for each province
1st population
i-th population
n-th population
MCMC for parameters
State Equation
save $\kappa^{(i)}$
for different κ model
impute $\kappa_t^{(i)}$ whenever missing
iterate
draw parameters via Gibbs
next model
calculate EDIC
Final Model
set prior
Run MCMC for Best Model

Two-Steps Modelling

Fitting the national model with national data only, and let $\ln M_{x,t} = \ln m_{x,t} | \boldsymbol{D}, \boldsymbol{E}$ and $B_x = \beta_x | \boldsymbol{D}, \boldsymbol{E}$. The model for each province is thus:

$$\ln M_{x,t} - \ln m_{x,t}^{(i)} = \alpha_x^{(i)} + B_x \cdot \kappa_t^{(i)} + \epsilon_{x,t}^{(i)}$$

We first run the MCMC algorithm for the national population (Hasting-within-Gibbs), then save the samples from $\{\ln \boldsymbol{M}, \boldsymbol{B}\}$ to construct the sub-population model.

Two-Steps Modelling, National Component

- $\ln m_{x,t}$, Metropolis-Hasting using a Gaussian proposal distribution that maximizes the full conditional of $\ln m_{x,t}$ and a tuning parameter to have an acceptance rate between $[30\%, 50\%]$.
- κ_t are drawn using sequential Kalman filter discussed in Koopman and Durbin (2000).
- s_x^2 and σ^2 follow Inverse Gamma distributions.
- α_x and β_x with coefficients follow Gaussian distributions with constraint embedded into the prior.
- μ follows Gaussian distribution.

Two-Steps Modelling, Provincial Component

We provide an uninformative prior for $\kappa_t^{(i)}$ without assuming any specific structural form for it, and setting uninformative priors to other parameters.

$$\pi(\kappa_t^{(i)}) \propto 1, \quad \pi(\alpha_x^{(i)}) \propto 1,$$
$$\left(s_x^{(i)}\right)^2 \sim \text{InvGamma}(a_{x,\text{prior}}, b_{x,\text{prior}}), \forall x, t$$

where $a_{x,\text{prior}}, b_{x,\text{prior}}$ are selected based on the national population.

Model Selection via EDIC

The original definition of DIC is

$$\text{DIC} = \mathbb{E}_{\theta|y}[D(\theta)] + \underbrace{\mathbb{E}_{\theta|y}[D(\theta)] - D(\hat{\theta})}_{=p_D}$$

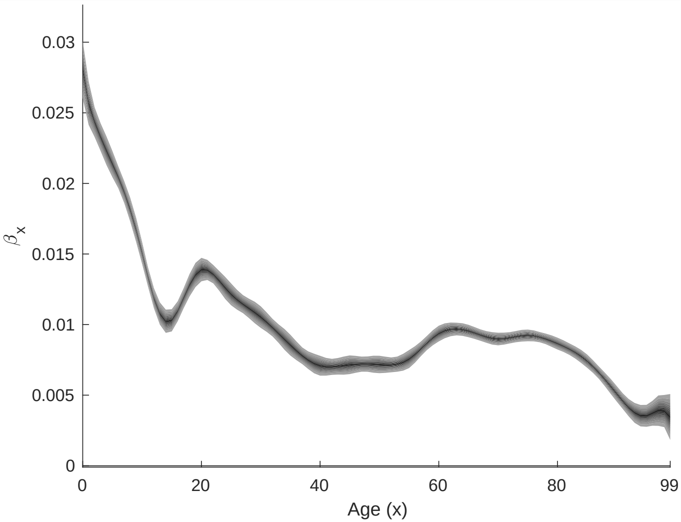
However, for provincial modeling, we rely on the data generating process (DGP, $g(y)$) instead, Li et al. (2017, 2020) provide the expected DIC selection

$$\text{EDIC} \approx \frac{1}{M}\sum_{i=1}^{M}\left[\left(\underbrace{\frac{1}{N}\sum_{j=1}^{N} D(\boldsymbol{\theta}^{(i,j)}; \boldsymbol{y}^{(i)})}_{=\overline{D(\boldsymbol{\theta}^{(i)})}}\right) + \left(\underbrace{\frac{1}{N}\sum_{j=1}^{N} D(\boldsymbol{\theta}^{(i,j)}; \boldsymbol{y}^{(i)}) - D(\hat{\boldsymbol{\theta}}^{(i)})}_{=\overline{P_D(\boldsymbol{y}^{(i)})}}\right)\right].$$

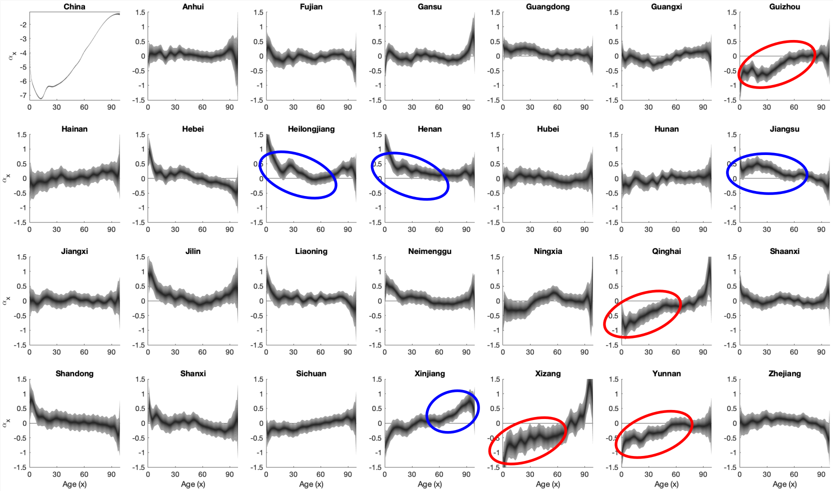
To reduce the computational time, we only draw one sample $\theta^{(i)} = \theta^{(i,1)}$ for each generated sample $\boldsymbol{y}^{(i)}$, such that $N = 1$.

Illustrations

β_x



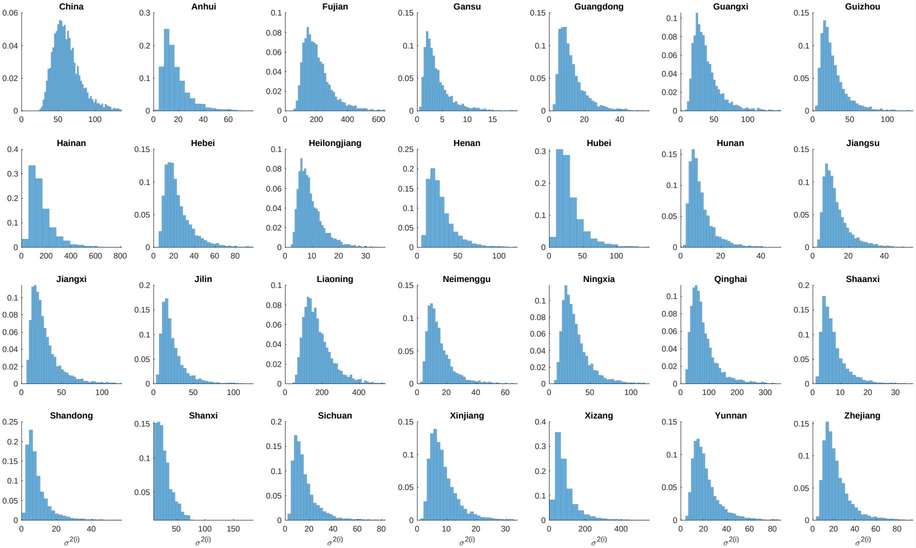
$\alpha_x^{(i)}$


China
Anhui
Fujian
Gansu
Guangdong
Guangxi
Guizhou
Hainan
Hebei
Heilongjiang
Henan
Hubei
Hunan
Jiangsu
Jiangxi
Jilin
Liaoning
Neimenggu
Ningxia
Qinghai
Shaanxi
Shandong
Shanxi
Sichuan
Xinjiang
Xizang
Yunnan
Zhejiang
α_x
Age (x)

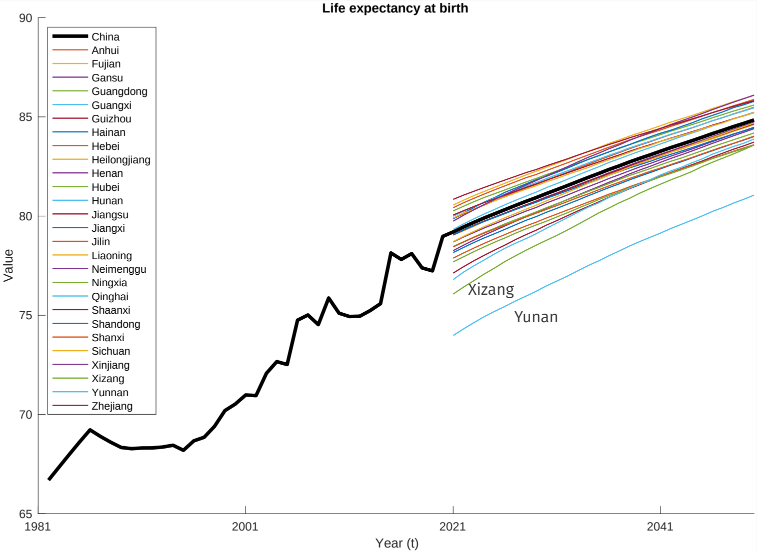
$\kappa_t^{(i)}$

Province	$\kappa_t^{(i)}$	Province	$\kappa_t^{(i)}$
Xi Zang	AR(1) without intercept	He Bei	AR(1) without intercept
Xin Jiang	RW without drift	He Nan	AR(1) with intercept
Yun Nan	RW without drift	Zhe Jiang	AR(1) with intercept
Nei Meng	AR(1) without intercept	Hai Nan	AR(1) without intercept
Ji Lin	RW without drift	Hu Bei	AR(1) without intercept
Si Chuan	AR(1) with intercept	Hu Nan	RW without drift
Ning Xia	RW without drift	Gan Su	AR(1) without intercept
An Hui	AR(1) with intercept	Fu Jian	AR(1) without intercept
Shan Dong	AR(1) without intercept	Gui Zhou	AR(1) without intercept
Shan Xi	AR(1) with intercept	Liao Ning	AR(1) without intercept
Guang Dong	AR(1) with intercept	Shaan Xi	AR(1) without intercept
Guang Xi	AR(1) without intercept	Qing Hai	AR(1) without intercept
Jiang Su	AR(1) without intercept	Hei Long Jiang	RW without drift
Jiang Xi	AR(1) with intercept		

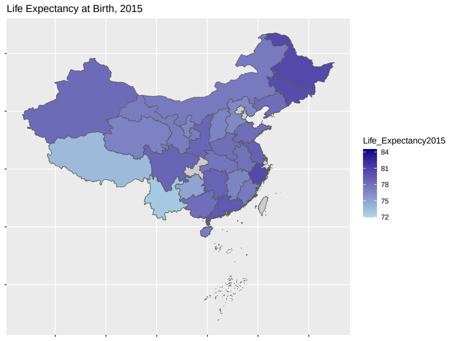
$\left(\sigma^{(i)}\right)^2$



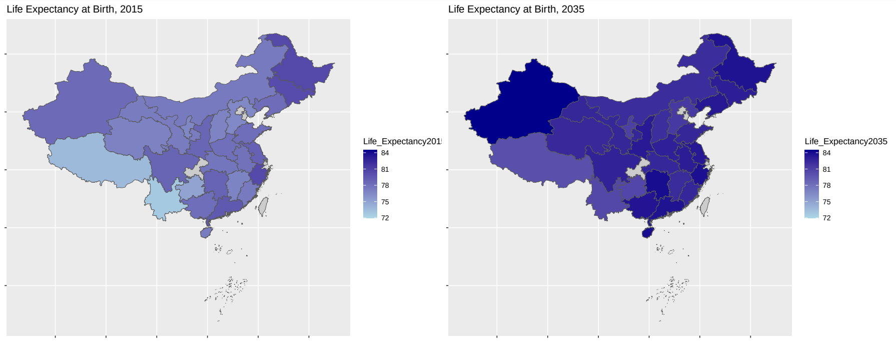
Life Expectancy at Birth



Life Expectancy at Birth, 2015



Life Expectancy at Birth, 2015/2035



Conclusion and Discussions

Conclusion

We established a two-step modeling approach under the Bayesian framework designed for China's national and provincial mortality. We utilize the relative modeling that exploits all information available, is parsimonious, and yields biologically reasonable projections.

Discussion

1. Overdispersion parameter $\left(s_x^{(i)}\right)^2$
    - Smooth function of x
    - Alternative interpretation, the error term of $\ln m_{x,t}^{(i)}$ is

$$\hat{\epsilon}_{x,t}^{(i)} = \epsilon_{x,t} - \epsilon_{x,t}^{(i)}$$

2. Knots selection, Bayesian parameter shrinkage techniques can be applied.
3. Economic variable, such as GDP per capita.

Thank You!