Mortality Model for Multi-Populations: A Semiparametric Comparison Approach

Lei Fang
Wolfgang Karl Härdle
Juhyun Park

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin

http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
Motivation

Demographic Risk

- Low mortality, low fertility, global aging trend
- Mortality rate is the key to insurance and pension industry
Demographic key element: mortality

- Mortality rate: number of death/number of exposure, taken as the log transformation.
- Mortality rate: age-specific, male and female, (region-specific)
- Mortality change is more "stable" compared to fertility

Note: in following graphs, rates in different years are plotted in rainbow palette so that the earliest years are red and so on.
Demographic Risk in Japan

Figure 1: Japan female mortality trend: 1947-2012
Demographic Risk in Japan

Figure 2: Japan fertility trend: 1947-2012
Demographic risk in China

- Small sample size: 17 years
- Aging trend is inevitable
- Regional similarities between Japan and China
Demographic Risk in China

Figure 3: China female mortality trend: 1994-2010
Demographic Risk in China

Figure 4: China male mortality trend: 1994-2010
Literature

Mortality Similarity

- Hanewald (2011): the Lee-Carter mortality index $k_t$ correlates significantly with macroeconomic fluctuations in some periods.

Semiparametric Comparison Model

- Grith et al. (2013): shape invariant model.
Multi-Population Mortality Modeling

China
- Is there mortality similarity between China and Japan?
- How can the mortality modeling and forecasting be improved via Japan?

Multi-Countries
- How do we generate a multi-population mortality model based on the common shape?
Outline

1. Motivation ✓
2. FDA-based mortality model
3. Semiparametric comparison model
4. Empirical results
5. Outlook
6. Reference
Lee-Carter (LC) Method

- A benchmark in demographics: Lee and Carter (1992)
- Idea: use SVD to extract a single time-varying index of mortality/fertility rate level
- Take mortality for analysis:

\[
\log\{y_t(x)\} = a_x + b_x k_t + \varepsilon_{x,t}
\]

- \(y_t(x)\): observed mortality rate at age \(x\) in year \(t\)
- \(a_x\): age pattern averaged across years
- \(b_x\): first PC reflecting how fast the mortality changes at each age
- \(k_t\): time-varying index of mortality level
- \(\varepsilon_{x,t}\): residual at age \(x\) in year \(t\)
Hyndman-Ullah (HU) Method

- Variant of LC method: presmooth, orthogonalize, forecast
- Estimate the smooth functions $s_t(x)$ through the data sets $\{x, y_t(x)\}$ for each $t$:

$$y_t(x) = s_t(x) + \sigma_t(x)\varepsilon_t$$

- $s_t(x)$ smooth function
- $\sigma_t(x)$ smooth volatility function of $y_t(x)$
- $\varepsilon_t$ i.i.d. random error
Hyndman-Ullah (HU) Method

Use functional principal component analysis (FPCA)

\[ s_t(x) = \mu(x) + \sum_{k=1}^{K} \beta_{t,k} \phi_k(x) + e_t(x) \]

- \( \mu(x) \) mean of \( s_t(x) \) across years
- \( \phi_k(x) \) orthogonal basis functional PCs
- \( \beta_{t,k} \) uncorrelated PC scores
- \( e_t(x) \) is residual function with mean zero
Demographic Data

- Japan
  Mortality: age-specific (0,110+), male and female
  Extracted ages: (0,90)
  Years: 1947-2012
  Data Source: Human Mortality Database

- China
  Mortality: age-specific (0,90+), male and female
  Years: 1994-2010
  Data Source: China Statistical Year Book
Mortality trends comparison

- Time-varying indicator $k_t$ derived from Lee-Carter model presents similar pattern.

Figure 5: China mortality trend vs. Japan mortality trend: female, male.
Mortality trends comparison

Intuitive comparison: time delay between China and Japan female mortality trend.

Figure 6: Japan trend, Japan smoothed trend, China trend and China smoothed trends of no-delay, 20-, 23- and 25- year delay respect.
Semiparametric comparison model

- Use $k_t$ derived from LC model
  \[
  \log\{y_t(x)\} = a_x + b_x k_t + \varepsilon_{x,t},
  \]  
  (1)

- Infer China’s mortality trend via Japan’s trend
  \[
  k_c(t) = \theta_1 k_j\left(\frac{t - \theta_2}{\theta_3}\right) + \theta_4,
  \]  
  (2)

- $k_c(t)$ is the time-varying indicator for China
- $k_j(t)$ is the time-varying indicator for Japan
- $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T$ are shape deviation parameters
Understanding $\theta$

$\theta = (\theta_1, \theta_2, \theta_3, \theta_4)^T = (1, \theta_2, 1, \theta_4)^T$

- $\theta_1$ is the general trend adjustment, possibly selected as 1.
- $\theta_2$ is the time-delay parameter
- $\theta_3$ is the time acceleration parameter, possibly selected as 1.
- $\theta_4$ is the vertical shift parameter

**Figure 7:** Time delay $\theta_2 = 23$

**Figure 8:** Vertical shift $\theta_4 = -85$
Model estimation

Estimation procedure

\[
\min_{\theta} \int_{tc} \left\{ \hat{k}_c(u) - \theta_1 \hat{k}_j \left( \frac{u - \theta_2}{\theta_3} \right) - \theta_4 \right\}^2 w(u) du, \tag{3}
\]

- \( \hat{k}_c(t) \) and \( \hat{k}_j(t) \) are the nonparametric estimates of the original time-varying indicators, \( tc \) is the China data’s time interval
- the comparison region satisfies the condition

\[
w(u) = \prod_{tj} 1_{[a, b]} \left\{ \left( u - \theta_2 \right) / \theta_3 \right\},
\]

where \( t_j \) is the time interval of Japan’s mortality data, \( a \geq \inf(t_j) \) and \( b \leq \sup(t_j) \).
Initial choice of $\theta_2$ and $\theta_4$

- Potential linear relation between $\theta_2$ and $\theta_4$.

Figure 9: Loss surface of $\theta_2$ and $\theta_4$.

Figure 10: Contour of $\theta_2$ and $\theta_4$. 

Mortality Model for Multi-Populations
Time delay or vertical shift

- Stick with time delay influence $\theta_2$, and the optimal value is obtained around 23.

Figure 11: Loss function of $\theta_2$ with $(\theta_1, \theta_3, \theta_4)^\top = (1, 1, 0)^\top$
Algorithm

- Iterate based on the scheme (3)
- The prior estimates are $\theta = (1, 1, 23, 1, 0)^\top$ and the nonparametric estimates of $\hat{k}_c(t)$ and $\hat{k}_j(t)$
- Update $(\theta_1, \theta_2, \theta_3, \theta_4)^\top$
- Convergence is reached fast
Empirical results

**Goodness of Fit**

- Optimal $\theta = (1.160, 23.032, 1.000, -0.057)^T$

![Graph showing Goodness of Fit: Japan trend, Japan smoothed trend, China trend, China smoothed trend and fitted trend (black dots).](image)

**Figure 12:** Goodness of Fit: Japan trend, Japan smoothed trend, China trend, China smoothed trend and fitted trend (black dots).
Forecast

- Forecasting $k_t$ for China

\[
k_c(t + i) = \theta_1 k_j \left\{ \frac{(t + i) - \theta_2}{\theta_3} \right\} + \theta_4,
\]

- $\theta = (1.160, 23.032, 1.000, -0.057)^T$
- $t = 1994, 1995, \ldots, 2010; i = 1, 2, \ldots, 20$

Figure 13: Forecast of China’s mortality trend from 2011 to 2030.
Outlook

- Longer forecasting period
- Variance and confidence interval
- Mortality trend delay with economic development
- Multi-populations model with common feature
Mortality Model for Multi-Populations: A Semiparametric Comparison Approach

Lei Fang
Wolfgang Karl Härdle
Juhyun Park

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin

http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
References

M. Grith, W. Härdle and J. Park
Shape Invariant Modeling of Pricing Kernels and Risk Aversion

K. Hanewald
Explaining mortality dynamics: the role of macroeconomic fluctuations and cause of death trends

W. Härdle and J.S. Marron
Semiparametric comparison of regression curves
References

R. J. Hyndman and H. Booth
Stochastic Population Forecasts using Functional Data Models for Mortality, Fertility and Migration
International Journal of Forecasting, 2008

R. J. Hyndman and Md. S. Ullah
Robust Forecasting of Mortality and Fertility Rates: A Functional Data Approach
Computational Statistics and Data Analysis, 2007

R. D. Lee and L. R. Carter
Modeling and Forecasting U.S. Mortality
Journal of the American Statistical Association, 1992

Mortality Model for Multi-Populations