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

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

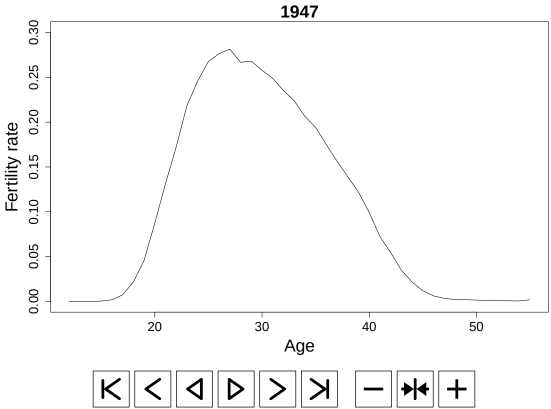

Figure 2: Japan fertility trend: 1947-2012

Demographic risk in China

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

Demographic Risk in China

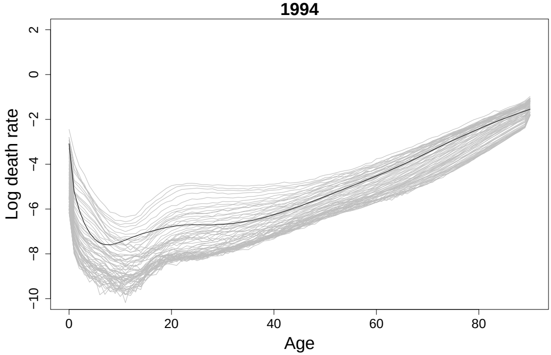

Figure 3: China female mortality trend: 1994-2010, Japan's historical female mortality is displayed as grey zone.

Demographic Risk in China

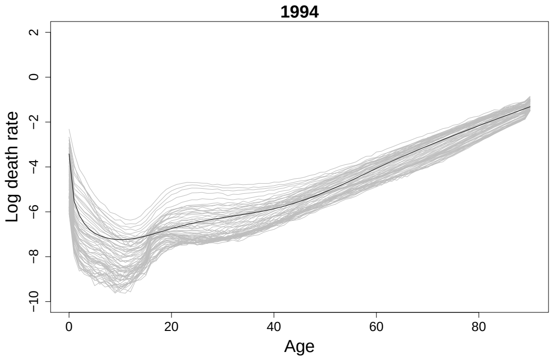

Figure 4: China male mortality trend: 1994-2010, Japan's historical male mortality is displayed as grey zone.

Literature

Mortality Similarity

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

Semiparametric Comparison Model

- Härdle and Marron (1990): Semiparametric comparison of regression curves
- 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. Classic mortality models
3. Semiparametric comparison model
4. Empirical analysis
5. 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 $y_t(x)$
  - ε_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

Mortality Analysis

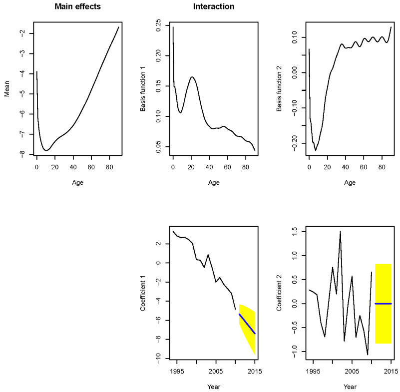

Figure 5: China's female mortality decomposition by HU Method: yellow areas represent the 95% confidence intervals for the coefficients forecast.

Mortality trends comparison

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

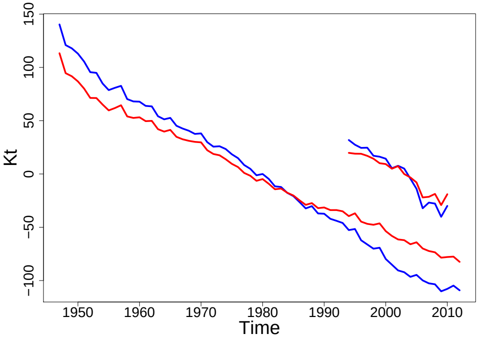

Figure 6: China mortality trend (short curves) vs. Japan mortality trend (long curves): female, male.

Semiparametric comparison model
two-country case

Take China and Japan for example

- Use k_t derived from LC model

$$\log\{y_t(x)\} = a_x + b_x k_t + \varepsilon_{x,t}, \tag{1}$$

- Infer China's mortality trend via Japan's trend

$$k_c(t) = \theta_1 k_j \left(\frac{t - \theta_2}{\theta_3} \right) + \theta_4, \tag{2}$$

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

Model estimation

- Estimation procedure

$$\min_{\theta} \int_{t_c} \left\{ \hat{k}_c(u) - \theta_1 \hat{k}_j \left(\frac{u - \theta_2}{\theta_3} \right) - \theta_4 \right\}^2 w(u) du, \qquad (3)$$

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

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

where tj is the time interval of Japan's mortality data, $a \geq inf(tj)$ and $b \leq sup(tj)$.

Algorithm

- Iterate based on the scheme (3)
- Set up the prior estimates $\theta^0 = (\theta_1^0, \theta_2^0, \theta_3^0, \theta_4^0)^\top$ and the nonparametric estimates of $\hat{k}_c(t)$ and $\hat{k}_j(t)$
- Update $(\theta_1, \theta_2, \theta_3, \theta_4)^\top$
- Reach convergence

Semiparametric comparison model multi-country case

- $k_i(t)$ is a derived time-varying mortality indicator for country i, with $i \in \{1, ..., n\}$, $n = 36$ stands for 36 countries.
- The curves can be represented in the form

$$k_i(t) = \theta_{i1} k_0 \left(\frac{t - \theta_{i2}}{\theta_{i3}} \right) + \theta_{i4}, \tag{4}$$

  - $k_i(t)$ is the time-varying indicator for country i
  - $k_0(t)$ is a reference curve, understood as common trend
  - $\theta = (\theta_{i1}, \theta_{i2}, \theta_{i3}, \theta_{i4})^\top$ are shape deviation parameters

Estimation of Common Trend

- Synchronization

$$k_i(\theta_{i3}t + \theta_{i2}) = \theta_{i1}k_0(t) + \theta_{i4}, \tag{5}$$

- Identification conditions (normalize)

$$T^{-1}\sum_{i=1}^{N}\theta_{i1} = T^{-1}\sum_{i=1}^{N}\theta_{i3} = 1, \tag{6}$$

$$T^{-1}\sum_{i=1}^{N}\theta_{i2} = T^{-1}\sum_{i=1}^{N}\theta_{i4} = 0 \tag{7}$$

- Common trend curve

$$k_0(t) = T^{-1}\sum_{i=1}^{N}k_t(\theta_{i3}t + \theta_{i2}) \tag{8}$$

Initial Value and Algorithm

- Choose a group of countries with bigger sample size and set their average curve $k_0(t)$ as initial reference curve
- Repeat the iteration of two-country case and generate initial θ^0 for the other countries
- Get the common trend based on formula (8)
- Iterate based on the above procedures
- Update $(\theta_1, \theta_2, \theta_3, \theta_4)^\top$
- Reach convergence

Demographic Data

- China
  Mortality: age-specific (0,90+), male and female
  Years: 1994-2010
  Data Source: China Statistical Year Book
- The other 35 countries
  Mortality: age-specific (0,110+), male and female
  Extracted ages: (0,90)
  Years: it differs from 14 years (Chile) to 261 years (Sweden)
  Data Source: Human Mortality Database

Mortality trends comparison

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

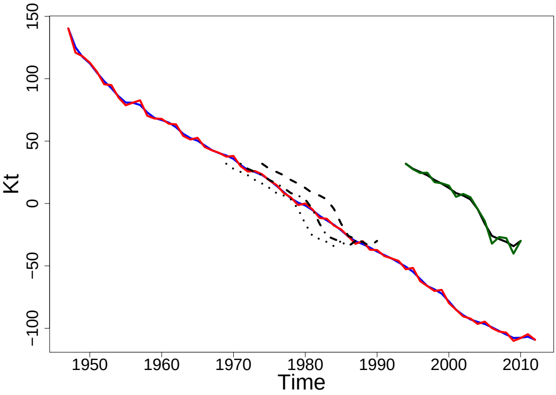

Figure 7: Japan trend, Japan smoothed trend, China trend and China smoothed trends of no-delay, 20-, 23- and 25- year delay respect[illegible]

Understanding θ

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

- θ_1 is the general trend adjustment, possibly selected as 1.
- θ_2 is the time-delay parameter
- θ_3 is the time acceleration parameter, possibly selected as 1.
- θ_4 is the vertical shift parameter

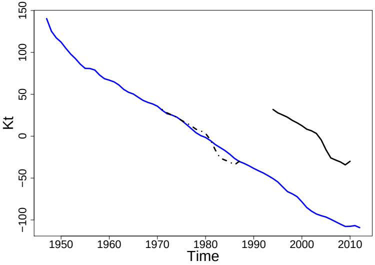

Figure 8: Time delay $\theta_2 = 23$

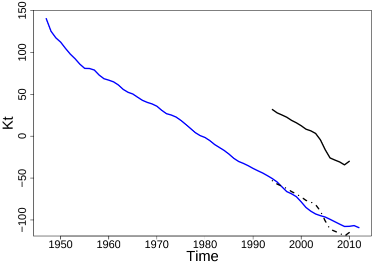

Figure 9: Vertical shift $\theta_4 = -85$

Initial choice of θ_2 and θ_4

- Potential linear relation between θ_2 and θ_4.

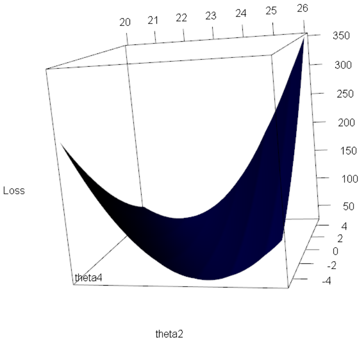

Figure 10: Loss surface of θ_2 and θ_4.

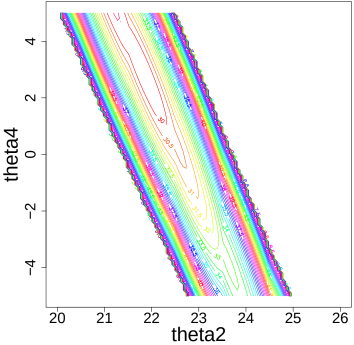

Figure 11: Contour of θ_2 and θ_4.

Time delay or vertical shift

- Stick with time delay influence θ_2, and the optimal value is obtained around 23.

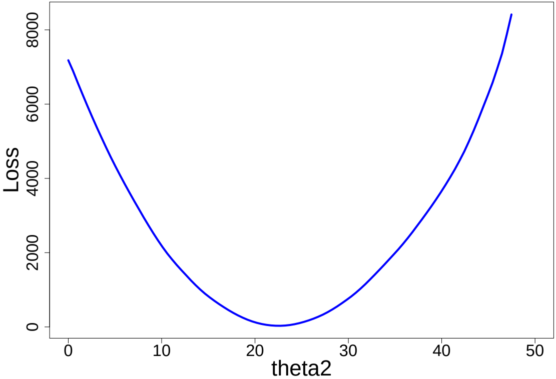

Figure 12: Loss function of θ_2 with $(\theta_1, \theta_3, \theta_4)^\top = (1, 1, 0)^\top$.

Goodness of Fit

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

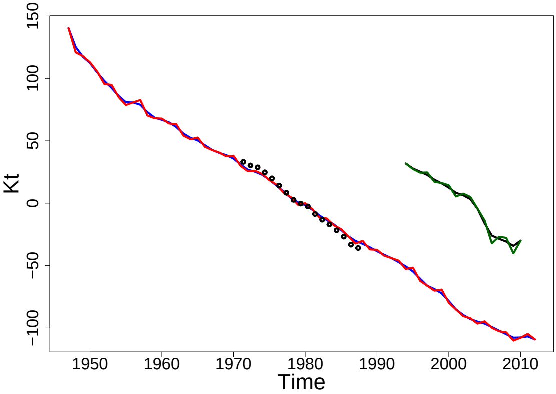

Figure 13: Goodness of Fit: Japan trend, Japan smoothed trend, China trend, China smoothed trend and fitted trend (black dots).

Forecast

- Forecasting k_t for China

$$k_c(t+i) = \theta_1 k_j \left\{ \frac{(t+i) - \theta_2}{\theta_3} \right\} + \theta_4, \tag{9}$$

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

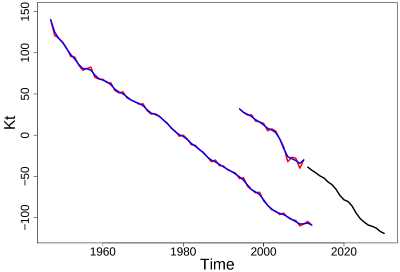

Figure 14: Forecast of China's mortality trend from 2011 to 2030.

Multi-Populations Case

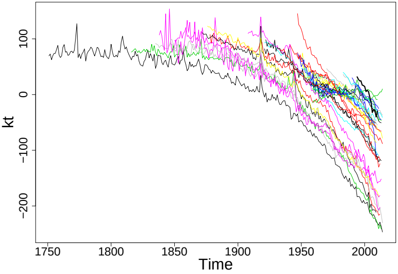

Figure 15: Original mortality trend among 36 countries

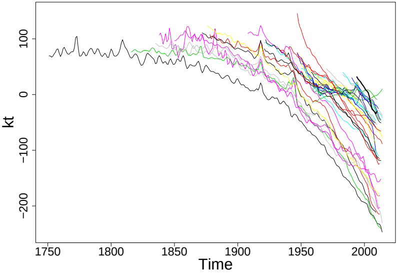

Figure 16: Original mortality trend among 36 countries

Multi-Populations Case

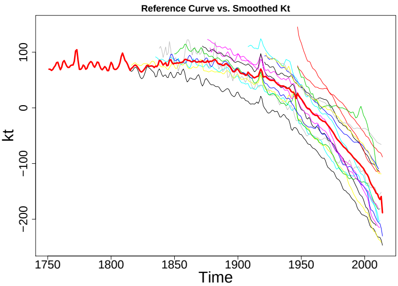

Figure 17: Reference curve vs. original smoothed $k_i(t)$

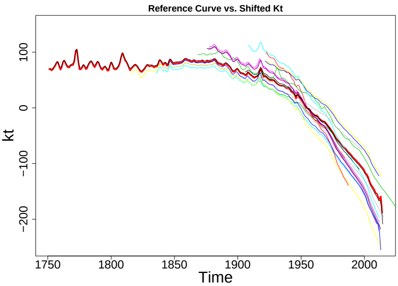

Figure 18: Reference curve vs. shifted $k_i(t)$

Multi-Populations Case

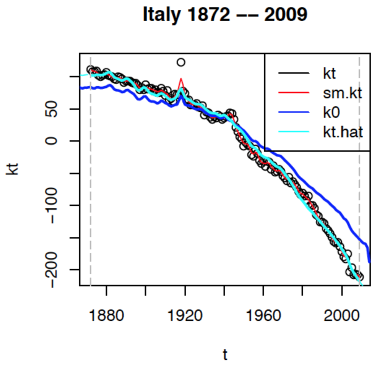

Figure 19: Italy shifted $\hat{k}_t$ according to reference curve k_0.

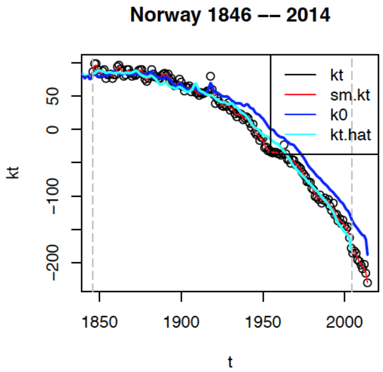

Figure 20: Norway shifted $\hat{k}_t$ according to reference curve k_0.

Outlook

- Global common mortality trend
- Confidence interval for forecast with multi-populations mortality model
- Comparison with classical mortality methods

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
  Journal of Financial Econometrics, 2013

- K. Hanewald
  Explaining mortality dynamics: the role of macroeconomic fluctuations and cause of death trends
  North American Actuarial Journal,2011

- W. Härdle and J.S. Marron
  Semiparametric comparison of regression curves
  Annals of Statistics, 1990

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