Using Graduation to Model Mortality of Small Areas

Thirteenth International Longevity Risk and Capital Markets Solutions Conference

Sep 21, 2017

Hsin-Chung Wang, Aletheia University

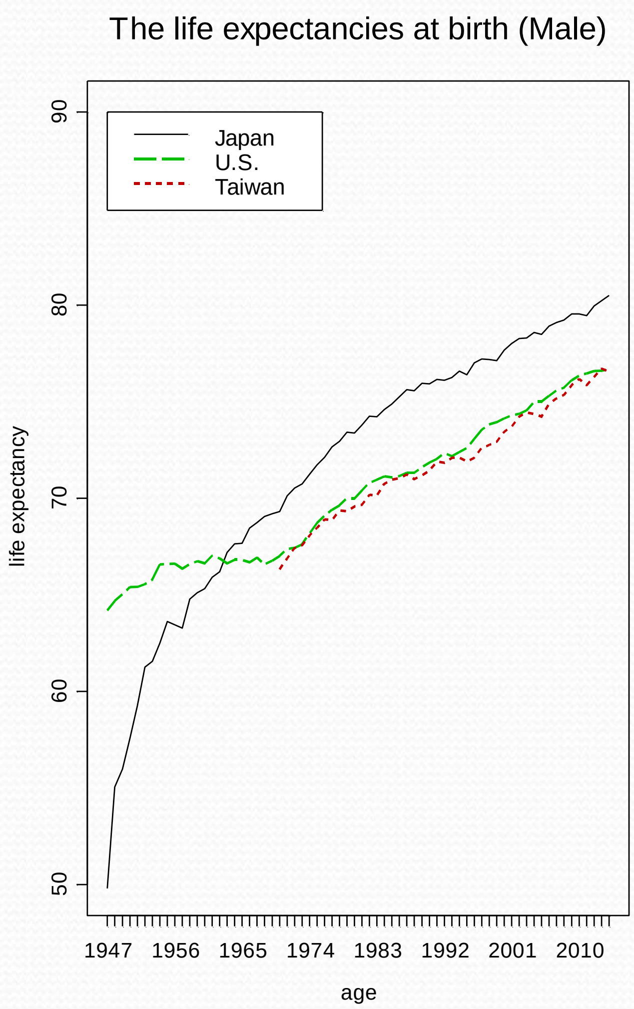
Jack C. Yue, National Chengchi University

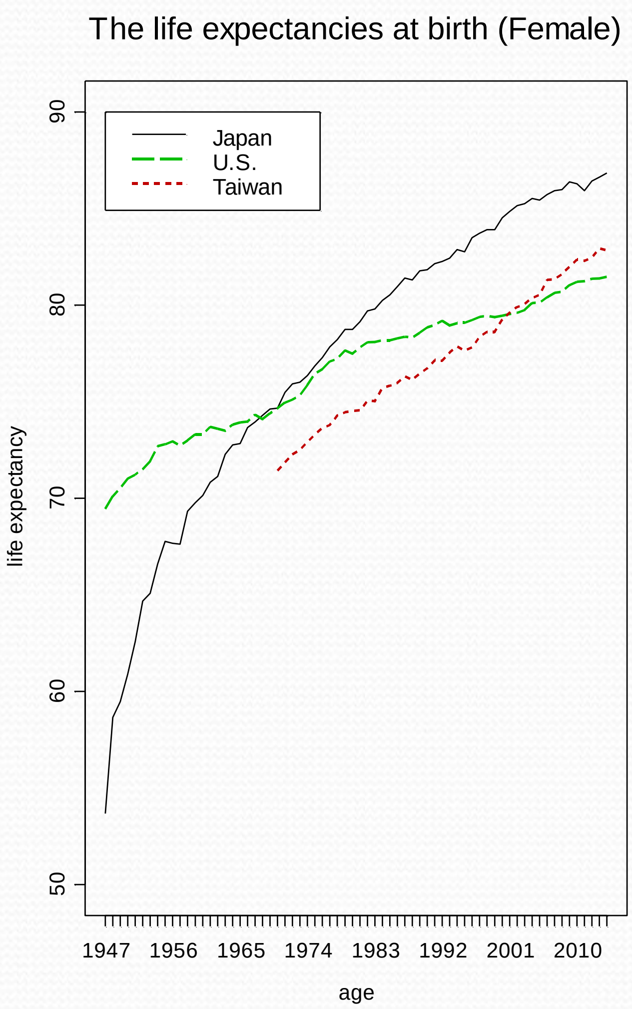
Tzu-Yu Wang, National Chengchi University

Outline

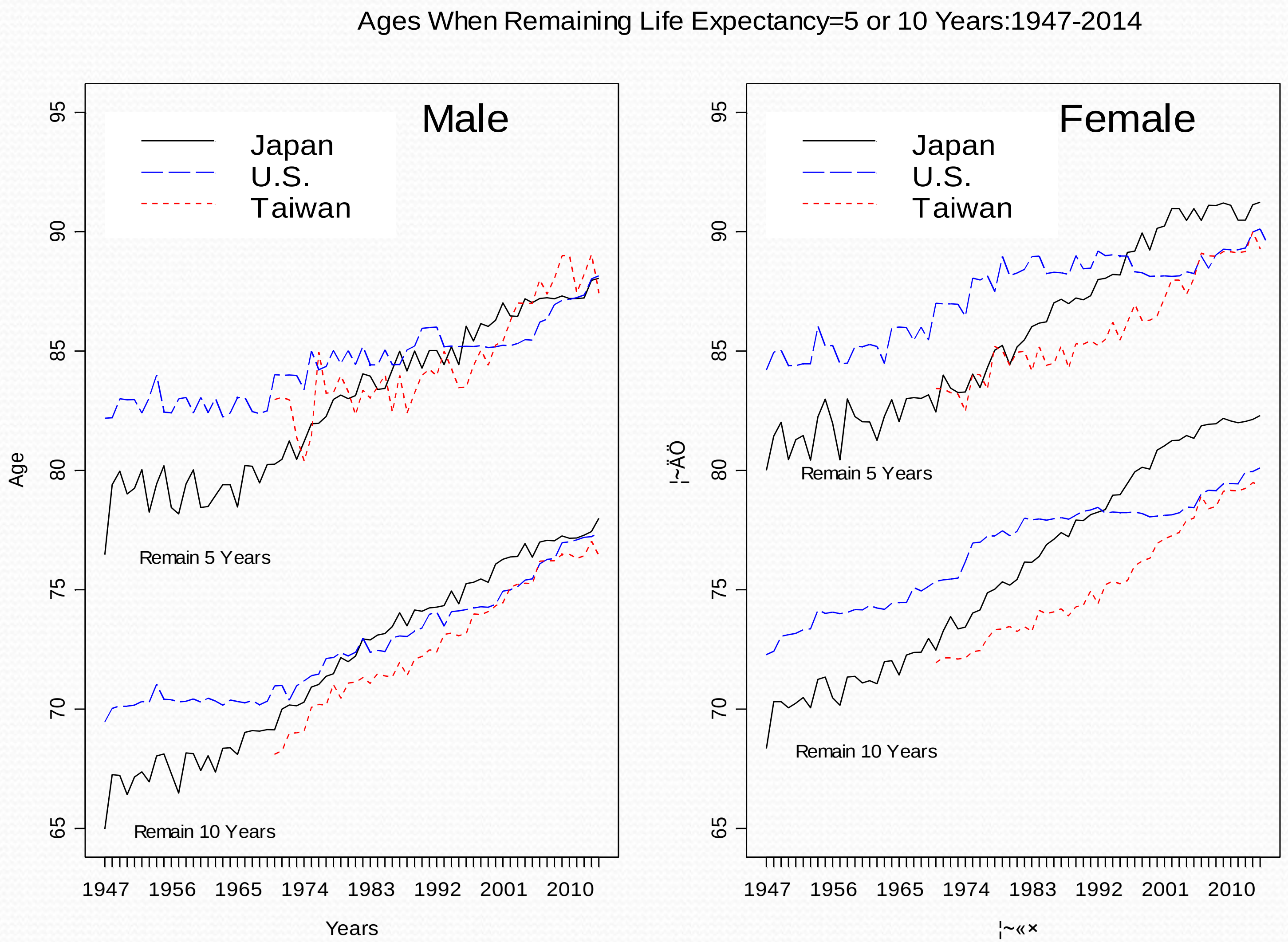
- Small Populations and Data Quality
- Graduation and Mortality Models
- Simulation and the Proposed Approach
- Comparisons to the Mortality Models
- Conclusion and Discussions

Trend of Life Expectancy in Japan, U.S. &Taiwan





Remaining Life Expectancy in Japan, U.S. &Taiwan



●Living longer（population aging）

- The future life expectancy and mortality rates of the elderly are of interest.
- The social welfare (especially, pension and national health insurance) for the elderly has become a popular issue around the world.
- migration is another common global trend in the 21st century, partly to compensate the shortage of labor force (i.e., not restricted to marriage migration).

- Longevity risk

1.Building mortality models is one of the possibilities for dealing the longevity risk.

➔There are not enough elderly data. It is more difficult for countries with small populations.

2.How to balance the losses of population and work force?

➔Population migration is one of the key factors.

- Derived migration phenomenon (welfare effects and labor migration of local government policy)

population distribution inequality in city and county

-Urbanization became more obvious

-Fewer international immigrations in Taiwan

➔small population in county levels.

- Data Quality

Ex1: The population records of highest attained age in Taiwan are 100 years old in 1998 (95 years old in 1995), i.e., limited data points for modeling the elderly mortality rates.

●Ex2: Response rates of Taiwan’s 2003 Elderly Census

➔Elderly Data quality is not good.

	Age 89	Ages 90~94	Ages 95~99	Ages 100+	Total
Total Population	12,597	23,898	11,190	2,484	50,169
Response	92%	91%	88%	66%	90%
Non-response	8%	9%	12%	34%	10%

- Increasing population size is more intuitive.
  - Aggregate more years of data
  - Combine data from countries with similar mortality profile (how to measure?)

However, it is often not possible to collect more data and thus the graduation methods are used instead.

Note: Variance $\propto$ 1/(Sample size) and it may require lots of data aggregation to become stable.

Life Table Construction

- Smoothing the mortality rates (or graduation) is often necessary in constructing life tables to reduce the mortality fluctuations between ages.
- Traditional graduation often "combines" data from adjacent ages around the target age.

Ex: moving weighted average (London, 1985), can be treated as augmenting mortality data from neighboring ages.

Ex: Bayesian graduation: using the mortality rates of reference population for smoothing.

Limitations of Graduation Methods

●The graduation results depend on the chosen methods (& parameters), as well as on the reference population(s).

➔It is not easy to predict mortality rates using the traditional graduation methods.

●Sample size is critical in choosing appropriate graduation methods.

➔Aggregating 3, 5, or 9 years of data are often necessary in practice.

Graduation vs. Mortality Models

- Mortality models can be treated as a class of graduation methods.

➔They can also be used for mortality prediction.

- The need for increasing the sample size (or graduation) also appears in improving the parameter estimation of mortality models.

Ex: The coherent Lee-Carter model by Li and Lee (2005) can reduce the estimation errors by referencing the mortality data from populations with similar mortality improvements.

Biased Estimates of Mortality Models

●Small population size can also cause problems in parameter estimation for mortality models.

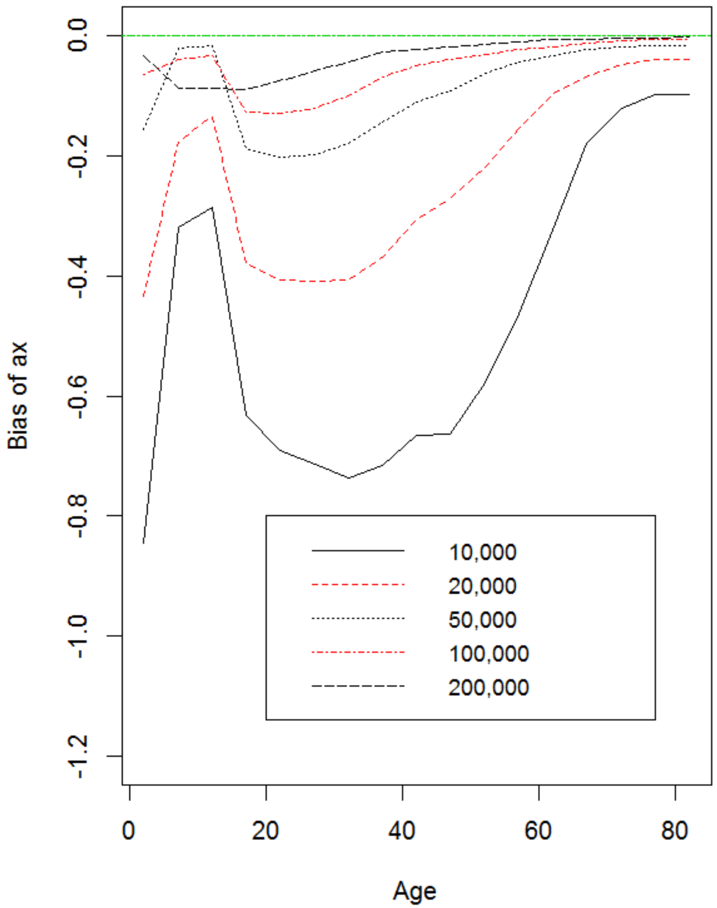
Ex: According to our simulation : Lee-Carter model have biased estimates for age-related parameters α_x and β_x in small populations sizes, especially when the size is less than 200,000.

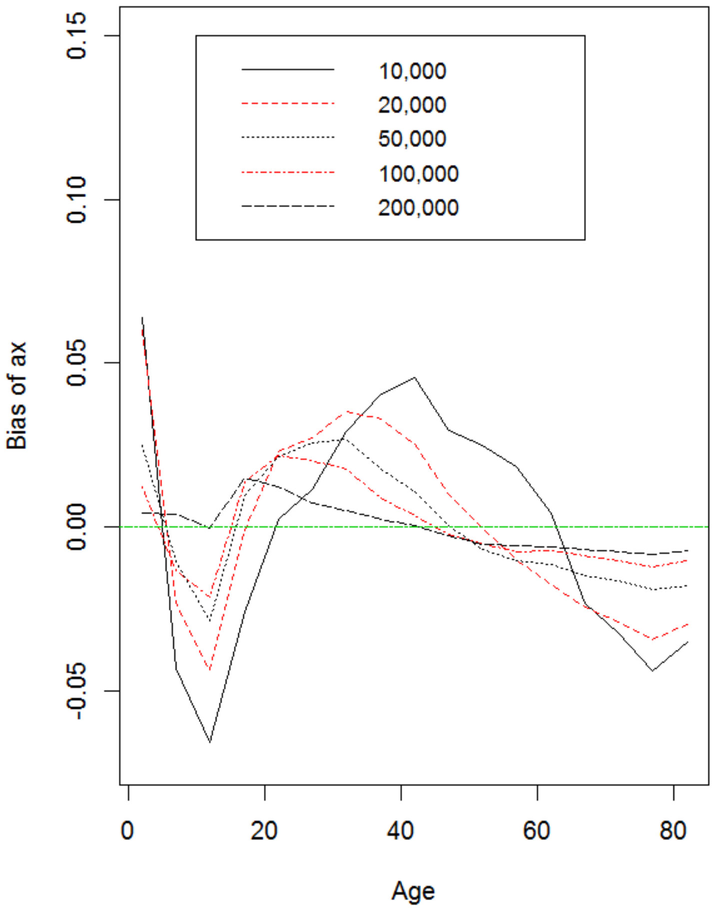
Ex: applying the Lee-Carter to small populations (such as Nordic countries and Denmark) would produce estimation bias in age-related parameters (Booth et al., 2006; Jarner and Kryger, 2011).

Simulation setting

- We use Taiwan mortality data in 1990-2009 to derive the parameters of Lee-Carter model , with the age range 0-84.
- We generate numbers of deaths based on Poisson assumption, producing simulated age-specific mortality rates under 10,000, 20,000, …, 2 million, and 5 million population sizes.
- Calculate the averages and standard errors of Lee-Carter model parameters and from 1,000 simulation runs
- Comparison criterion: $\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\%$

Bias of Parameters' Estimates of the Lee-Carter Model





Increasing the Sample Size?

- Increasing the sample size is the most effective way to deal with the problem of small population.

Question: How do we increase the sample size?

- Traditional graduation is to include data from nearby ages (e.g., same age for 3 or 5 consecutive years, ages x–1~x+1 or x–2~x+2 for single year)
- Borrowing information from larger populations has been used previously.-Li and Lee (2005).

But it is difficult to decide if the small and large populations are similar.

Graduation Methods

- Whittaker and the partial SMR can reduce the fluctuation of mortality rates.
- Graduation methods can stabilize the mortality rates and possibly reduce the bias of parameter estimation of mortality models.

➔We proposed using the graduation methods (Whittaker and Partial SMR) to borrow information from larger populations.

- Whittaker

→Minimizing the sum of Fit and Smoothness:

$$\mathrm{M} = \mathrm{F} + hS = \sum_{x=1}^{n} w_x (v_x - u_x)^2 + h \sum_{x=1}^{n-z} (\Delta^z v_x)^2$$

- Partial SMR (Standard Mortality Ratio)

  Lee (2003) proposed using the partial SMR (connection between large and small areas) to modify the mortality rates of small area:

$$v_x = u_x^* \times \exp\left(\frac{d_x \times \hat{h}^2 \times \log(d_x / e_x) + (1 - d_x / \sum d_x) \times \log(\mathrm{SMR})}{d_x \times \hat{h}^2 + (1 - d_x / \sum d_x)} \right)$$

$$\mathrm{SMR} = \frac{\sum_x d_x}{\sum_x n_x \cdot u_x^*} \qquad \hat{h}^2 = \max\left(\frac{\sum \left((d_x - e_x \times SMR)^2 - \sum d_x \right)}{SMR^2 \times \sum e_x^2} \, , \; 0 \right)$$

Graduation + Mortality Models

- According to the data aggregation, we can classify the graduation methods into 4 groups, *same area or not* & *one year or more*.

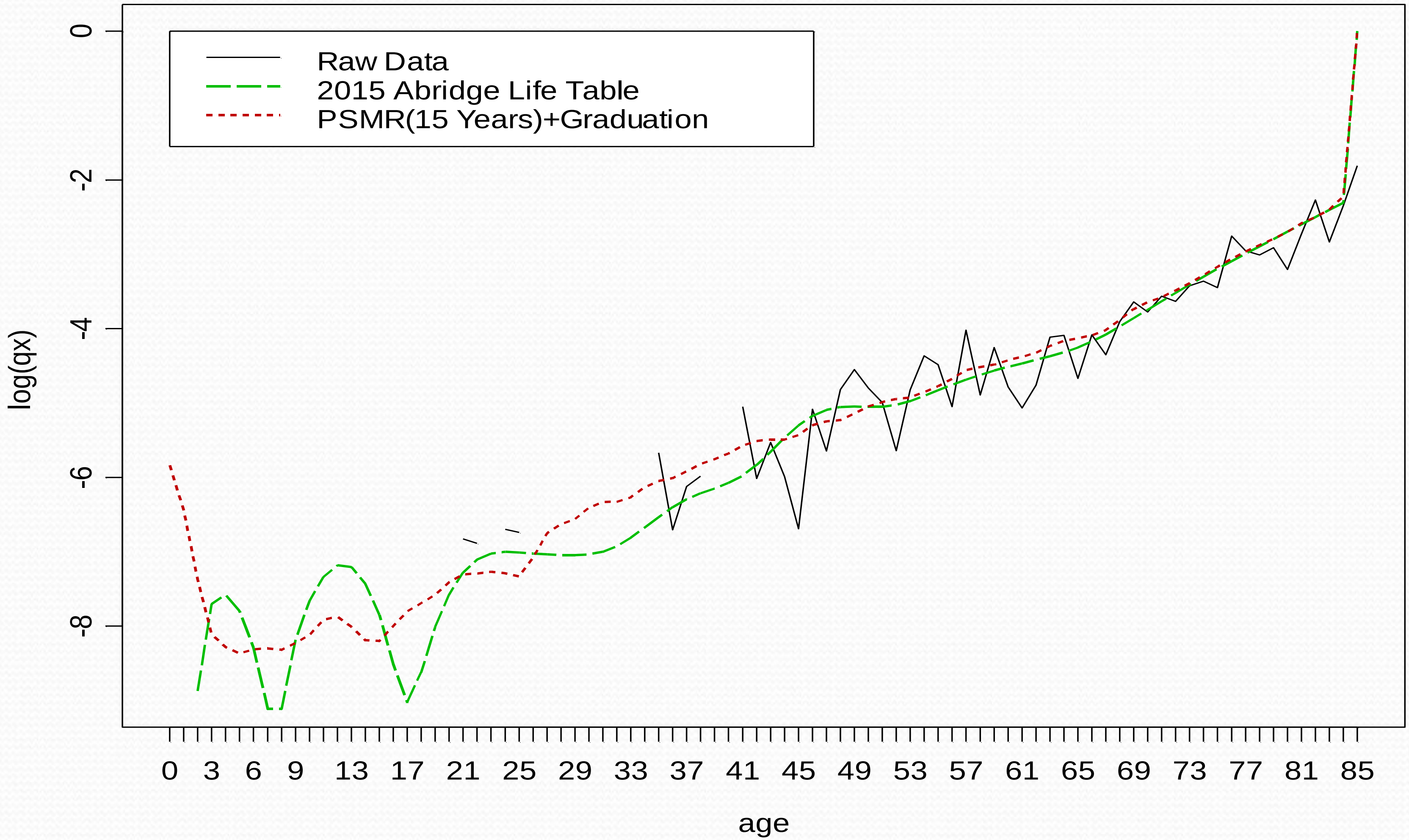
Note: We focus on (same area, multiple years) and (different area, one year)

	Same Area	Different Area
one year	MWA, Greville's 9-term formula	PSMR, Whittaker, Bayesian
multiple years	Block Bootstrap, LC model	Li-Lee model

- We can aggregate the historical mortality data for the target population (age-wise), treating the aggregation as the reference population.
- Compute the age-specific mortality ratios and then apply graduation methods to these ratios. This can achieve smoother mortality rates for small populations (e.g., Pen-Hu County, population about 100,000).

Note: Smaller populations, better results! time.

Peng-Hu County (Male)



Comparison the partial SMR and the Lee-Carter model

- MAPE of Simulation (Taiwanese Male, 1990-2009, 5-age)

(Unit: %)	10,000	20,000	50,000	100,000	200,000	500,000
Raw	68.23	50.59	32.90	22.88	16.28	10.27
Lee-Carter	33.57	23.67	15.53	10.97	8.66	6.05
Partial SMR	14.31	11.75	9.68	8.70	8.09	7.50

Comparison the partial SMR and the Lee-Carter model

- MAPE of Simulation (Taiwanese Male, 1990-2009, Single-age)

(Unit: %)	10,000	20,000	50,000	100,000	200,000	500,000
Raw	125.56	101.45	73.01	54.89	39.40	24.60
Lee-Carter	69.77	52.70	31.11	19.93	13.30	8.32
Partial SMR	18.02	14.82	12.47	11.39	10.72	10.19

Comparison the partial SMR and the Lee-Carter model

- For both the 5-age and single-age groups, the partial SMR has smaller MAPE than the Lee-Carter model when the population size is not more than 200,000.
- The advantage is more obvious for smaller population size and single-age group.
- It is interesting that the partial SMR does not suffer much for smaller population size and its MAPE is always smaller than 20%.

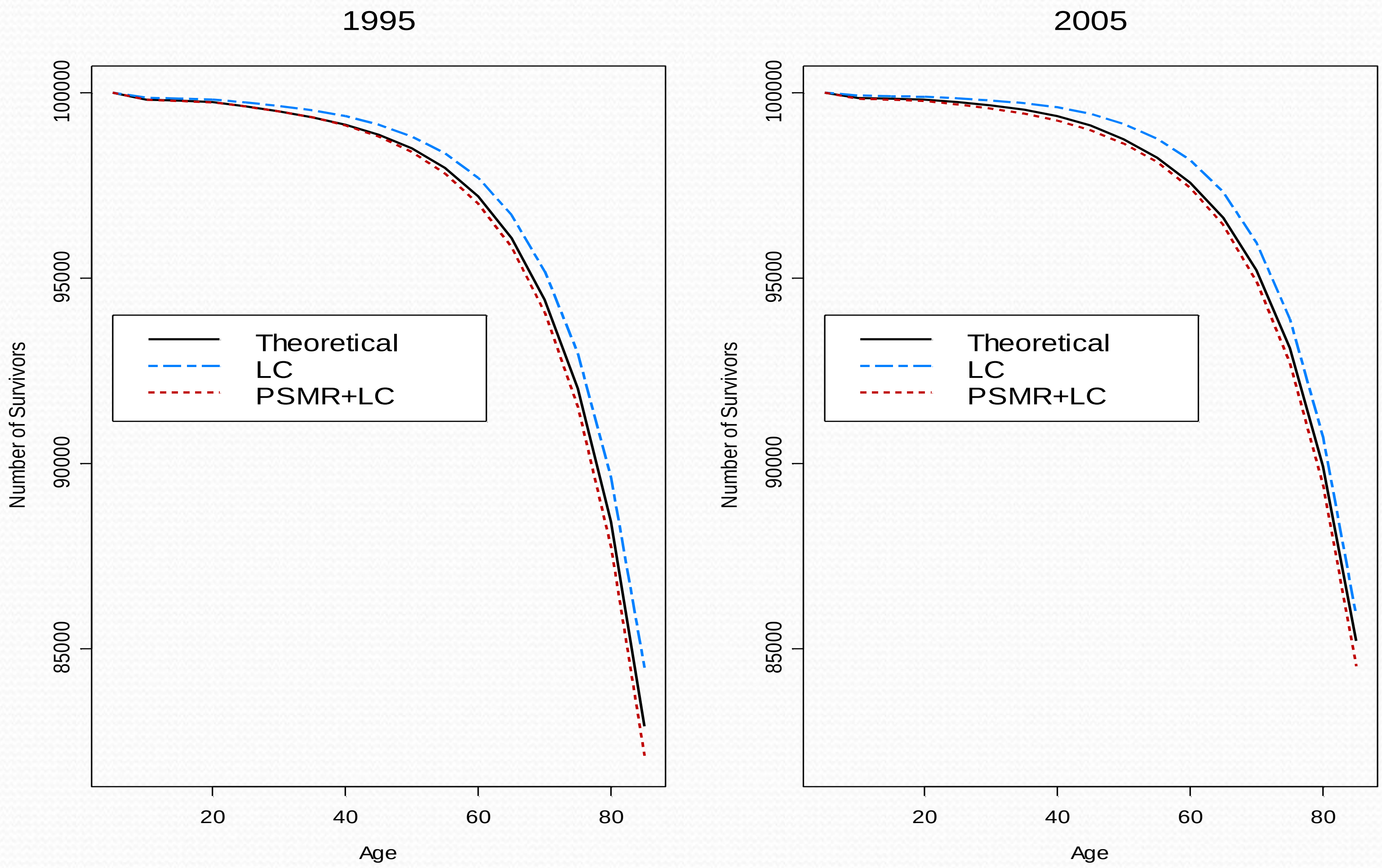
➔It seems that the partial SMR can be used to handling mortality estimation of small population.

Use the survival curve to validate the proposed approach

- The role of survival curve is similar to life expectancy and thus can link to longevity.
- We first use the theoretical survival curves in 1995 and 2005 for population size 10,000, together with their estimates via the Lee-Carter method and proposed approach, to demonstrate the differences of estimation methods

➔The areas between the theoretical survival curves and those of the proposed partial SMR are significantly smaller, which indicates that the proposed approach has a better performance in estimating the life expectancy.

Use the survival curve to validate the proposed approach



Mean Squared Error (MSE) of the parameters estimates for the Lee-Carter model

- The proposed partial SMR does provide smaller MSE's for both parameters
- The MSE's of the partial SMR do not change much for different population sizes

(Unit: 10^{-3} for α_x & 10^{-4} for β_x)

		10,000	**20,000**	**50,000**	**100,000**	**200,000**	**500,000**
α_x	Lee-Carter	169.02	47.19	10.55	4.50	2.44	0.91
	Partial SMR	3.00	1.37	0.78	0.63	0.57	0.56
β_x	Lee-Carter	20.19	13.76	8.61	4.79	2.63	1.32
	Partial SMR	0.27	0.23	0.15	0.10	0.07	0.06

Proposed Approaches

- Use the large population or aggregate of historical data from small population as the reference group.

➔The mortality curves of large and small populations (individual year) should look similar.

➔PSMR/Whittaker Ratio can be used to reduce the fluctuations of single-year mortality rates.

- Instead of raw data, we can graduate the mortality rates first, before applying the mortality models.

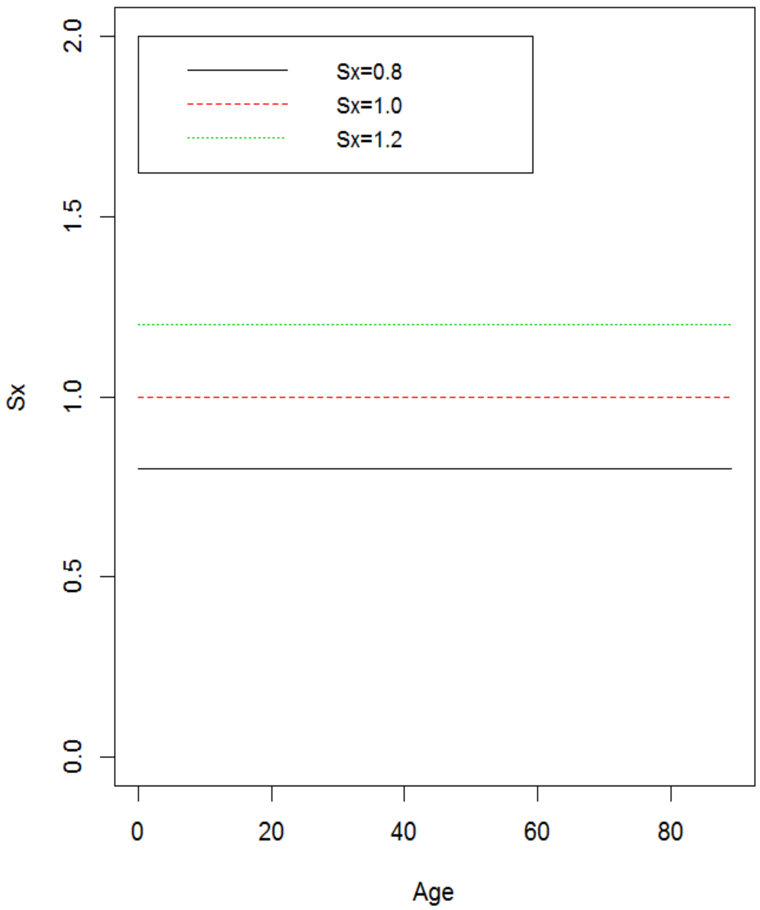
➔PSMR/Whittaker Ratio + mortality models.

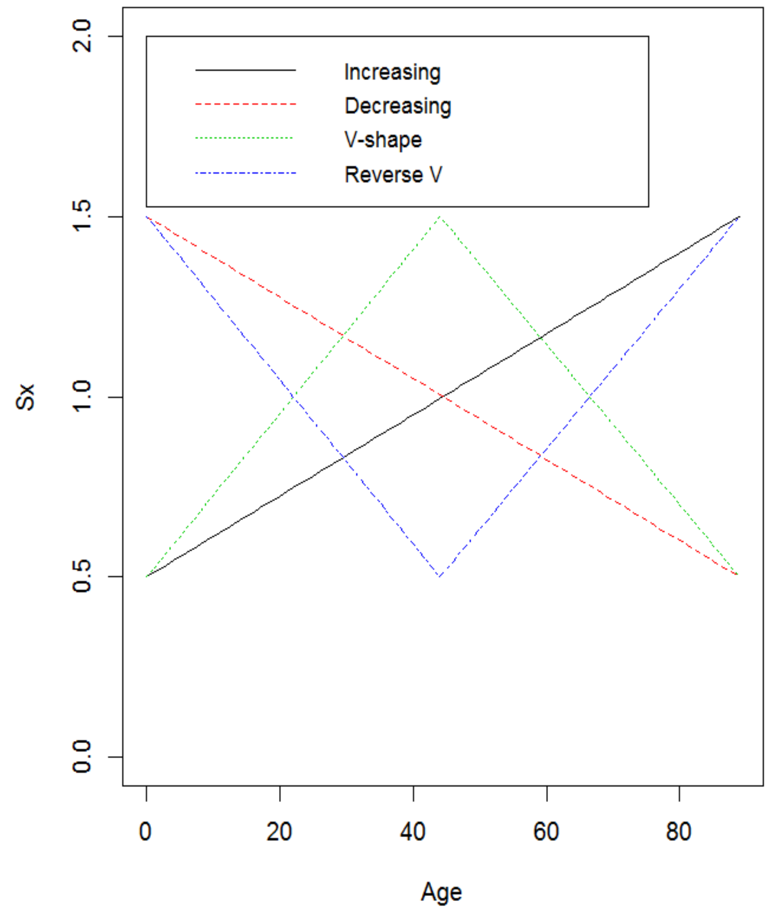
Simulation Setting

- We use Taiwanese male mortality data in 1996-2015 to derive the parameters of Lee-Carter model , with the 5-age group (0-4, 5-9, …, 95-99).
- We generate numbers of deaths based on Poisson assumption, 1,000 simulation runs to produce simulated age-specific mortality rates under 100,000, and 2 million population sizes.
- The size of reference population is 2 million or aggregate 20-year data and the mortality rates of reference population satisfy the Lee-Carter model.
- The mortality rates of small population follow one of 7 mortality scenarios:

➔Similar to the reference group (3 cases) and Differ to the reference group (4 cases

Seven Mortality Scenarios





Simulation Setting (cont.)

- Cae1:The mortality ratio $S_x = \frac{q_x^s}{q_x^R}$ where q_x^s and q_x^R are the mortality rates of small and reference populations.
- Cae2:The mortality ratio $S_x = \frac{\alpha_x^s}{\alpha_x^R}$ where α_x^s and α_x^R are the intercept parameters in the Lee-Carter model for the small and reference populations.
- Cae3:The mortality ratio $S_x = \frac{\beta_x^s}{\beta_x^R}$ where β_x^s and β_x^R are the age-related slope parameters in the Lee-Carter model for the small and reference populations.

MAPE of Lee-Carter Model Graduation (mortality ratio)

$S_x = \frac{q_x^s}{q_x^R}$	**Sx=0.8**	**Sx=1**	**Sx=1.2**	**Increasing**	**Decreasing**	**V**	**Rev-V**
Raw	29.31	27.11	25.32	29.29	27.33	27.68	28.66
LC	17.11	15.32	14.08	18.52	16.48	14.96	18.94
Li-Lee	13.29	12.25	11.75	15.69	11.49	13.20	13.79
PSMR+LC	5.02	4.68	4.49	39.24	24.26	21.36	24.17
Whitakker ratio+ LC	10.28	8.52	8.61	10.50	10.35	9.00	8.61
PSMR(Sum)+ LC	10.40	9.69	9.03	10.41	9.51	9.69	10.14
Whitakker ratio(Sum)+ LC	8.97	8.62	8.51	8.84	8.99	9.19	8.58

MAPE of Lee-Carter Model Graduation (Different α_x)

$S_x = \dfrac{\alpha_x^s}{\alpha_x^R}$	**Sx=0.8**	**Sx=1**	**Sx=1.2**	**Increasing**	**Decreasing**	**V**	**Rev-V**
Raw	66.42	27.04	15.66	317.41	18.46	28.34	173.75
LC	53.80	15.35	9.14	311.41	14.41	14.95	169.06
Li-Lee	49.44	12.26	7.52	303.80	8.49	13.25	160.53
PSMR+LC	67.67	4.72	19.71	1399.72	8.46	98.85	456.44
Whitakker ratio+ LC	42.67	8.65	5.66	41.30	5.91	9.95	36.02
PSMR(Sum)+ LC	16.58	9.68	6.75	24.35	7.34	9.80	17.45
Whitakker ratio(Sum)+ LC	11.79	8.72	5.57	9.06	5.94	9.12	7.38

MAPE of Lee-Carter Model Graduation (Different β_x)

$S_x = \frac{\beta_x^s}{\beta_x^R}$	**Sx=0.8**	**Sx=1**	**Sx=1.2**	**Increasing**	**Decreasing**	**V**	**Rev-V**
Raw	26.99	27.17	27.14	27.14	27.10	27.12	27.09
LC	15.34	15.22	14.98	15.03	15.64	15.22	15.30
Li-Lee	12.65	12.22	12.37	12.82	12.96	12.80	12.80
PSMR+LC	4.79	4.72	4.72	6.29	6.49	5.84	5.68
Whitakker ratio+ LC	8.44	8.60	8.92	8.26	8.65	8.51	8.29
PSMR(Sum)+ LC	10.34	9.74	9.06	11.97	9.37	9.45	11.22
Whitakker ratio(Sum)+ LC	8.77	8.65	8.76	8.16	9.10	8.75	8.35

Simulation Results

- Unsurprisingly, the MAPEs of Lee-Carter and Li-Lee models are obviously smaller than those of raw observations, since the mortality rates satisfy the Lee-Carter model.
- In addition, the MAPE's of Li-Lee model are always smaller than those of Lee-Carter model. It seems that the Li-Lee model is a fine modification to the Lee-Carter model
- It seems that even if the small and reference populations have quite different β_x, using the idea of coherent group to increase the population size still can reduce the estimation error of mortality rates for the small populations.

Simulation Results

- Case2: Different α_x (same β_x) :

➔We find that the mortality structure are distorted in mortality scenarios 3 (Increasing) and 7 (Rev-V)

➔ MAPE by LC model, Li-Lee model, PSMR+LC and Whitakker ratio+ LC model (under the reference population sizes is 2 million) are all higher than that in the other scenarios.

➔If we use aggregate the historical mortality data for the reference population, that is the small and large populations are similar, then PSMR(Whittaker)+LC model can

reduce the estimation errors.

Conclusion

- The idea of increasing sample size can be used in small area estimations.

→We suggest using the aggregate of historical data as the reference population.

→The proposed approach has smaller estimation errors for small areas.

Discussions and Future Study

- Can we construct a selection criterion when there are more than one reference population?
- How do we deal with the case when the trend of parameters α_x , β_x , and κ_t are all different.
  ➔ Interaction effect?
- Can we use the graduation methods to other mortality models (e.g., Age-period-cohort model) or find other approaches?

Thank you for your attention.