A Study of Smoothing Methods on Small Area Population Projection

Shuoh Jin
Jack C. Yue
National Chengchi University

Dec. 18, 2011

Summary

- Motivation
- Graduation Methods
- Simulation Studies
- Empirical Simulation
- Conclusion & Discussions

Population Aging in Taiwan

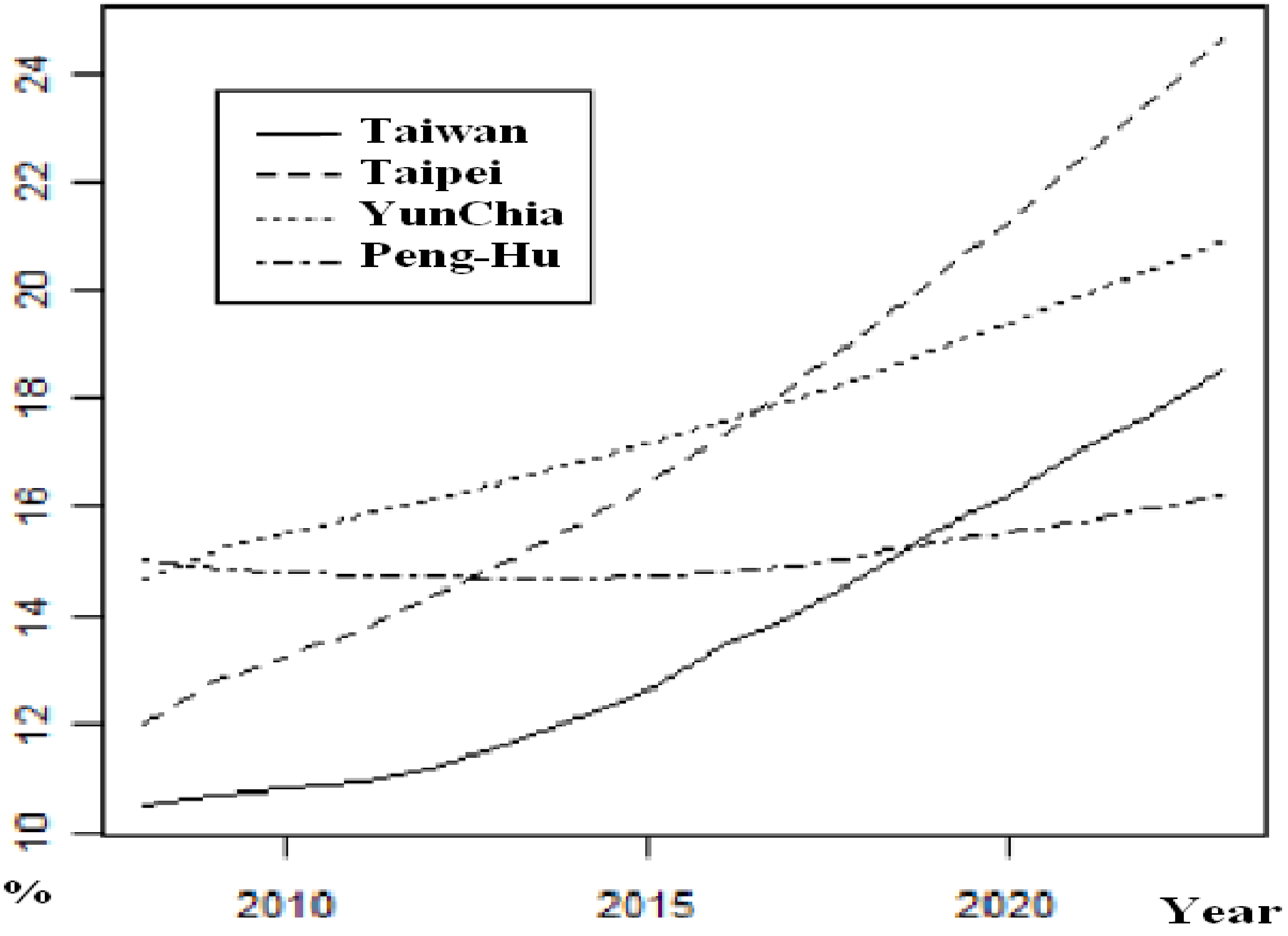
- The population aging is a common phenomenon in Taiwan, but the problem is not the same for different counties.

→The population aging in Taipei city is faster.

→Small area population projection is essential.

- It is difficult to derive a reliable projection for small areas, and even more difficult if the social attributes within the small areas vary widely.

Proportion of Ages 65+ in Taiwan (unit: %)



Important Factors

- Difficulties for small area population projection can be attributed into three directions:

→Population size;

→Number of base years;

→Projection horizon.

Note: Smith and Sincich (1990) studied the influences of “base year” and “projection horizon” for simple projection methods (e.g. linear & exponential) on large populations.

Study Objective

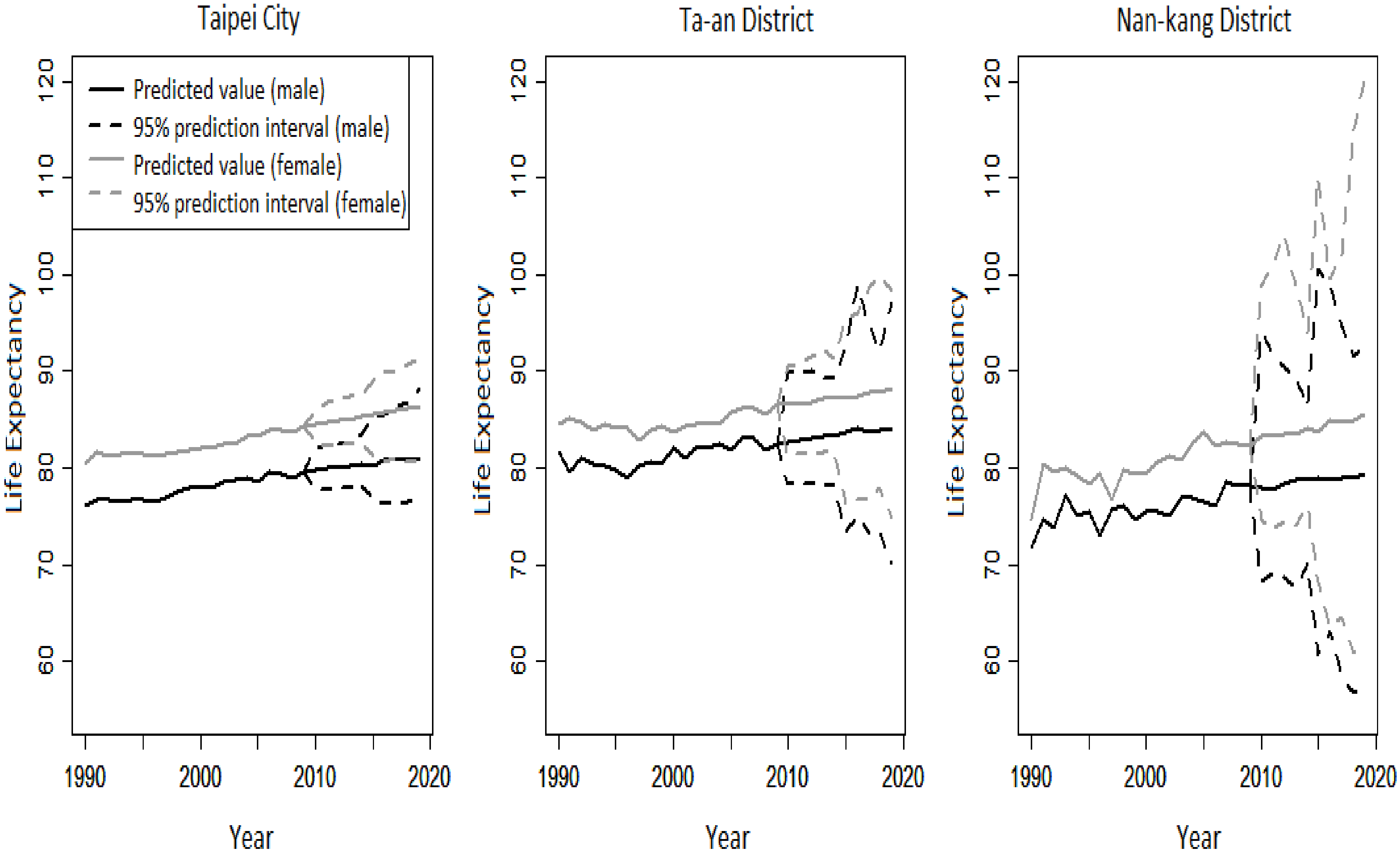
- Use smoothing techniques to improve the stability and accuracy of small area population projection.

→Consider Cohort Component method.

- Basically, smaller population size would have larger projection errors.

→e.g., We use the projection of life expectancy in Taipei city as a demonstration.

Life Expectancy Projection in Taipei City



Note: The projection is based on 1,000 block bootstrap simulation runs.

What is Graduation ?

- *Graduation may be defined as the process of obtaining from* ***irregular*** *set of observed values of a dependent variable a corresponding smooth set of values* ***consistent in a general way*** *with the observed values.*

Note: Graduation will possibly alter the original data properties and it should be applied with care.

Graduation Methods on Taiwan Elderly Mortality

Graduation (Smoothing) Methods

- Graduation has been studied for more than one century. Mathematical formulas were popular in early days (e.g., moving average) but they often suffer to the weight problem.
- Methods incorporating weight are considered:

→Whittaker (Henderson) method

→Mortality Ratio (Whittaker) method

→Partial SMR (Lee, 2003)

Graduation Methods (cont.)

- Whittaker-Henderson method:

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

→The idea of Whittaker method is similar to the Penalized Likelihood Estimation (PLE), by setting the objective function to

$$L(\theta \mid data) + \frac{\lambda}{2} J(\theta) \ \text{ or } \ \sum_{i=1}^{n} \left(\frac{Y_i - \theta(x_i)}{\sigma_i} \right)^2 + \frac{\lambda}{2} \int \ddot{\theta}(x)^2 \, dx$$

Graduation Methods (cont.)

- Mortality Ratio (Whittaker) method:

Applying Whittaker on mortality ratio $r_x = \dfrac{u_x}{u_x^*}$,

u_x^* is the mortality rates from the reference population.

Note: u_x & v_x are the raw & graduated rates.

Graduation Methods (cont.)

- Partial SMR (Standard Mortality Ratio)

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

where $$SMR = \frac{\sum_x d_x}{\sum_x n_x \cdot u_x^*}$$

$$h^2 = \max\left(\frac{\sum \left((d_x - n_x \times u_x^* \times SMR)^2 - \sum d_x \right)}{SMR^2 \times \sum (n_x \times u_x^*)^2}, 0 \right)$$

Cohort Component Method

- Cohort component method is the most popular population projection method.

→Use the results at year *t* to project for year *t+1*, i.e., $P(t+1) = S(t) \times P(t) + I(t) - E(t)$

where $P(t) = [{}_tP_0,\ {}_tP_1,\ \ldots,\ {}_tP_\omega]^T$ and

$$S(t) = \begin{bmatrix} 0 & 0 & \cdots & 0 & S_t(\alpha,\alpha+0.5)\cdot f_\alpha^t & \cdots & S_t(\beta,\beta+0.5)\cdot f_\beta^t & 0 & \cdots & 0 & 0 \\ S_t(0,1) & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ 0 & S_t(1,2) & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & S_t(w-1,w) & 0 \end{bmatrix}$$

Simulation-based Projection Method

- **Use Block Bootstrap to project fertility, mortality, migration.**

Estimation

Year	Ages 0~4	Ages 5~9	…	…
1990	$q_x(1990)$	$q_x(1990)$	…	…
⋮	⋮	⋮	⋮	⋮
2010	$q_x(2010)$	$q_x(2010)$	…	…

+

Projection

t	$\Delta q_x(t)$	$\Delta q_x(t)$	...	...
⋮	⋮	⋮	⋮	⋮
t+n	$\Delta q_x(t+n)$	$\Delta q_x(t+n)$	...	...

=

Year	Ages 0~4	Ages 5~9	...	...
2011	$q_x(2011)$	$q_x(2011)$	…	…
⋮	⋮	⋮	⋮	⋮

Lee-Carter Mortality Model

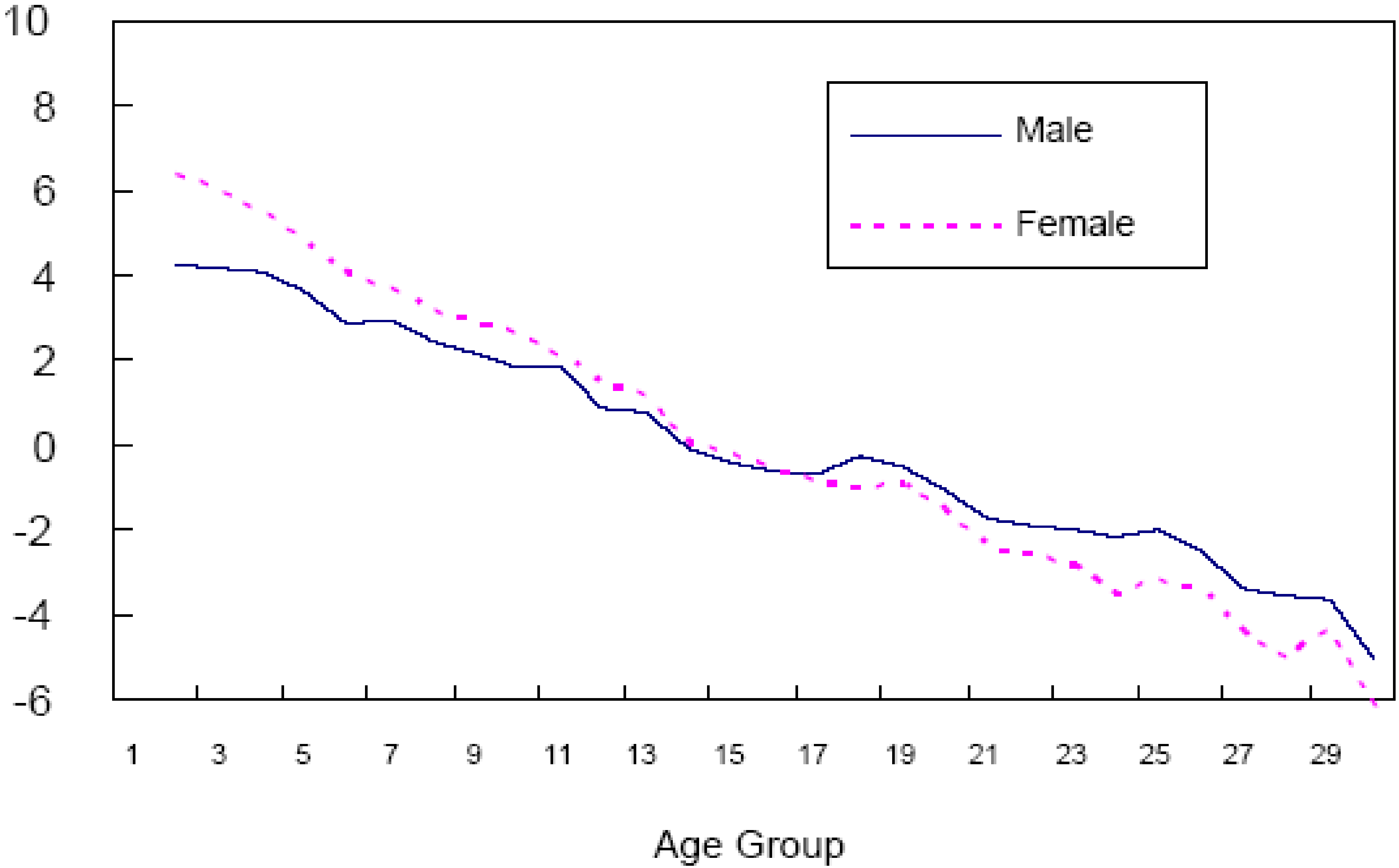
- The Lee-Carter Model (Lee and Carter, 1992), the central mortality rate should be consistent with the following equation

→

$$\ln\left(m_{x,t}\right) = \alpha_x + \beta_x \kappa_t + \varepsilon_{x,t}$$

α_x describes the average age-specific mortality, κ_t represents the general mortality level, and the decline in mortality at age *x* is captured by β_x.

κ_t Estimates of Lee-Carter Model in Taiwan



Simulation Setting

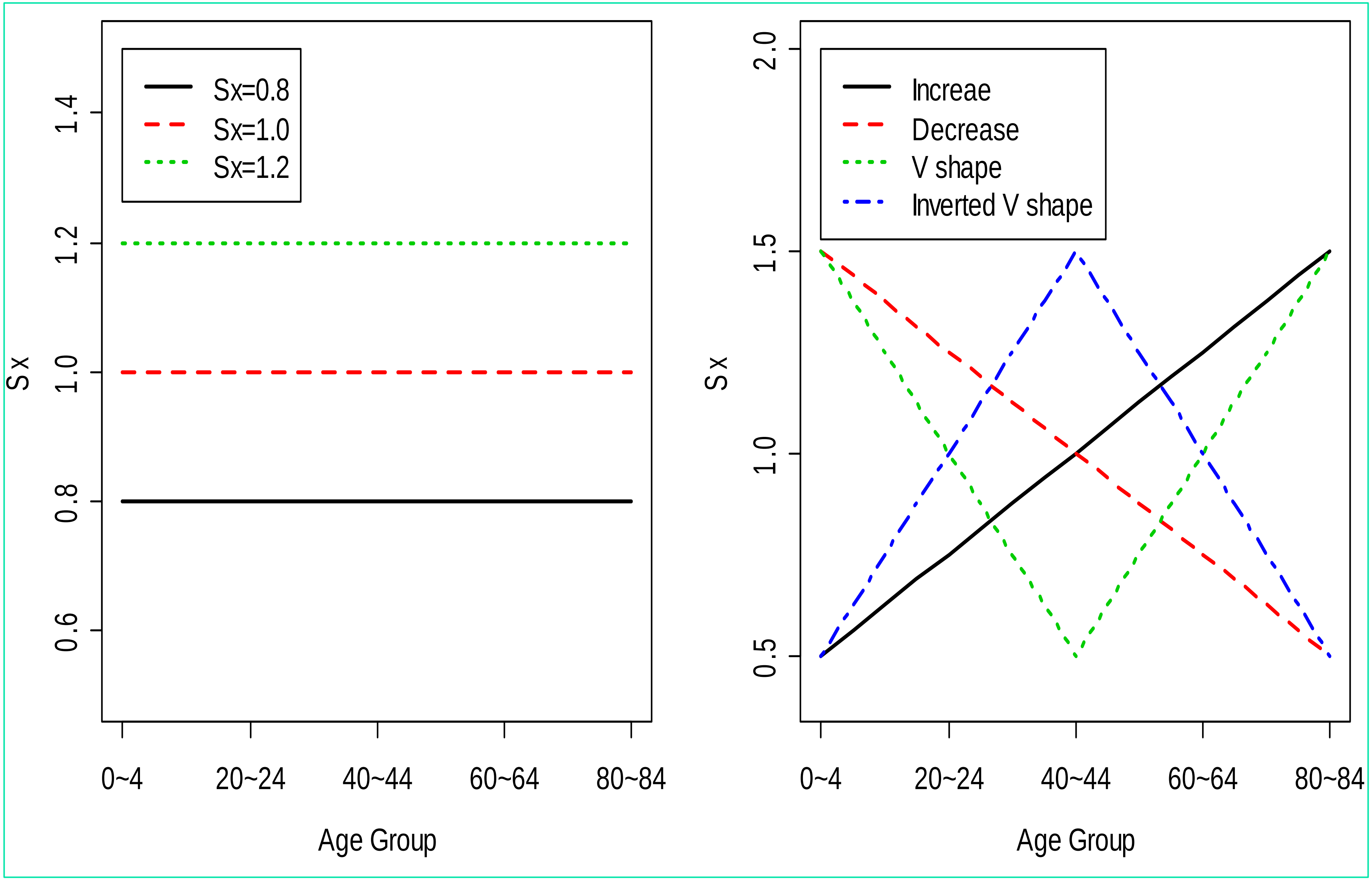
- The reference population is larger than the small population, and the mortality rates of reference population satisfy the LC model.
- The mortality rates of small population follow one of 7 mortality scenarios:

→Similar to the reference group (3 cases)

→Differ to the reference group (4 cases)

Note: We use mortality ratio $s_x = \frac{q_x}{q_x^*}$ to measure.

Seven Mortality Scenario



Simulation Setting (cont.)

- Taiwan is the reference population and counties in Taiwan are the small populations.

→5-age group (0-4, 5-9, …, 80-84)

→Training vs. Testing Periods

- Comparison criterion:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100\%$$

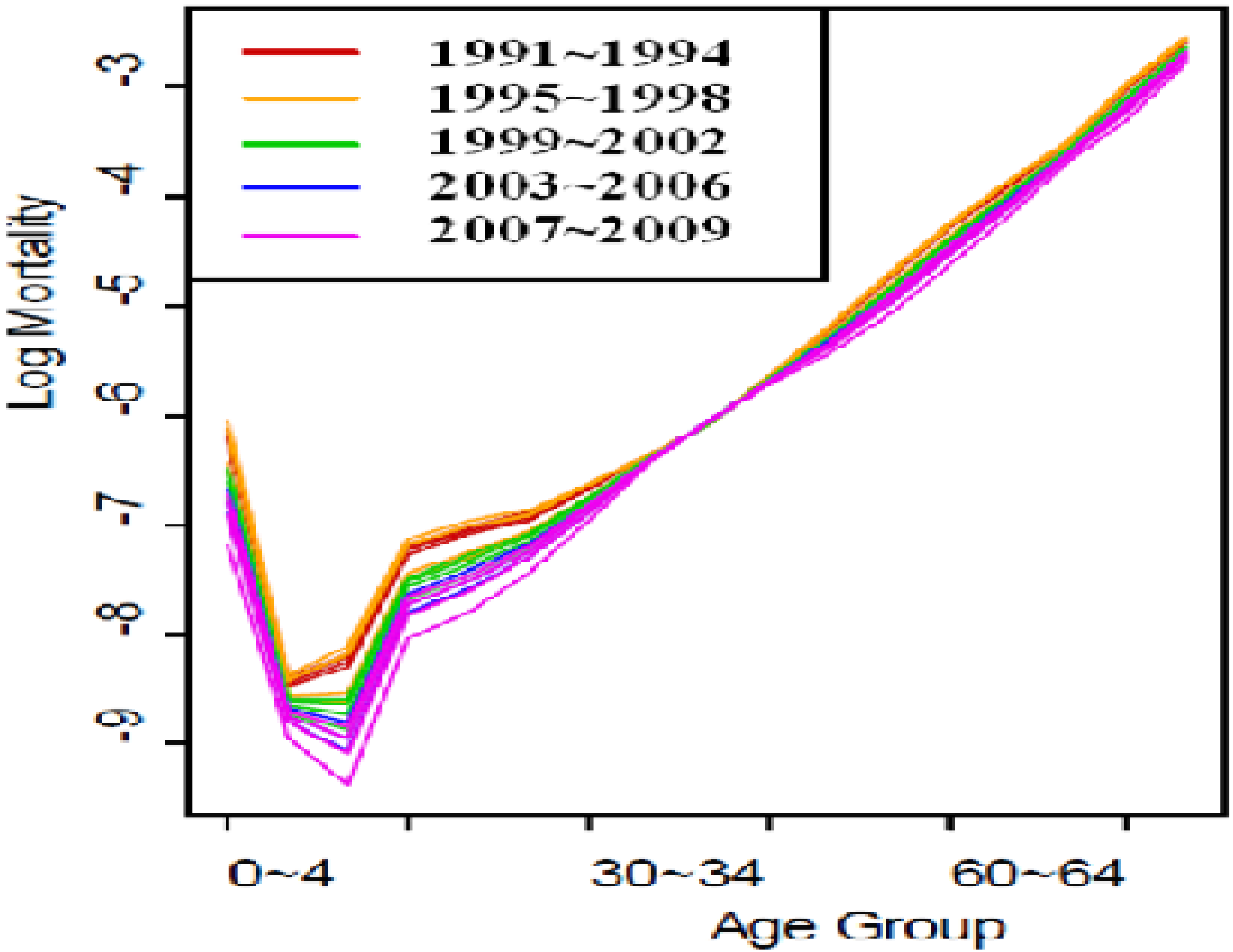
Simulation Study 1

- The first part of simulation is to check if the graduation methods can reduce the variance.
- The population of a sub-county in Taipei city (Xin-Yi Township), around 0.22 million, is treated as the target small population.

→ Taipei city (2.6 million) is the reference and its mortality rates are assumed to follow the Lee-Carter model.

→1,000 simulation runs (1991~2009)

Lee-Carter Mortality Rates (1991~2009)



Simulation Errors of Graduating Sub-county Data

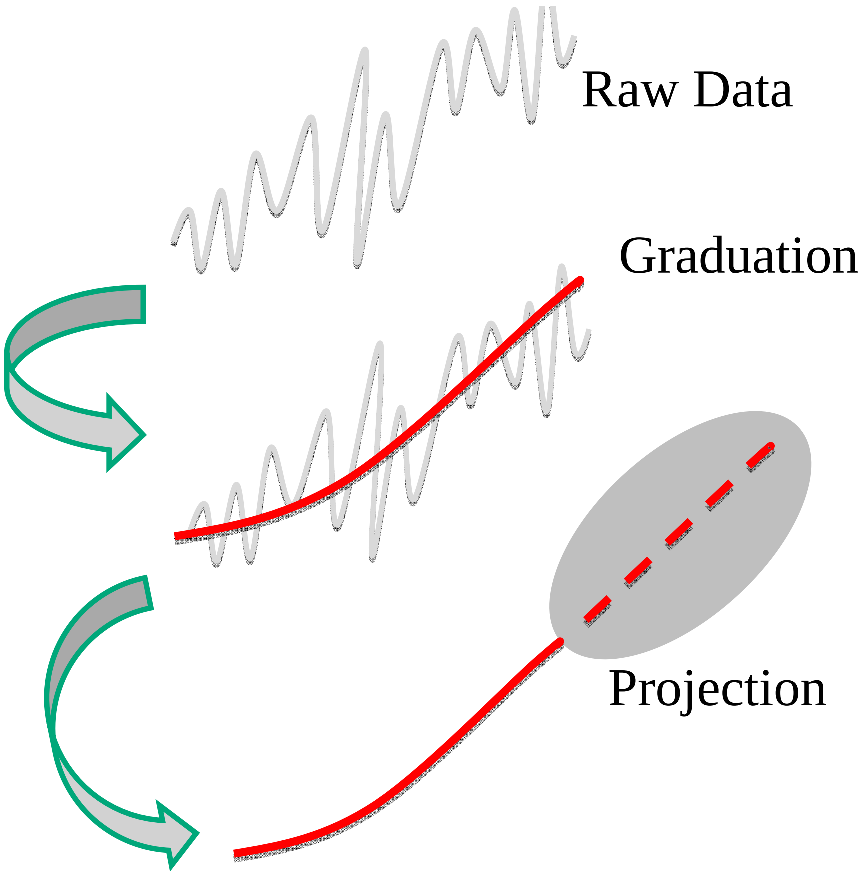
MAPE (%)

	80%	100%	120%	Incr.	Decr.	V	Λ
Raw	40.8	36.7	33.1	42.5	33.4	35.2	39.6
Whittaker	31.7	29.6	27.9	32.9	28.4	31.7	29.8
✓ Partial SMR	4.3	3.8	3.6	46.7	25.6	25.5	18.5
✓ Whittaker Ratio	12.7	12.7	12.0	32.8	22.3	13.5	16.0

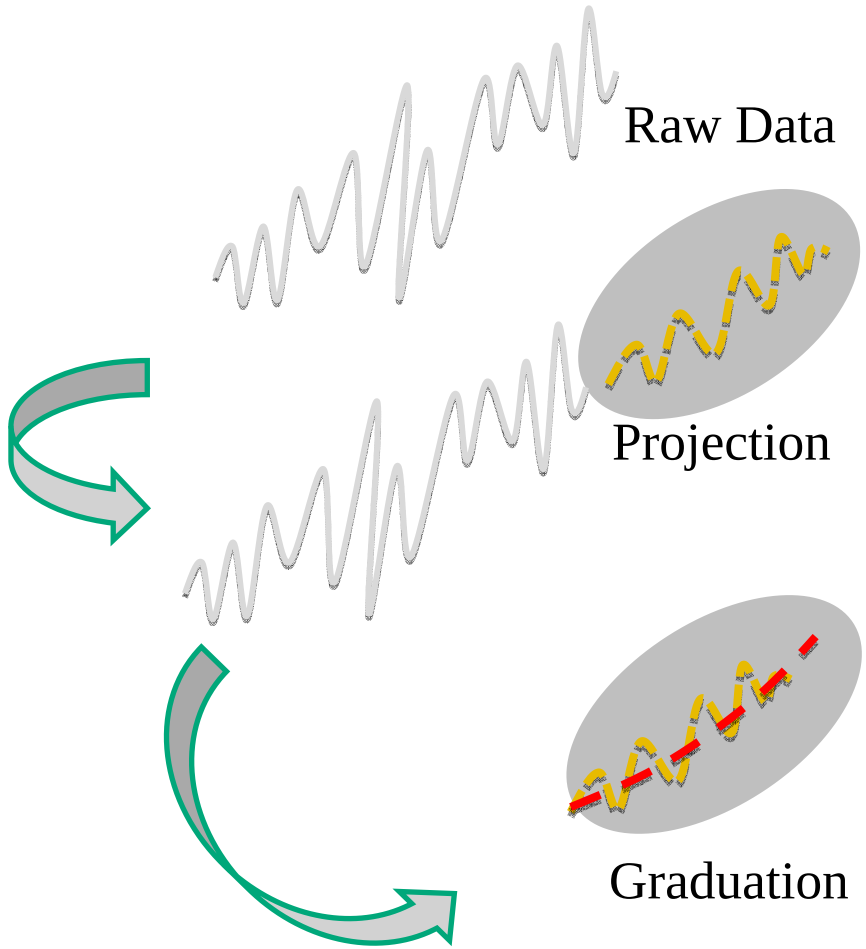
The Order of Projection (which first?)

Graduation →Projection

- Easier to handle



Projection →Graduation



Simulation Study 2

- The second part of simulation is to check if the graduation or projection shall go first.

→Partial SMR & Whittaker Ratio

→Suppose the Lee-Carter model is true

- Data: Training (1990~2004) & Testing (2005~2009; Block Bootstrap). Taipei city (2.6 million) is the reference and Xin-Yi Township (0.22 million) is the target .

→ 7 s_x scenarios are considered

Simulation Errors of Projection Order

MAPE (%)

	80%	100%	120%	Incr.	Decr.	V	Λ
W/O Graduation	42.4	41.1	36.6	39.4	36.3	39.4	36.2
Partial SMR G→P	16.9	16.8	16.3	37.5	30.0	27.5	22.9
Whittaker Ratio G→P	24.2	23.2	21.9	28.8	27.7	23.6	24.4
Partial SMR P→G	16.1	15.7	15.1	36.0	30.1	28.1	23.6
Whittaker Ratio P→G	21.1	20.2	20.1	24.6	28.5	22.7	23.3

Note: the projection order doesn't have big influence.

Simulation Study 3

- The third part of simulation is to evaluate the influence of “Block size,” “Population size,” “Base year,” and “Projection horizon”.

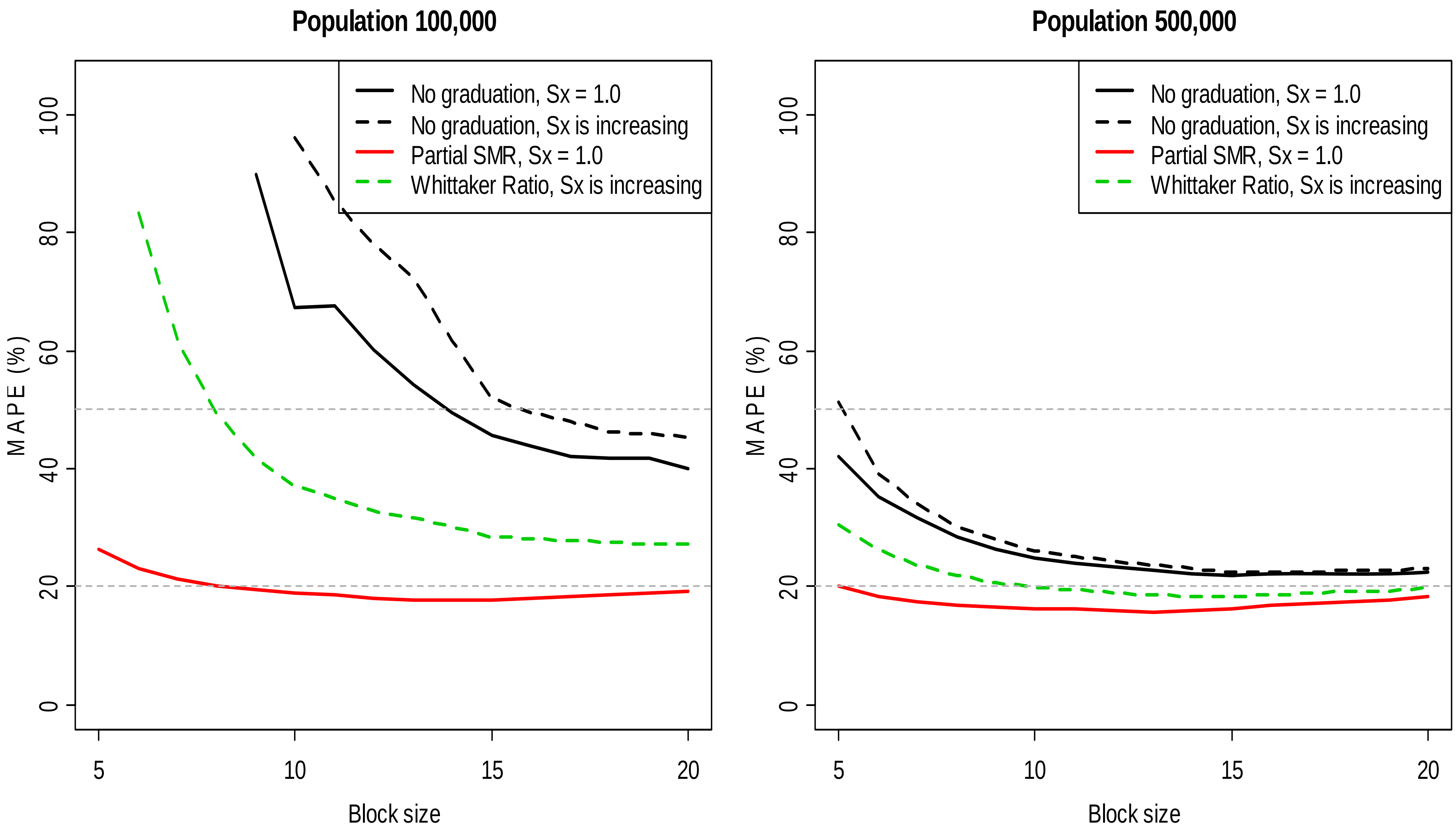
→Partial SMR & Whittaker Ratio

- Data: Training (1950~1979) & Testing (1980~2009; Block Bootstrap). Taiwan population (23 million) is the reference.

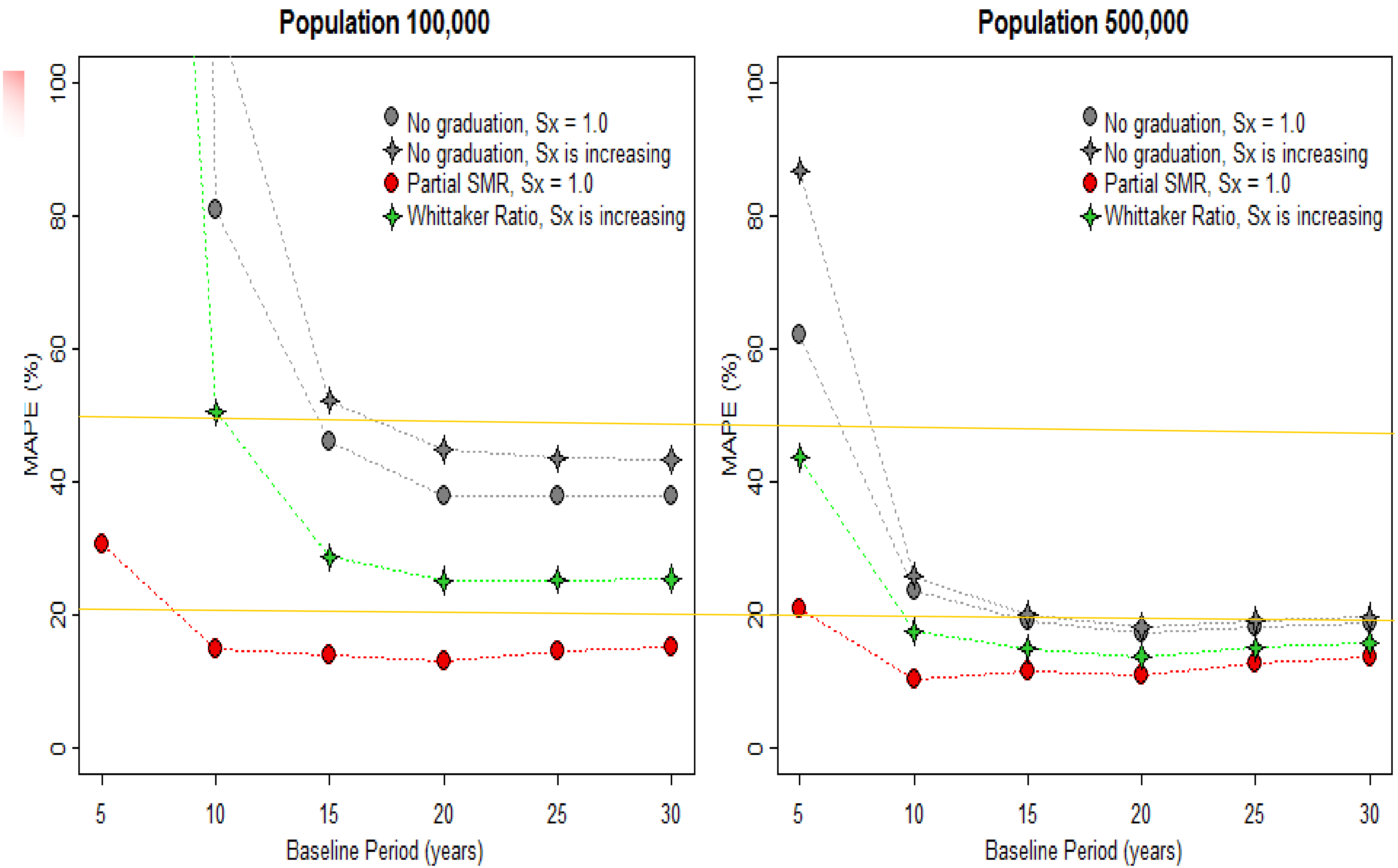
→ $S_x = 1$ & Increasing S_x is considered

→Several sizes of small population

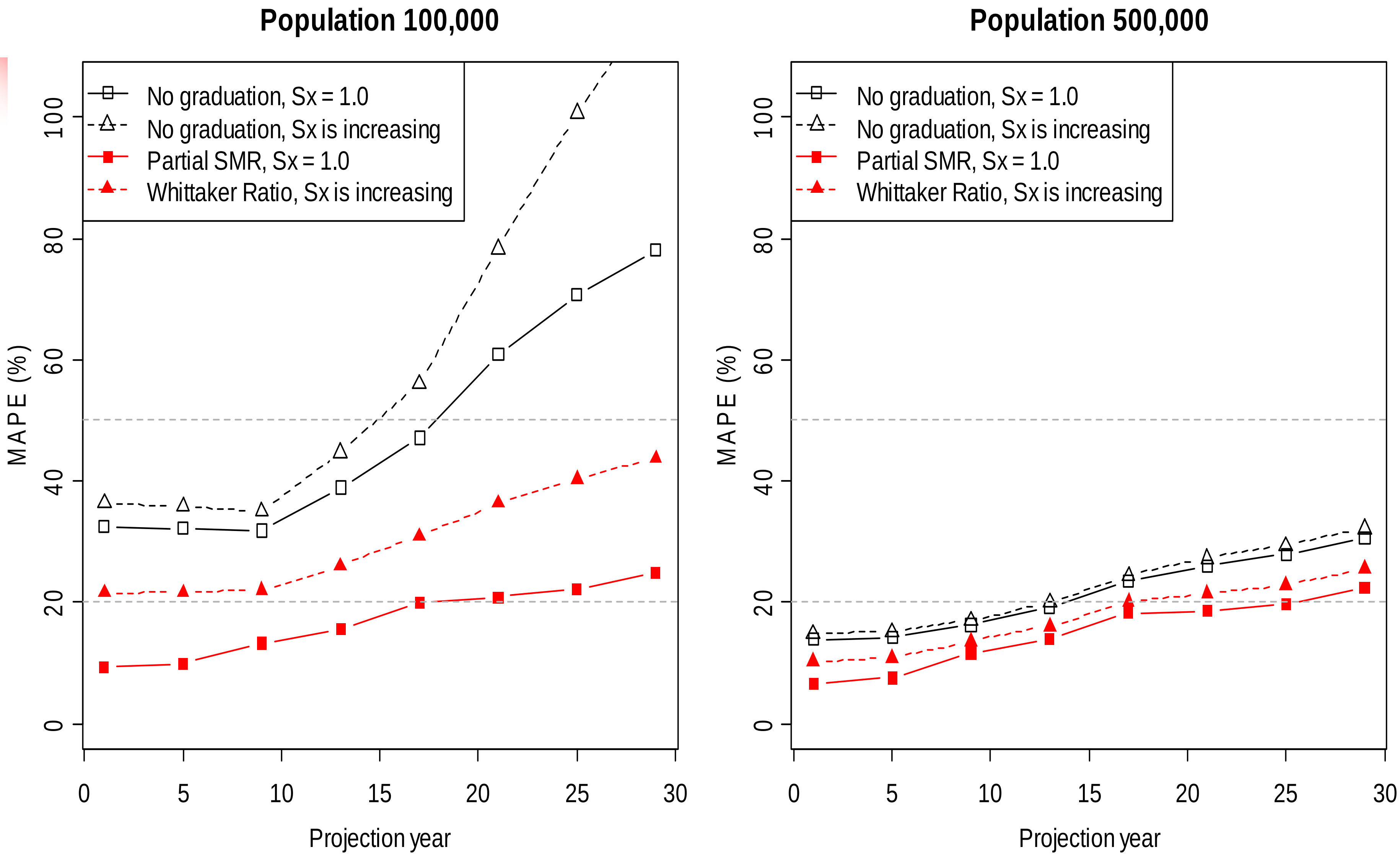
Influence of Block Bootstrap Size



Influence of Length of Base Years



Influence of Projection Horizon



Empirical Study

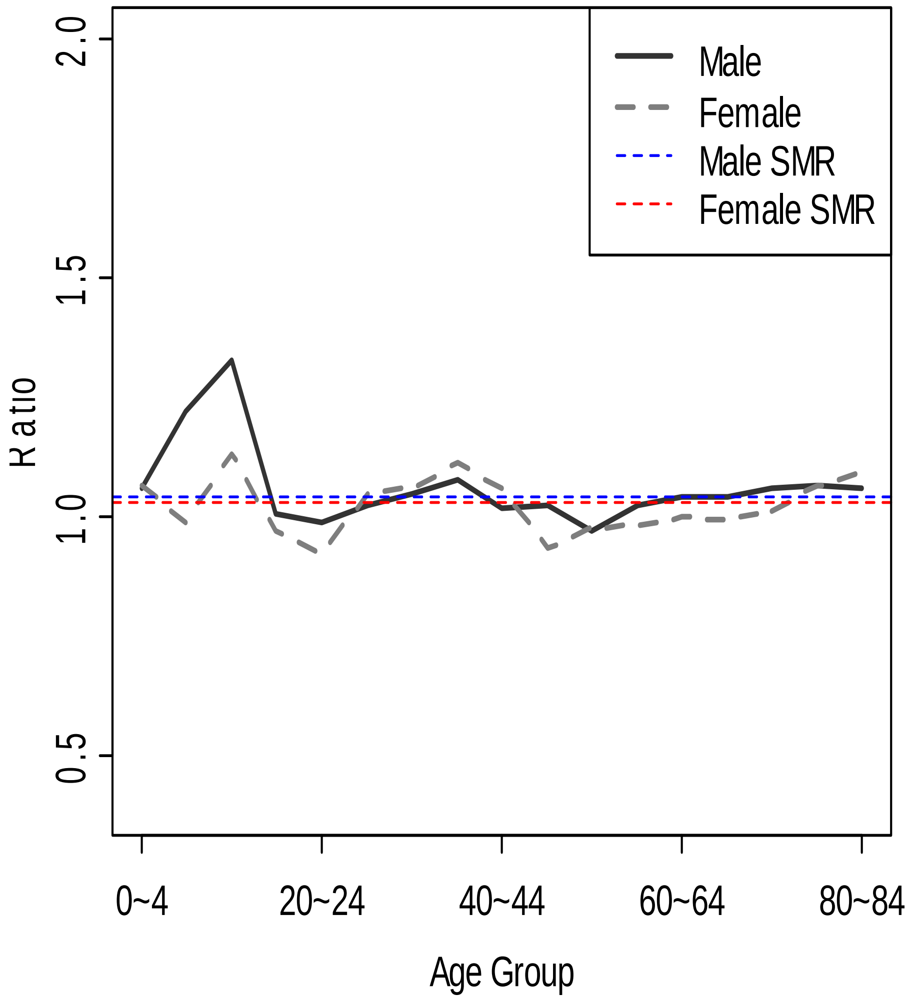
- We also conduct the population projection of 2 small counties in Taiwan (0.20 & 1.0 million). Taiwan population (23 million) is the reference.

→Training period: 1980~1994;

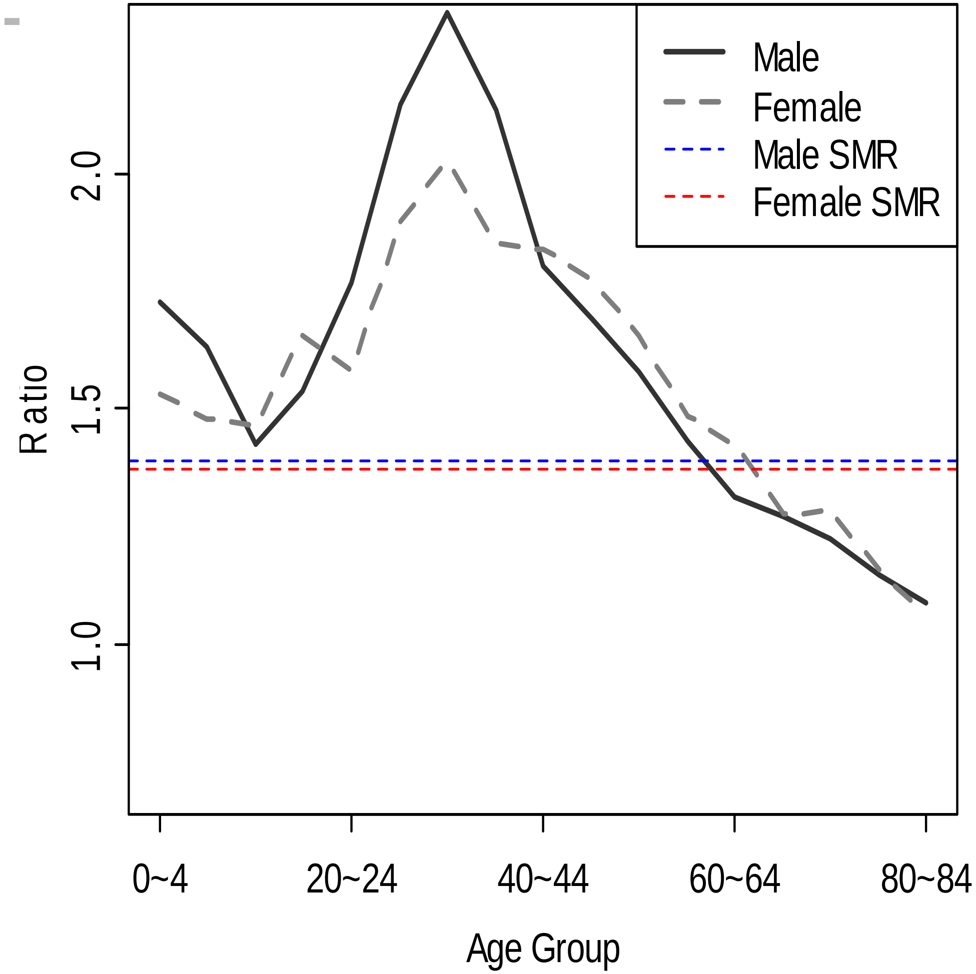
Testing period: 1995~2009

→5-age group (0-4, 5-9, …, 80-84)

Tainan County (1.0 million)



Taitung County (0.2 million)



Prediction Errors of Empirical Simulation

MAPE (%)

	Tainan County		Taitung County	
	Male	Female	Male	Female
W/O Graduation	43.5	43.9	70.0	102.4
Partial SMR	26.7	27.2	45.6	48.4
Whittaker Ratio	27.8	36.0	44.0	57.8

Conclusion

- More needs for small area population projection and the population size is a key factor.
- We propose using graduation to reduce the projection errors for small areas.

→The methods of mortality ratio & partial SMR can reduce the variance of prediction.

→The degree of influence on projection:

Population Size > Base Years > Projection Horizon

→Base years ≥ 15 & Projection Horizon ≤ 20.

Discussions

- About small area graduation:

→Include age-period-cohort (APC) relationship between small & reference areas in Partial SMR

→2-dim or 3-dim (APC) Whittaker Ratio

→Other methods (e.g., Bayesian or parametric)

- Apply graduation on a group of small populations, say, a group of heterogeneous populations?
- Combine graduation & multi-area migration model.

Thank you

for your attention!