Modeling Mortality for Countries with Small Populations

Hong-Chih Huang
Jack C. Yue
Hui-Ting Wang
National Chengchi University
Sept. 9, 2011

Summary

- Motivation
- Mortality Models for Small Population
- Proposed Approach
- Empirical Study
- Discussions

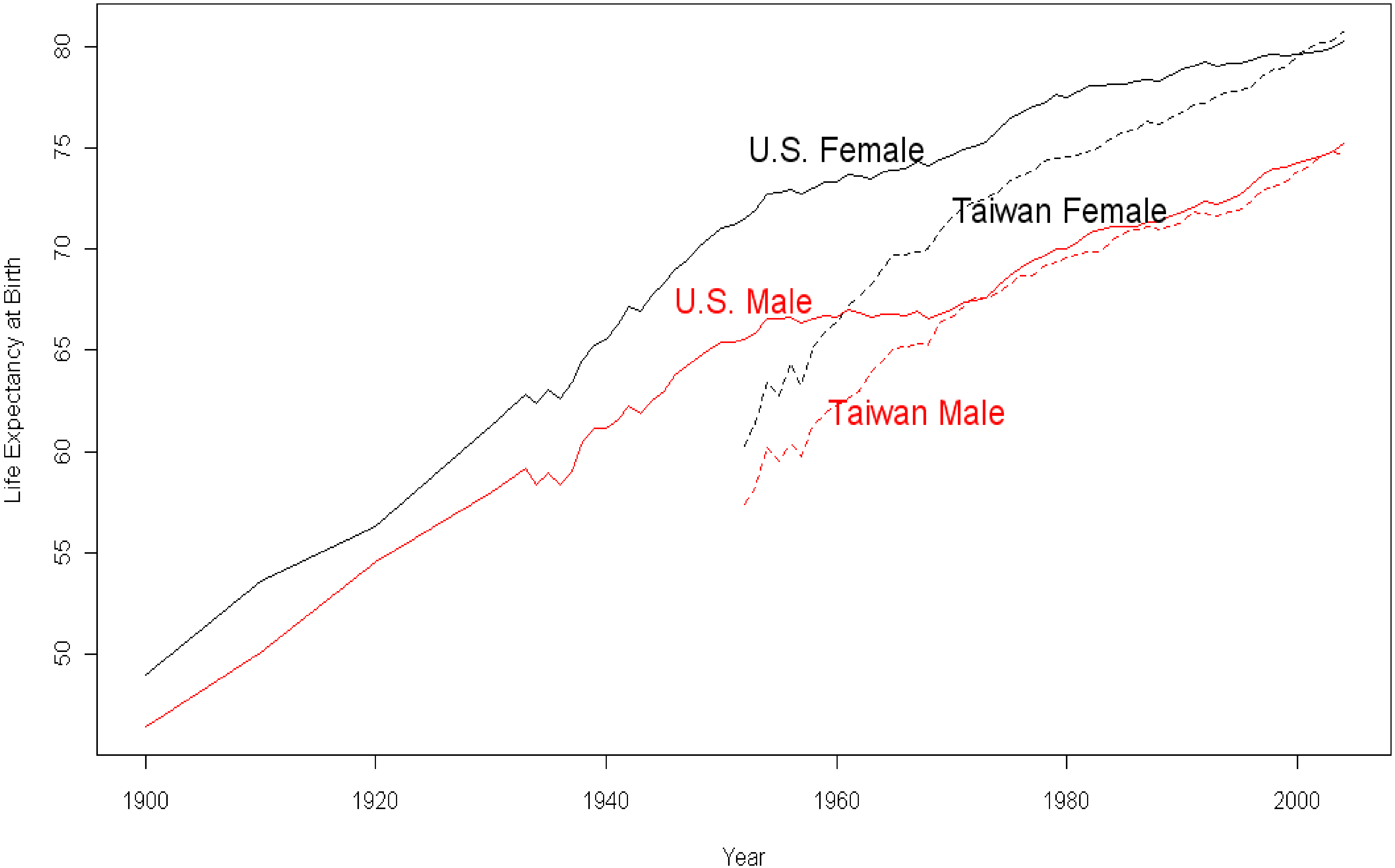
Prolonging Life Expectancy

- The average life-span of western countries has being increased significantly since turning to the 20th century.

→Mortality improvements in Asia are much faster and greater, and the life expectancies of many Asian countries are about the same as those of Western countries.

→ Population size & Rapid improvement contribute to the difficulty of mortality study.

Life Expectancies of Taiwan and the U.S.



Description of Taiwan Population

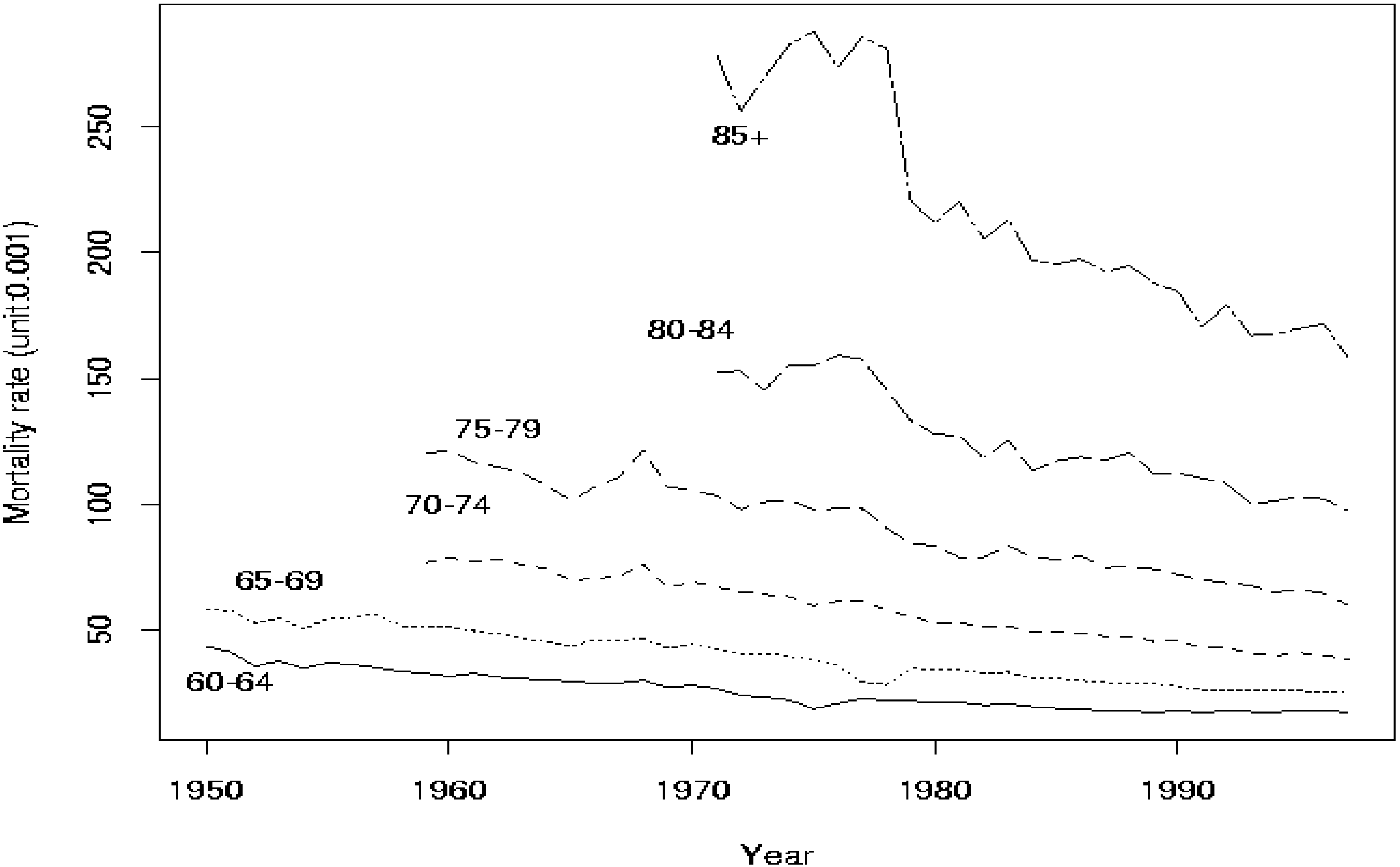
- Taiwan's population is about 23 million and the elderly data (especially the oldest-old) are not available for early years.

→Data for ages 85 & beyond are fewer than 40 years.

→The population for ages 100 & beyond is less than 2,000.

- It is not easy to model the elderly mortality.

Highest Recorded Age in Taiwan



Small Population & Mortality Model

- The most obvious way for reducing variance is to increase the sample size, or to use reference groups to modify (i.e., smooth) estimates.

→Accumulating 3 to 5 years of data

→Graduation or life-table construction techniques

→Coherent model (e.g., Li & Lee, 2005)

→Bayesian approach

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.

Coherent Mortality Model

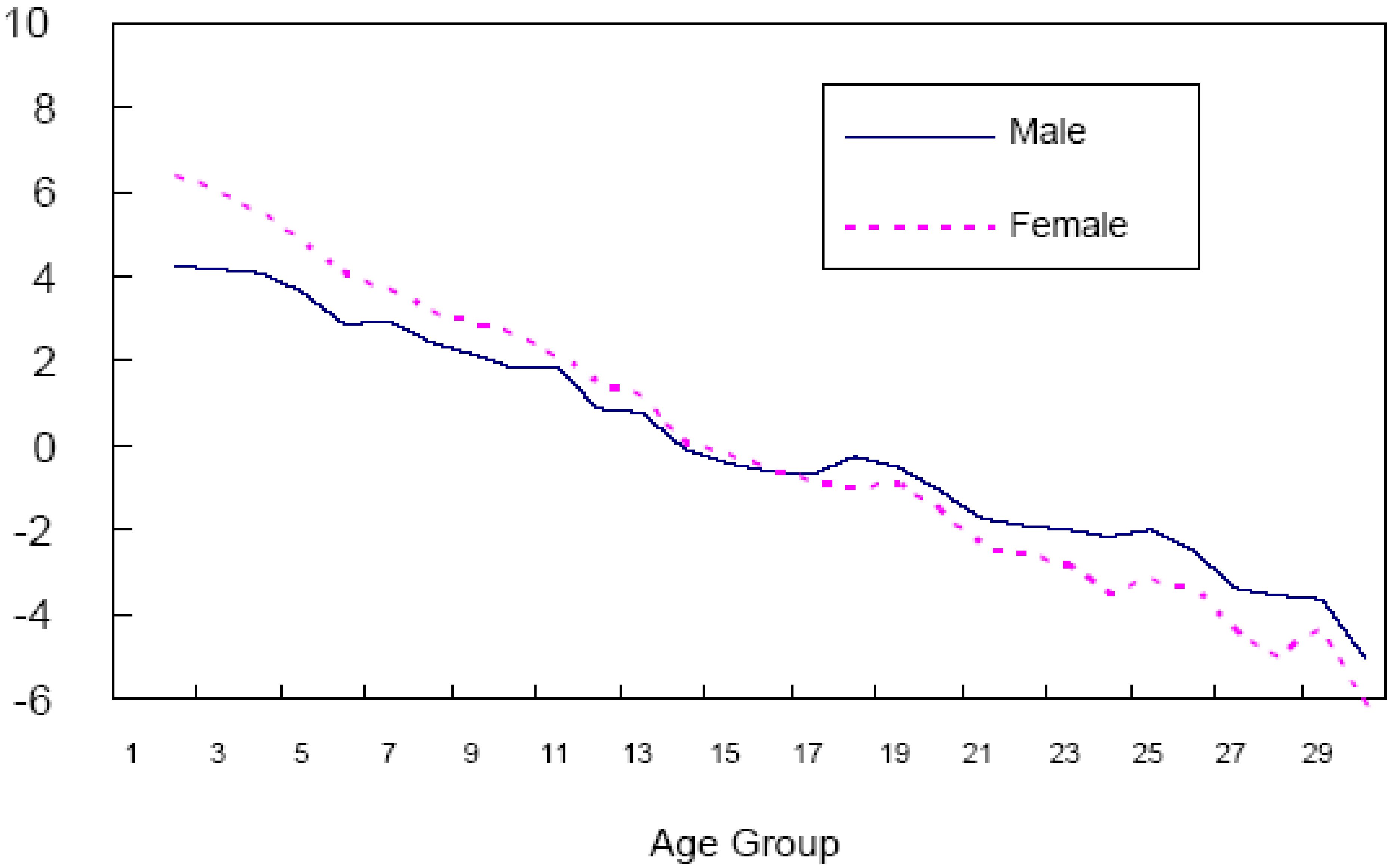
- Coherent approach is to use a (larger) reference group to refine parameter estimation.

→Li and Lee (2005) assume that the small population have same LC parameters (e.g.,κ_t) as the large population.

→Still, the small population has its parameters with smaller variation.

e.g., κ_t →Random walk; κ_t^* →AR(1)

κ_t Estimates of Lee-Carter Model in Taiwan



Proposed Approach

- We propose a method for increasing sample size, by combining populations with similar attributes as the small population.

→Using Cluster Analysis to choose countries with similar attributes

→Using Principal Component Analysis to select important variables

Note: Taiwan is the target (small) population.

Cluster Analysis (CA)

- First, define the mortality improvement

$$r_x(t) = \frac{q_x(t+1)}{q_x(t)}$$

where $q_x(t)$ is the mortality rate of age x at time t.

→Compute the differences of $r_x(t)'s$ between Taiwan and varies countries. If the differences of a country form only one cluster, it will be treated as in the same group as Taiwan.

Criteria for Number of Clusters

- Ward's minimum variance criterion is used.

→Distance within a cluster is the variance of all individuals to cluster center.

→Distance between clusters are the Euclidean distance between two cluster centers.

- We shall maximize the average distance between clusters and minimize the average distance within clusters.

Criteria for Cluster Number (Cont.)

- **AIC** (Akaike Information Criterion) or **BIC** (Bayesian Information Criterion) to decide the optimal number of clusters.
- In addition, we can compute the ratio of the average distance for k clusters to that for $k+1$ clusters. (4^{th} column: ratio of distance)

→A larger value indicates that the average distance between clusters is larger.

Table 2-1. Optimal Number of Clusters for Mortality Similarity between Taiwan & U.S.

# of Cluster	AIC	BIC	Ratio of distance measure
1	2131.692	2283.369	---
2	1925.121	**2228.475**	1.580
3	1822.094	2277.125	**1.750**
4	1774.748	2381.456	1.070
5	1873.571	2631.956	1.303
6	**1529.755**	2439.817	1.089
7	1627.571	2689.310	1.229
8	1722.177	2935.593	1.082
9	1790.095	3155.188	1.073
10	1895.513	3412.282	.939

Note: U.S. and Taiwan are not similar in mortality.

Principal Component Analysis (PCA)

- PCA is a popular tool for data reduction and it is especially useful in our case since each country has a lot of $q_x(t)$ values. Traditionally, usually 1, 2, or 3 PC's are chosen.

→Lee-Carter model: 1 PC

→Yang et al. (2010): 2 PC's

→Heligman and Pollard (1980): 3 PC's

Note: More PC's are expected if a lot of countries are chosen from the CA.

Empirical Study

- Consider the data of 30 countries from Human Mortality Database (HMD) and select countries share similar mortality attributes as Taiwan.

→Lee-Carter model is treated as the benchmark.

→Mean absolute percentage error (MAPE) is the criterion for model performance,

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|Y_i - \hat{Y}_i\right|}{Y_i}\times 100\%$$

Step 1. Choose Countries with similar attributes

Table 3-1. Cluster analysis of mortality improvement for single age (1970-1999)

Country	Cluster #	Country	Cluster #	Country	Cluster #
Australia	2	France	**1**	Poland	1
Austria	**1**	Hungary	2	Portugal	2
Belarus	3	Ireland	**1**	Russia	2
Belgium	2	Italy	2	Slovakia	2
Bulgaria	2	Japan	2	Spain	1
Canada	2	Latvia	2	Sweden	3
Czech Republic	2	Lithuania	3	Switzerland	3
Denmark	2	Netherlands	2	UK	2
Estonia	2	New Zealand	2	USA	3
Finland	2	Norway	**1**	Ukraine	2

Table 3-2. Data Periods for six selected countries

Countries	**Fitting period**	**Prediction period**
Austria	1970~1999	2000~2008
France	1970~1999	2000~2007
Ireland	1970~1999	2000~2006
Norway	1970~1999	2000~2007
Spain	1970~1999	2000~2006
Taiwan	**1970~1999**	**2000~2008**

Step 2. Select PC's from the chosen countries

Table 3-3. Results of PCA for Taiwan male and female

	Male		Female	
PC#	**Eigen-value**	**Cumulative Variation (%)**	**Eigen-value**	**Cumulative Variation (%)**
1	11.399	56.996	14.324	71.619
2	4.681	80.402	2.730	85.271
3	1.323	87.016	.889	89.714
4	.632	90.177	.628	92.855
5	.477	92.564	.341	94.559
6	.417	94.651	.224	95.679
7	.287	96.085	.199	96.673
8	.184	97.004	.155	97.445
9	.118	97.597	.112	98.003
10	.108	98.137	.097	98.489

Cross-Validation

- The data are divided into two parts:

→In-sample (fitting period or training data, 1970-1999)

→Out-sample (testing data, 2000-2008)

- The data are mortality rates of 5-age groups, ranging from 0 to 99 year-old.

Note: Although AIC & BIC can be used, we suggest using the MAPE for comparison.

Training Errors (Male)

Table 3-4. Fitting MAPE of the male for six countries (unit：%)

Country	# of Principal Component							LC
	1	2	3	4	5	6	7	
Austria	8.91	8.95	5.73	5.38	4.43	4.00	3.56	4.38
France	9.71	6.89	4.61	4.23	3.72	2.76	2.66	3.81
Ireland	17.20	8.39	8.23	6.27	6.00	5.10	4.48	6.51
Norway	9.62	7.25	6.65	5.83	5.34	4.56	4.43	5.52
Spain	9.62	7.60	6.51	5.82	4.57	4.07	3.117	6.41
Taiwan	14.11	8.25	7.51	6.77	5.89	4.30	3.41	5.75
Total	**11.53**	**7.89**	**6.54**	**5.72**	**4.99**	**4.13**	**3.62**	**5.40**

Training Errors (Female)

Table 3-5. Fitting MAPE of the female for six countries (unit：%)

Country	# of Principal Component							LC
	1	2	3	4	5	6	7	
Austria	8.76	7.95	5.95	4.88	4.53	4.36	3.71	4.75
France	11.68	5.75	5.07	3.65	2.88	2.15	2.03	2.85
Ireland	16.05	9.12	7.50	7.20	6.12	5.43	4.85	7.06
Norway	10.27	8.11	7.39	6.41	5.96	5.48	5.04	6.78
Spain	9.19	7.93	7.02	4.12	3.59	2.81	2.54	4.75
Taiwan	9.47	7.52	5.78	5.34	4.14	3.74	3.33	4.46
Total	**10.90**	**7.73**	**6.46**	**5.27**	**4.54**	**4.00**	**3.58**	**5.11**

Testing Errors (Male)

Table 3-7. MAPE values of forecasting results for males of six countries (unit：%)

	# of Principal Component							LC
Country	**1**	**2**	**3**	**4**	**5**	**6**	**7**	
Austria	10.48	9.99	9.58	8.94	7.62	7.81	8.85	9.55
France	14.27	13.58	12.76	10.31	9.29	8.15	8.29	11.61
Ireland	23.25	21.82	21.12	19.02	20.09	21.22	21.29	22.72
Norway	13.69	14.08	14.00	13.68	13.71	13.49	12.80	13.38
Spain	10.57	9.79	9.77	8.62	8.75	8.66	9.13	14.04
Taiwan	22.26	15.10	15.70	15.13	14.75	14.90	14.86	18.13
Total	**15.73**	**13.93**	**13.71**	**12.54**	**12.23**	**12.22**	**12.40**	**14.72**

Testing Errors (Female)

Table 3-8. MAPE values of forecasting results for females of six countries (unit：%)

Country	# of Principal Component							LC
	1	2	3	4	5	6	7	
Austria	13.03	11.77	11.88	10.33	10.12	9.74	9.67	10.88
France	13.67	14.34	13.34	10.63	8.26	7.54	7.64	10.03
Ireland	16.06	14.34	14.35	14.74	13.93	14.00	13.70	15.50
Norway	14.95	14.97	15.59	13.86	12.96	12.93	13.70	14.58
Spain	7.07	6.58	5.48	5.38	5.68	5.90	5.91	9.62
Taiwan	12.59	14.52	12.48	14.90	14.78	14.76	15.36	15.37
Total	**12.95**	**12.86**	**12.28**	**11.74**	**11.07**	**10.91**	**11.11**	**12.69**

Conclusion

- Modeling mortality rates for small populations become more popular and several models were proposed.
- We propose Cluster Analysis to increase the sample sizes for small populations and Principal Component Analysis for data reduction and mortality prediction.

→The proposed method has smaller MAPE than the Lee-Carter model for Taiwan data.

Discussions

- Reference groups for A & B can be different, even A & B share similar mortality attributes.
- Mortality improvements are not homogeneous at all ages for the target and reference populations.
- Other data reduction methods (Functional PCA)
- Other methods for reducing the variance

→Combining the idea of coherent model and cluster analysis

→Graduation

Thank you for your attention!