修勻學(Graduation) — 高齡死亡率修勻

授課教師: 余清祥教授

課程日期:2023年11月29日

資料下載:

http://csyue.nccu.edu.tw



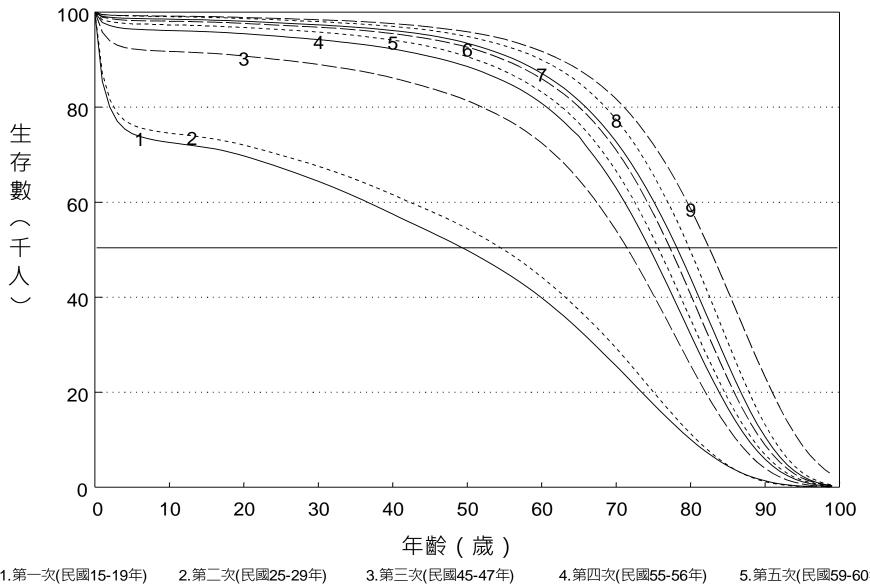
關於高齡死亡率的修勻

- ■因為高齡人口較少,傳統作法多為仰賴關係模型(Relational Model),例如:Gompertz及高馬氏(Gompertz-Makeham)模型、Coale-Kisker模型、羅吉斯模型等。
- →也有人使用隨機模型(Stochastic Model),例如: Cairns-Blake-Dowd (CBD)、Booth-Maindonald-Smith (BMS)、Hyndman-Ullah (HU)模型。

各國簡易生命表的高齡編算方法比較

小區域方法	國家	特色
Brass Logit 模型	紐西蘭、澳洲、美國 拉丁美洲裔	解決死亡數為 0 或不合理 的波動現象
Greville 法	加拿大	另案處理 90 歲以上高齡 死亡率,死亡率及平均餘 命均計算其邊際誤差
Gompertz 假設	臺灣	死力的對數值隨年齡直線 上升,可藉由迴歸合併高 齡人數增加模型穩定度





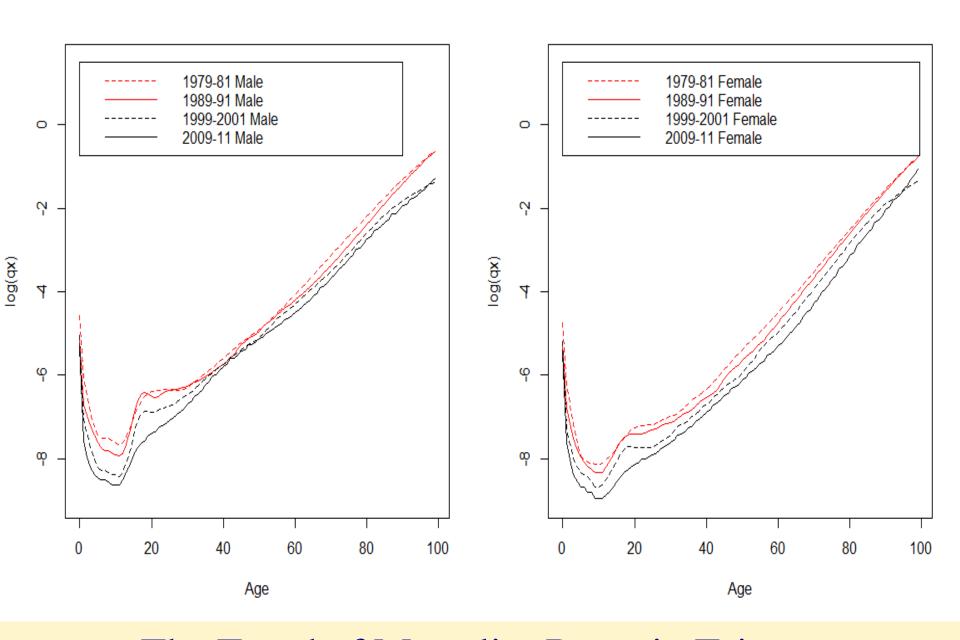
1.第一次(民國15-19年) 6.第六次(民國64-65年)

7.第七次(民國69-70年)

3.第三次(民國45-47年) 8.第八次(民國78-80年)

4.第四次(民國55-56年) 9.第九次(民國88-90年)

5.第五次(民國59-60年)



The Trend of Mortality Rates in Taiwan

死亡率變化對於出生時平均壽命差異之貢獻(日本女性)

年份	1951~1	.960至	1961~1	970至	1971~1	1980至	1981~1	990至
年齡	1961	~1970	1971~	1980	1981	~1990	1991~	1995
0	1.502	25.6	0.619	16.3	0.279	8.5	0.086	4.9
1-4	0.911	15.5	0.149	3.9	0.085	2.6	0.020	1.1
5- 9	0.217	3.7	0.065	1.7	0.037	1.1	0.006	0.3
10-29	0.809	13.8	0.273	7.2	0.151	4.6	0.045	2.6
30 - 49	0.915	15.6	0.497	13.1	0.298	9.0	0.117	6.7
50-69	1.027	17.5	1.109	29.2	0.891	27.0	0.369	21.1
70-89	0.478	8.2	1.064	28.0	1.474	44.7	1.000	57.3
90+	0.006	0.1	0.024	0.6	0.084	2.5	0.103	5.9
總和	5.865	100%	3.800	100%	3.300	100%	1.746	100%

死亡率變化對於出生時平均壽命差異之貢獻(台灣女性)

年份	1956~	1958至	1966~	1976至	1970~	1971至	1975~	1976至	1980~1	1981至
年齡 \	1966~	-1976	1970~	-1971	1975	~1976	1980~	1981	1989~	1991
0	1.369	25.7	0.384	20.8	0.383	18.8	0.159	13.8	0.263	13.4
1-4	1.459	27.4	0.396	21.4	0.113	5.5	0.178	15.5	0.116	5.9
5-9	0.160	3.0	0.051	2.8	0.056	2.7	0.003	0.3	0.009	0.5
10-29	0.462	8.7	0.220	11.9	0.137	6.7	0.044	3.8	0.074	3.8
30-49	0.752	14.1	0.208	11.3	0.352	17.2	0.135	11.7	0.265	13.5
50-69	0.772	14.5	0.375	20.3	0.548	26.8	0.216	18.8	0.760	38.9
70-89	0.338	6.3	0.206	11.2	0.446	21.9	0.420	36.5	0.457	23.4
90+	0.012	0.2	0.006	0.3	0.004	0.2	0.002	0.2	0.012	0.6
總和	5.324	100%	1.846	100%	2.041	100%	1.150	100%	1.956	100%

死亡率變化對於出生時平均壽命差異之貢獻(瑞典女性)

年齢年份	1951~1 1961~	.960至 ~19 7 0		1970至 ~1980	1971~1 1981	.980至 ~1990	1981~1 1991~	
0	0.304	14.8	0.300	16.0	0.156	9.0	0.146	10.7
1-4	0.101	4.9	0.047	2.5	0.031	1.8	0.024	1.8
5- 9	0.015	0.7	0.036	2.0	0.036	2.1	0.009	0.7
10-29	0.091	4.4	0.040	2.1	0.097	5.6	0.066	4.8
30-49	0.208	10.1	0.085	4.5	0.162	9.3	0.144	10.6
50-69	0.667	32.5	0.428	22.8	0.395	22.7	0.259	19.0
70-89	0.645	31.4	0.899	47.9	0.828	47.7	0.655	48.1
90+	0.024	1.2	0.043	2.3	0.032	1.8	0.060	4.4
總和	2.055	100%	1.878	100%	1.737	100%	1.363	100%

資料來源:林麗芬、林作民、吳蕙菁(2003、人口學刊)

什麼是Gompertz Law?

- ■由Gompertz於1825年提出,根據19世紀 英格蘭、瑞典、法國年齡介於20至60歲 的成年人死亡資料,發現瞬間死亡率的 對數值隨年齡直線上升。
- →瞬間死亡率滿足 $\mu_x = BC^x$, B > 0, C > 1, 也可轉化成生存機率的限制式:

$$log(p_{x+1})/log(p_x) = C \circ$$

■ 根據Gompertz的瞬間死亡率 $\mu_x = BC^x$ 假設,可推得:

$$p_{x} = e^{-\int_{x}^{x+1} \mu_{t} dt} = e^{-\int_{x}^{x+1} BC^{t} dt}$$
$$= e^{-BC^{x}(C-1)/\log C}$$

因此,

$$\log p_x = -BC^x(C-1)/\log C$$

或是

$$\log(-\log p_x) = \log B + \log(C - 1) - \log(\log C) + x \log C$$
$$= \alpha + \beta x$$

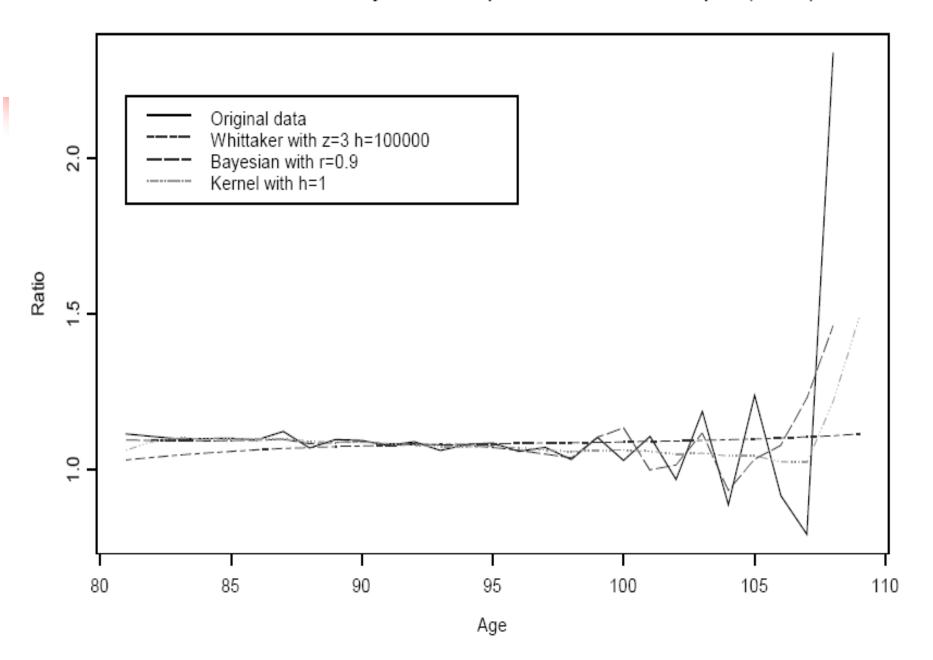
Gompertz Law的優缺點

- 優點:
- →因為屬於參數模型,參數估計相對容易, 推估未來的死亡率較為直接。
- 缺點:
- →老年資料品質不一,假設不易驗證
- →尚無統一的估計與檢定參數的方法

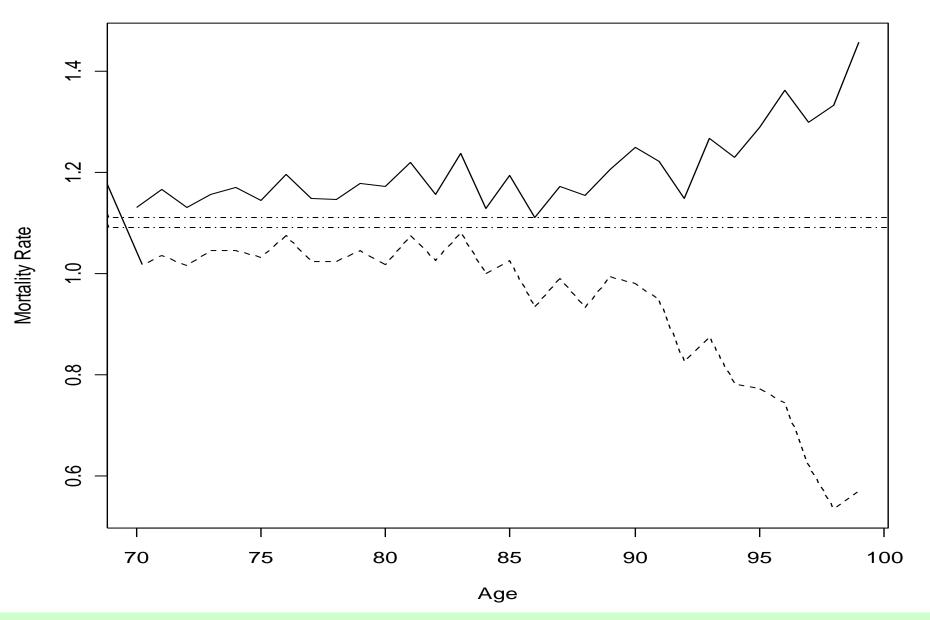
蒐集老年資料的問題

- 因為資料量不足,蒐集老年人資料可能 需要較長的時間(最高年龄組?)
- →為了累積足夠的資料,可能會混合不同母體、不同來源的資料。
- 資料量不足會使參數估計值較不穩定
- → 為了避免混合不同母體的資料而縮短蒐 集時間,也會面臨變異數過大的問題。

Is the force of mortality the Gompertz function?---Japan(male)

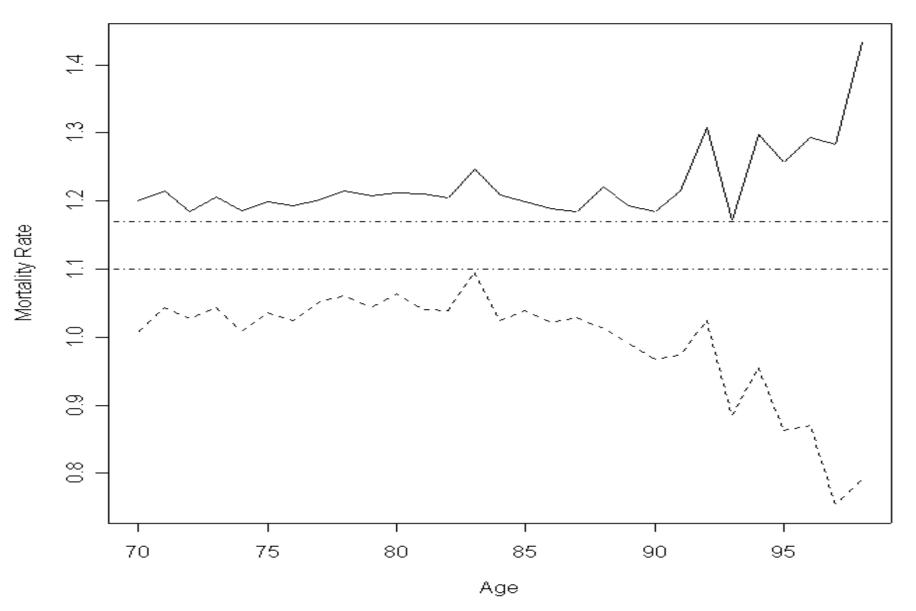


Bootstrap C.I. of Male Mortality Rates

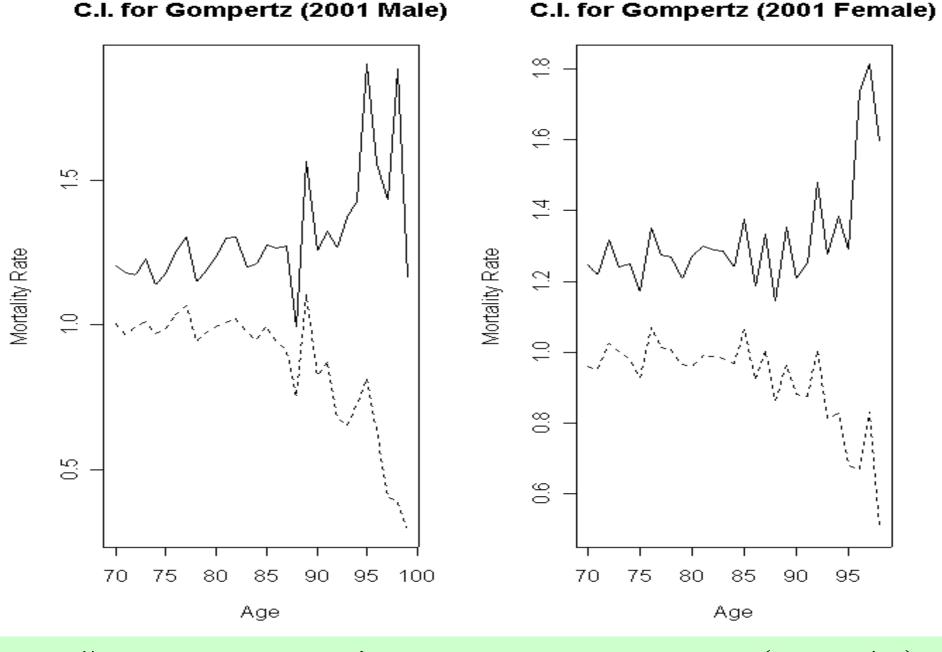


台灣男性死亡率1000次Bootstrap模擬(1999至2001年)

Bootstrap C.I. of Female Mortality Rates



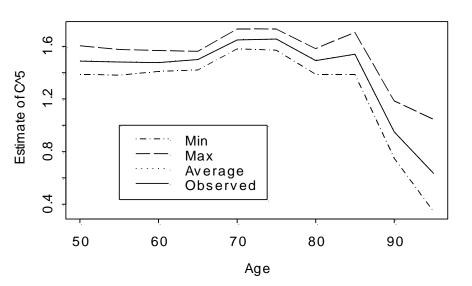
台灣女性死亡率1000次Bootstrap模擬(1999至2001年)

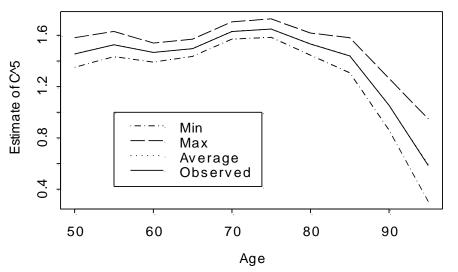


台灣男女兩性死亡率1000次Bootstrap模擬(2001年)

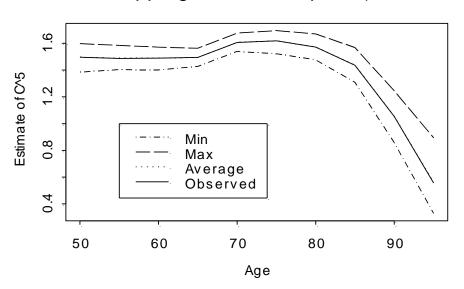


Bootstrapping C.I. for Gompertz (1999 Male)



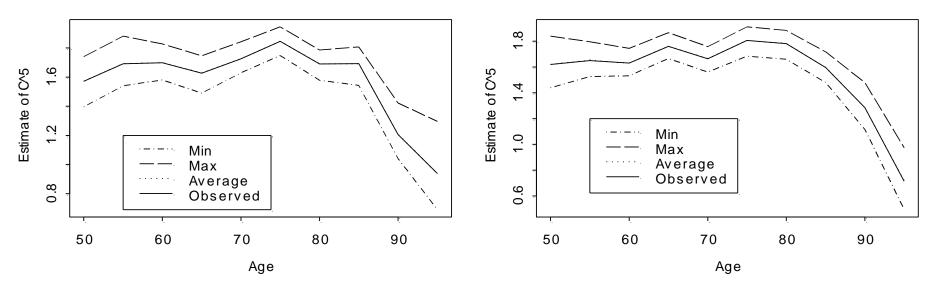


Bootstrapping C.I. for Gompertz (2000 Male)

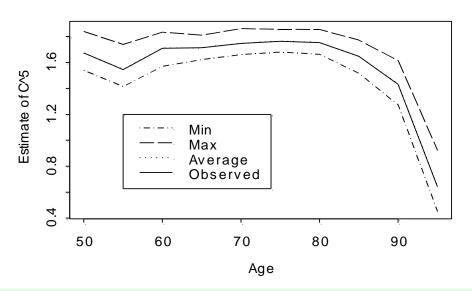


1998至2000年男性高齡死亡率的模擬結果

Bootstrapping C.I. for Gompertz (1998 Female) Bootstrapping C.I. for Gompertz (1999 Female)



Bootstrapping C.I. for Gompertz (2000 Female)



1998至2000年女性高齡死亡率的模擬結果

三種Gompertz參數估計方法

1. Weighted least square (WLS)

$$\min_{\alpha,\beta} \sum_{x} w_{x} (\log(-\log p_{x}) - \alpha - \beta x)^{2}$$

2. Nonlinear Maximization (NM)

$$\min_{B,C} \sum_{x} w_{x} \left(p_{x} - e^{-BC^{x}(C-1)/\log C} \right)^{2}$$

3. Maximum Likelihood Estimation (MLE)

$$\min_{B,C} \sum_{x} [(n_{x} - d_{x})BC^{x}(C-1)/\log C - d_{x} \log(1 - e^{-BC^{x}(C-1)/\log C})]^{2}$$

- Example 1. Given x > 0, find the minimum of f(x) = x + 1/x.
- \rightarrow In R, we use the function "nlminb" to find the minimum of f(x).

```
f = function(x) \{ x+1/x \}

nlminb(start=2,obj=f,lower=0)
```

The minimum 2 is attained at x = 1, with 7 iterations. (Larger starting values, more iterations?)

Starting	2	5	8	10	20	50	100
Iterations	7	9	11	10	10	13	16

Note: "nlm" and "optim" can also be used.

- Example 1. (continued) Given x, y > 0, find the minimum of f(xy) = xy + 1/xy.
- → Similarly, we use the command "nlminb."

```
f = function(x) \{ x[1] * x[2] + 1/(x[1] * x[2]) \}

nlminb(start = c(2,2), obj = f, lower = c(0,0))
```

Note: The minimum 2 is attained at xy = 1 and there are infinitely many solutions. Thus, unlike the 1-dim case, the number of iterations does not necessarily increase with larger initial values.

Example 2. We use Japanese and Swedish elderly data (both sets are age-specific data) in <u>Human Mortality Database</u> to check the Gompertz assumption, use the weighted least squares to find the estimates of parameters.

Note:
$$\mu_{x} = BC^{x} \& p_{x} = e^{-\int_{0}^{1} \mu_{x+t} dt}$$

$$\Rightarrow {}_{5}p_{x} = -BC^{x}(C^{5} - 1)/\log C$$

$$or \log p_{x+1}/\log p_{x} = C.$$

- Of course, we can also apply least squares to $log p_x$ directly, since $log p_{x+1} / log p_x = C$.
- →Alternatively, we can consider the optimization, via minimizing

$$\min_{B,C} \sum_{x} {}_{5}l_{x} ({}_{5}p_{x} + BC^{x}(C^{5} - 1)/\log C)^{2}.$$

Similarly, we can use Maximum likelihood to find the values of B and C as well.

• Question: Which approach is the best?

Table 3 Simulated loss of parameter C for WLS, NM, and MLE

(unit: 0.001)

		Jap	oan	Swe	eden
Method: Weights		Male	Female	Male	Female
WLS:	n_x	.000146	.000093	.000859	.000515
	$\sqrt{n_{_X}}$.000280	.000138	.001168	.000616
	$\log(n_x)$.007751	.003696	.009002	.005687
	1	.039220	.013698	.028503	.018376
NM:	n_x	.000138	.000088	.000759	.000497
	$\sqrt{n_x}$.001465	.000950	.004239	.003402
	1	.127820	.005851	.099599	.090268
MLE	1	.000126	.000074	.000729	.000416

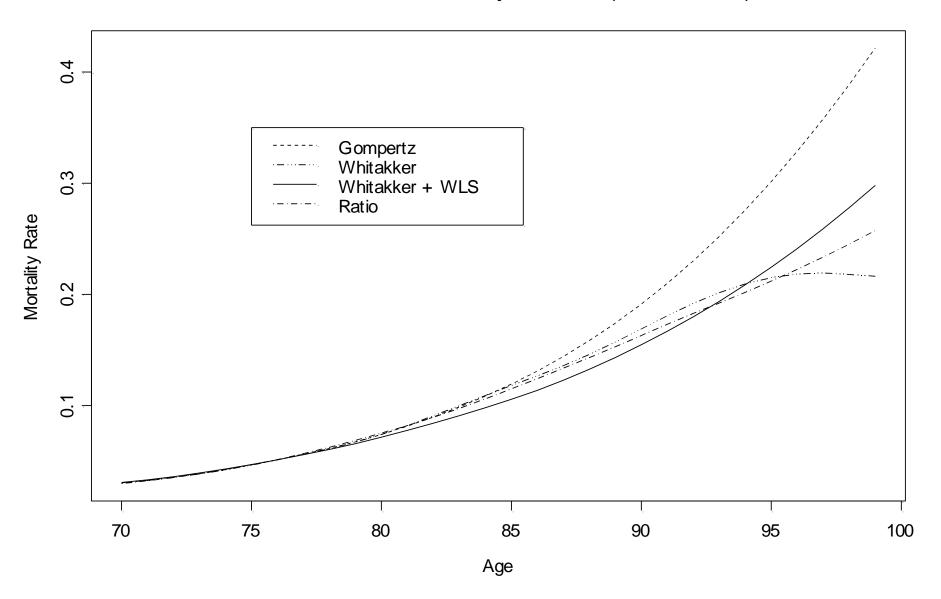
Table 4 Number of rejections among 1,000 simulations for WLS, NM, and MLE

		Ja	npan	Sw	eden
Method:	Method: Weights		Female	Male	Female
WLS:	n_{x}	0	0	0	0
	$\sqrt{n_{_X}}$	0	0	0	0
	$\log(n_x)$	480	371	114	76
	1	790	689	392	349
NM:	n_{x}	0	0	0	0
	$\sqrt{n_{_X}}$	159	145	46	51
	1	943	934	836	831
MLE	1	0	0	0	0

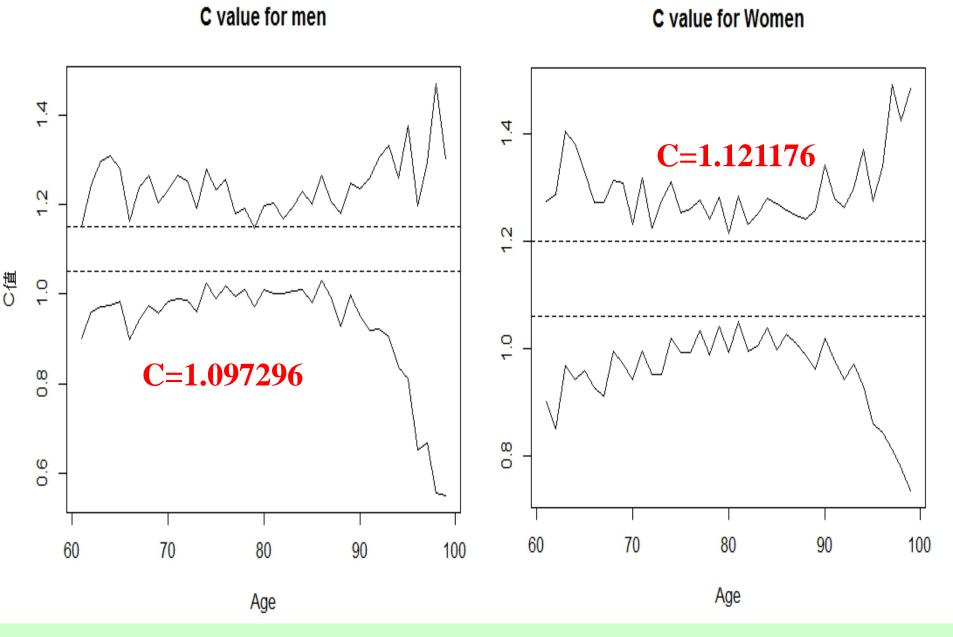
Table 5 Number of rejections among 1,000 simulations when the null hypothesis is wrong

			pan	Sweden		
Parameter	Parameter: Difference		Female	Male	Female	
B:	1%	1,000	1,000	15	95	
	0.5%	188	485	0	0	
	0.2%	0	0	0	0	
	0.1%	0	0	0	0	
C:	0.1%	1,000	1,000	1,000	1,000	
	0.01%	1,000	1,000	5	25	
	0.005%	46	198	0	0	
	0.001%	0	0	0	0	

Taiwan Male Mortality Rates (1999-2001)



台灣高齡男性死亡率編算結果比較



2009至2011年臺灣高齡死亡率的模擬檢定

使用Gompertz法則的建議

- Gompertz 法則是否適合描述高龄死亡率, 時而有支持或反對這個理論與實證研究出現,至今尚無定論。
- →可先用Bootstrap或比值法檢驗,確定是否存在一個固定的參數C值。
- →建議可用WLS估計Gompertz 法則的參數。
- →若用於推估未來死亡率,可求出各年度的B 與C,再預測未來的B、C值。

- 95歲以上的編算:(比值法; Ratio method)
- →一般而言,高龄(65歲以上)老年人資料品 質已難掌控,高高龄(85歲以上)老年人甚 至可能無法確認。(Wilmoth, 1994)
- →美國因為屬於社會福利國家,老年人健康 有Medicare的保障,可由美國社會安全機 構獲得老年人資料。

$$q_x = \frac{1}{11} [(95 - x) q_x^c + (x - 84) q_x^M], \text{ for } 85 \le x \le 94$$

Medicare data: Ages 84 and higher Census data: Ages 95 and lower

 \rightarrow 再以 $\frac{q_{x+1}}{q_x}$ 的比值取得94歲以上的死亡率:

(1) 1979-81作法:

找出符合下式的最小年龄 x,

$$\frac{q_{y+1}/q_y-1}{q_y/q_{y-1}-1} < 0.9$$

x 歲以上代入下式,推算至111歲。

$$\frac{q_{y+1}/q_y-1}{q_y/q_{y-1}-1} = 0.9 \Rightarrow q_{y+1} = q_y \left[0.9 \left(\frac{q_y}{q_{y-1}} - 1 \right) + 1 \right], 110 \ge y \ge x + 1$$

(2) 1989-91作法:
女性部份要求
$$q_{y+1}/q_y > 1.06$$
;

男性部份要求
$$q_{y+1}/q_y > 1.05$$
。

逐步由最小年齡檢查,若某一年齡不符 合要求時,即以最小要求代入,並以修正 後的死亡率代替原先的數值。



(3) 2000年作法:

與之前作法不同,不預設相鄰兩個年齡的死亡率比值。2000年蒐集之前20年所有 $q_{x+1}(t)/q_x(t)$ 的比值,再計算20年間相同年齡的比值之平均,當作2000年比值的估計值。

高龄人口的死亡率模型

■現今常見的高齡死亡模型大多與Gompertz 模型類似,具有隨時間而呈現指數遞增或 遞減的趨勢,像是美國精算學會模型:

$$_{n}q_{x}(t+k)=(g_{x})^{k}_{n}q_{x}(t);$$

Lee-Carter 模型:

$$\ln(m_{x,t}) = a_x + b_x k_t ;$$

但這些模型都過於保守,有不少研究指出它們不適用於高齡人口的死亡率。

高龄人口死亡率的其它模型

Coale and Kisker (1990)

$$k_{x} = ln(m_{x}/m_{x-1})$$

→假設 kx 在 85歲以後為年齡的線性函數

$$k_x = k_{85} - (x - 85)s$$
, $x = 85, 86, \dots 110$

$$\rightarrow m_x = m_{85} \exp(\Sigma_{h=86\sim x} k_h)$$

Coale-Kisker方法

■ Coale and Kisker (1990)提出高龄死亡率的 編算方法,想法類似Gompertz假設:

$$m_x = m_{65} \cdot \exp\left[\sum_{y=66}^x k_x\right], \quad x = 66,67,...$$

一般假設死亡率比值 $k_x = \ln(m_x/m_{x-1})$ 與年齡無關,因此上式可簡化為

$$m_x = m_{65} \cdot \exp[(x - 65)k], \quad x = 66,67,...$$

如果假設 $m_x \cong \mu_x = BC^x = B \cdot \exp[x \cdot \ln C]$ 則參數 k 與 Gompertz 假設的 $\ln C$ 相同。

Coale-Kisker方法(續)

■ Coale and Kisker以下式估計參數 k:

$$k_x = k_{x-1} - R$$
 \gtrsim $R = \frac{26k_{85} + \ln m_{84} - \ln m_{110}}{325}$

因為

$$k_{85} + \dots + k_{110} = 26k_{85} - R(1 + \dots + 25)$$

但上述的估計方法過於依賴三個年齡的死亡率資料,也有人提議修正公式:

$$k_{85}^* = \frac{k_{82} + \dots + k_{88}}{7} \not \supset m_{84}^* = \frac{m_{82} + \dots + m_{86}}{5}$$

註:男女性的 m_{10} 在CK方法分別假設1.0及0.8。

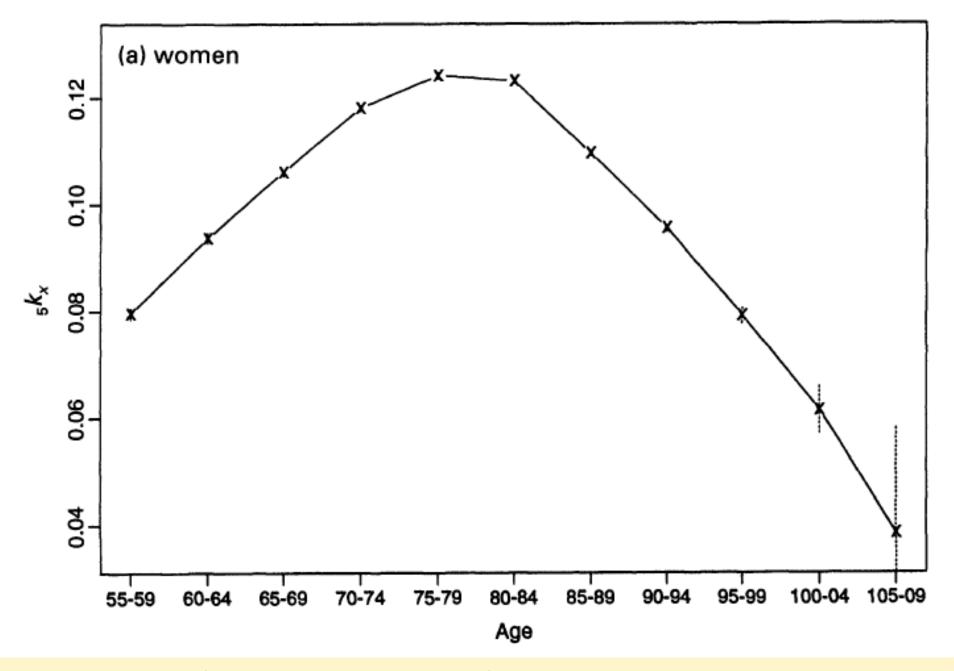
Coale-Kisker方法(續)

不合理現象。

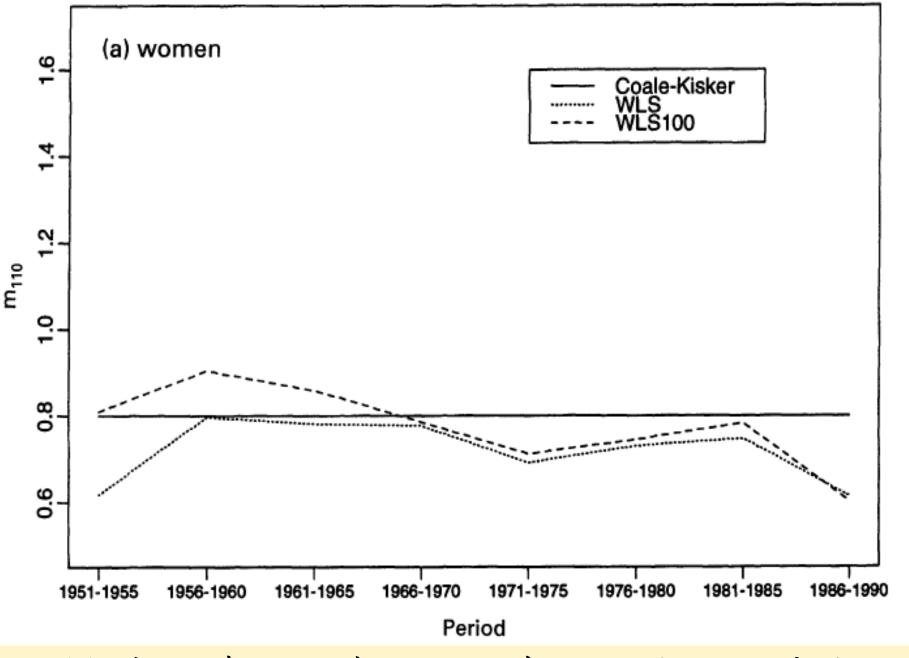
■加權最小平方法(WLS)仿造Gompertz的估計方式,但使用中央死亡率與二次式:

$$\min_{\alpha,s} \sum_{x} w_{x} \left(\ln m_{x} - \alpha - k_{85}(x - 84) + \frac{(x - 84)(x - 85)}{2} s \right)^{2}$$

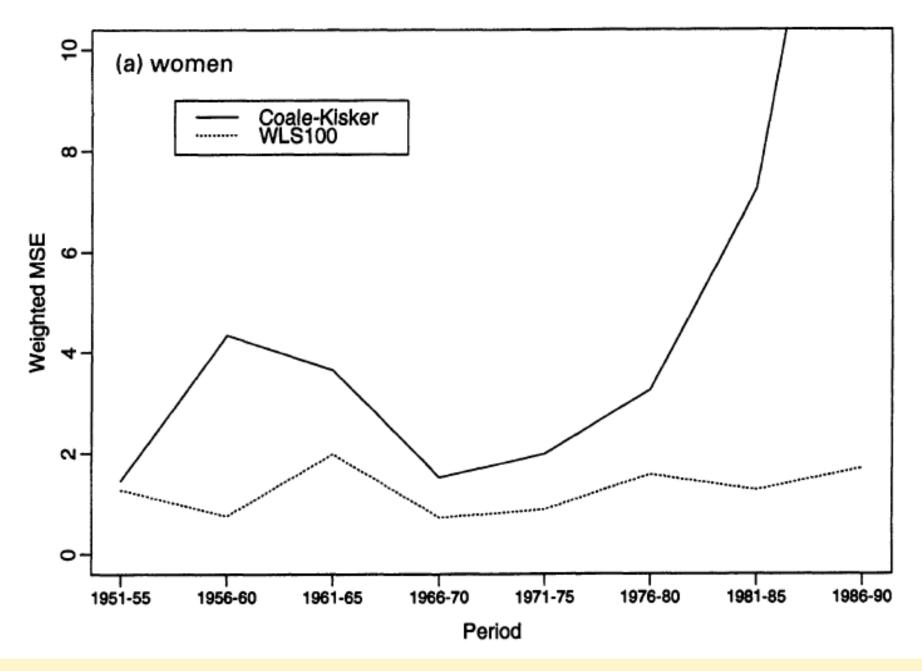
上式的權數可代入各年齡的人數(85~110歲)。 註:WLS的優點在於估計值不受離群值影響; 以84歲及110歲頭尾死亡率估計參數,未考慮 其他年齡的死亡率資訊;以各年齡分別估計 k_x 值也有類似問題,可能因樣本數不足而產生



1971-1990年五龄組的死亡率比值信賴區間(日本女性)



CK模型110歲死亡率的上限實證估計值(日本女性)



CK模型的Goodness-of-fit 誤差比較(日本女性)

Wang & Yue (2011) Model

- Discount Sequence (bandit problems)
- → Defined a random variable N with

$$\alpha_n = P(N \ge n), \quad \gamma_n = \sum_{i=n}^{\infty} \alpha_i.$$

The sequences $(\alpha_1, \alpha_2, ...)$ and $(\gamma_1, \gamma_2, ...)$ are regular, if the following conditions are satisfied

$$\alpha_n \cdot \alpha_{n+2} \le (\alpha_{n+1})^2 \text{ or } \gamma_n \cdot \gamma_{n+2} \le (\gamma_{n+1})^2$$



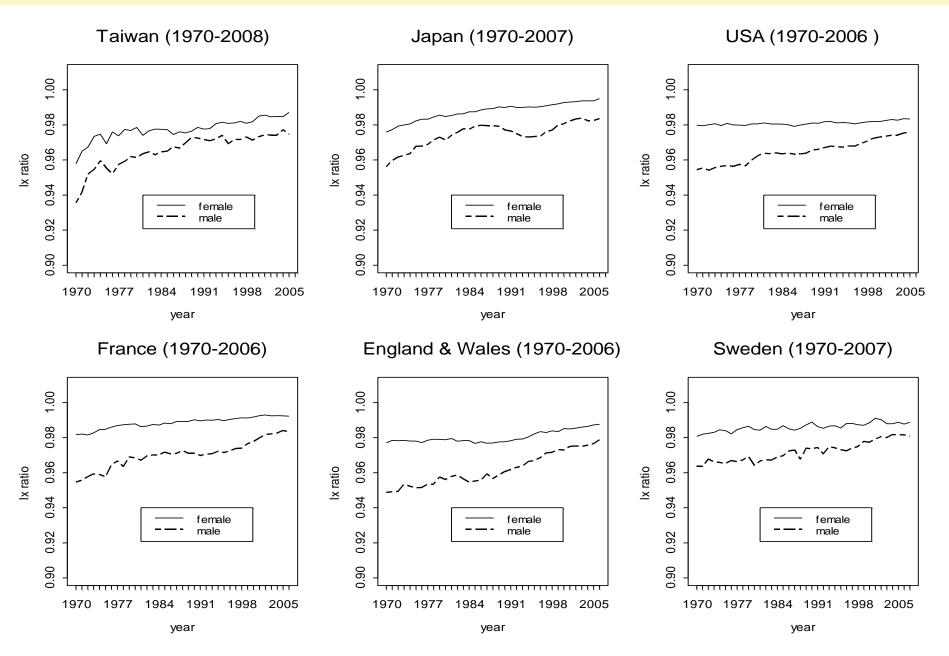
Wang-Yue Model (Conti.)

Let the survival time T and α_x as the survival function $S(x) = P(T \ge x)$. Then, the relationship between Discount Sequence and mortality is

$$l_n \cdot l_{n+2} \le (l_{n+1})^2$$
 or $\frac{l_n \cdot l_{n+2}}{(l_{n+1})^2} \le 1$

where l_n is the number of survivors at age n in life table.

臺灣、日本、美國的しな比值



Wang-Yue Model的估計

- Modeling the Discount Sequence Ratio:
- →Brownian Motion Stochastic Differential Equations Model (BMSDE)

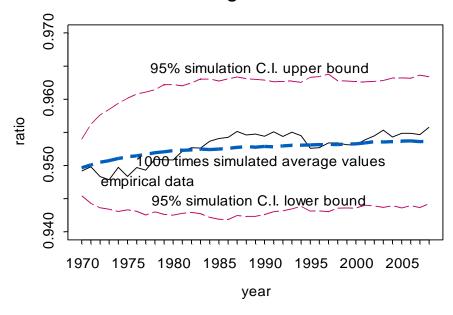
$$dr_x(t) = k(\theta - r_x(t)) \cdot dt + \sigma \cdot dz, \quad dz \sim N(0, dt)$$

k: is the strength, strong or weak, for $r_x(t)$ returning to the average level θ

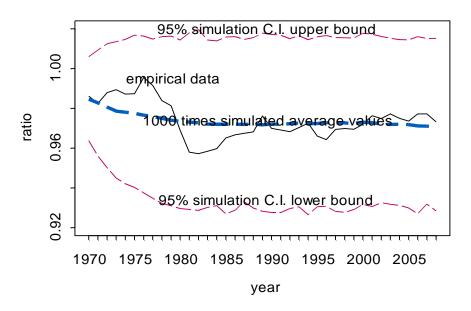
 θ : the long-term average level of $r_x(t)$.

 σ : Diffusion coefficient

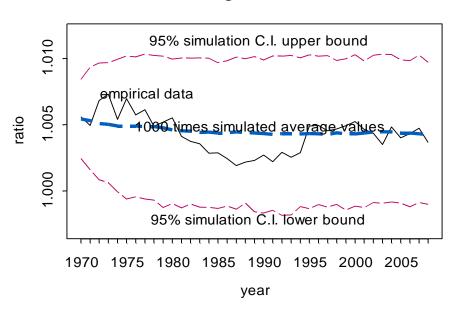
male age 0 ex ratio



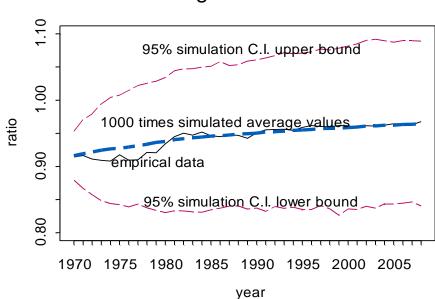
male ages 60~64 ex ratio



male age 0 lx ratio



male ages 60~64 lx ratio



CBD(Cairns, Blake, & Dowd)模型

• Cairns et al. (2006):

logit
$$(q_{x,t}) = k_t^{(1)} + k_t^{(2)}(x - \bar{x}) + \varepsilon_{x,t}$$

其中, $k_t^{(1)}$ 、 $k_t^{(2)}$:期間效應(Period Effect)

 $k_t^{(1)}$: 死亡率水準(Level of Mortality)

 $k_t^{(2)}$: 斜率(Slope)

 $\varepsilon_{x,t}$: 隨機誤差



- 依據各國高齡資料,分別以估計、預測(交叉驗證)比較各模型。
- →估計比較加入模型參數個數為考量,使用 BIC及AIC:BIC=-2log(L)+klog(n) 其中L為最大概似估計量,k為模型中參數個 數,n為觀察值個數。
- →預測比較將資料分成訓練(Training)、測試 (Testing)兩組,訓練資料的參數估計值代入 測試資料,計算測試資料的誤差。

高齡模型的估計比較(1/2)

- 先以臺灣、日本、美國的歷年高齡資料,測試常見的幾種高齡模型,以加入參數個數的指標為比較標準(BIC及AIC),資料格式1970年至2009年、60~99歲五齡組。
- →為確定模型配適與資料格式無關,也考量單齡或五齡組、年齡分組(60-99歲及85-99歲)及資料期間(1970-2009年、1980-2009年、1990-2009年),以BIC較佳前三名次數為評斷標的。

臺灣、日本、美國高齡死亡率估計結果

lat mil	台灣		日	本	美國		
模型	男性	女性	男性	女性	男性	女性	
Gompertz	III -645	III -822	I -1162	II -1182	II -1338	II -1318	
CK	-466	-548	-268	-410	I -1421	I -1661	
DSW	I -1208	I -1418	II -887	I -1221	-406	-616	
LC	II -757	II -1020	III -805	III -961	III -932	III -1276	
RH	-581	-640	-1474	-652	-655	-737	
CBD	-228	-390	284	189	262	173	

註:I的結果最佳、II次之、III第三。

不同年齡組、資料時間的推薦模型(BIC)

年龄組	資料期間	順序	台灣		日本		美國	
			男性	女性	男性	女性	男性	女性
	1970- 2009	1	DSW	DSW	G	DSW	CK	CK
		2	LC	LC	DSW	G	G	G
		3	G	G	LC	LC	LC	LC
	1980 -2009	1	DSW	DSW	DSW	DSW	DSW	DSW
		2	G	LC	G	G	G	G
		3	LC	G	LC	LC	LC	LC
	1990 -2009	1	G	DSW	G	LC	DSW	DSW
		2	DSW	LC	DSW	DSW	LC	LC
		3	LC	G	LC	G	G	G

高齡模型BIC指標的前三名次數

模型	五龄組	單龄 (高龄)	單龄 (高高龄)	各模型 總數
Gompertz	18	0	4	22
CK	2	5	17	24
DSW	16	11	7	34
LC	18	17	4	39
RH	0	18	16	34
CBD	0	3	6	9
各年龄總數	54	54	54	162

註:紅字及綠字分別為顯著較大、顯著較小。

高齡模型的預測比較(2/2)

- 中長期預測誤差較大,僅考量短期預測,以 過去十年資料預測未來五年的死亡率。
- →台灣、日本、美國男性與女性60-99歲年齡死亡率,資料時間為1970~2009年。
- →總計有26種組合。
- →考量LC、CBD、DS三個模型,
- LC: ARIMA及區塊拔靴法
- CBD: ARIMA及區塊拔靴法
- · DS: 兩種區塊拔靴法

高龄模型五年預測誤差MAPE(單位%)

模型	台灣		日本		美國		T. 11.
	男性	女性	男性	女性	男性	女性	平均
DS1	9.9	8.9	5.2	5.2	3.7	4.3	6.2
DS2	8.3	7.4	4.7	5.0	3.4	4.1	5.5
LC1	8.3	7.3	4.3	4.3	3.5	4.0	5.3
LC2	9.2	7.5	4.6	4.1	3.5	4.5	5.6
CBD1	8.0	7.4	5.2	7.2	3.8	6.4	6.3
CBD2	8.3	6.7	5.1	7.1	4.2	7.0	6.4

註:數字加底色者為模型預測最佳者。

高龄模型預測前三名的總次數

模型	台灣		日本		美國		1.5 1 b)
	男性	女性	男性	女性	女性	男性	總次數
DS1	6	2	7	14	11	14	54
DS2	11	13	13	16	17	18	88
LC1	17	13	18	19	19	23	109
LC2	8	13	18	25	18	20	102
CBD1	17	15	12	4	9	3	60
CBD2	19	22	10	0	4	0	55

註:數字加底色者為模型在該國預測最佳者;紅字、綠字分別為顯著較佳、顯著較差者。

高齡死亡率的編算建議

- 臺灣現行的Gompertz法則,適合用於單年五 龄組資料,暫時無須更動。
- 本研究比較的關係模型、隨機模型都可用於 高龄死亡率,其中:
- →若有充足資料,亦即有最高編算年齡的歷史 資料,Lee-Carter模型為首選;
- →若高齡資料不足、或品質不佳,可採用可外 推的關係模型,折扣模型是不錯的選擇。