

# 修勻學(Graduation) — 參數(Parametric)修勻法

授課教師：余清祥教授

課程日期：2024年11月13日

資料下載：

<http://csyue.nccu.edu.tw>





## 參數(Parametric)修勻

---

- 實務上，理論模型較容易處理，也較容易用於預測未來的死亡率，更合適處理隨機生命表(Stochastic Life Table)。
- 近年較為常見的參數死亡模型有：
  - Lee-Carter Model
  - Reduction Factor Model (或GM Model)
  - APC Model

註：參考Generalized Age-Period-Cohort (GAPC)模型(R package 「StMoMo」)。



# Lee-Carter 模型

---

- Lee and Carter (1992)提出預測美國未來死亡率，模型如下：

$$\ln(m_{xt}) = \alpha_x + \beta_x \kappa_t + \varepsilon_{xt}$$

其中

$\kappa_t \rightarrow$  死亡率強度的變化量

$\alpha_x \rightarrow$  年齡組死亡率的平均曲線

$\beta_x \rightarrow$  年齡組相對死亡率的變化速度



# Singular Value Decomposition (SVD)

---

Lee-Carter模型的參數估計透過

$$\text{Minimize } \sum_{xt} (\ln(m_{xt}) - \alpha_x - \beta_x \kappa_t)^2$$

→ 通常以SVD法求解，也就是分解矩陣

$$(\ln(m_{xt}) - \alpha_x) = UPV^T$$

→ Renshaw 與Haberman (2003)提出加權估計法，以解決SVD在人數較少時遇到的求解問題



# Singular Value Decomposition

---

- In regression problem, we have

$$Y = X\beta + \varepsilon \Rightarrow U'Y = U'X\beta + U'\varepsilon$$

or equivalently, ( $\theta = V'\beta$  &  $X_1^*V = D$ )

$$\begin{aligned} Y^* &= \begin{pmatrix} X_1^* \\ 0 \end{pmatrix} \beta + \varepsilon^* = \begin{pmatrix} X_1^* \\ 0 \end{pmatrix} VV'\beta + \varepsilon^* \\ &= \begin{pmatrix} D \\ 0 \end{pmatrix} \theta + \varepsilon^* \end{aligned}$$

Note: SVD is often used for regression diagnostics, data reduction, and graphical clustering.



---

## ■ The Singular-Value Decomposition

→ Let  $X$  be an arbitrary  $n \times p$  matrix with  $n \geq p$ .

Then there exists orthogonal matrices  $U: n \times n$   
and  $V: p \times p$  such that  $U' X V = \tilde{D} = \begin{pmatrix} D \\ 0 \end{pmatrix}$ ,

where  $D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_p \end{pmatrix}$  with  $d_1 \geq d_2 \geq \dots \geq d_p \geq 0$ .



---

- Example of SVD:

→ You can check if the command “svd” in R returns correct outputs.

```
A=matrix(c(1:12),ncol=3)
```

```
aa=svd(A)
```

```
t(aa$u)%*%A%*%aa$v    # Diagonal!
```

```
aa$u%*%diag(c(aa$d))%*%t(aa$v)
```

```
t(aa$u)%*%A    # Upper triangular!
```



## SVD Interpretation of Lee-Carter Model

---

Applying the SVD, i.e.,  $(\ln(m_{xt}) - \alpha_x) = UPV^T$ , the matrix  $U$  represents the time component,  $P$  is the singular values, and  $V$  is the age component.

→  $K_t$  is derived from the first vector of the time-component matrix and the first singular value, and  $\beta_x$  is from the first vector of the age-component matrix. Other vectors correspond to the residuals.



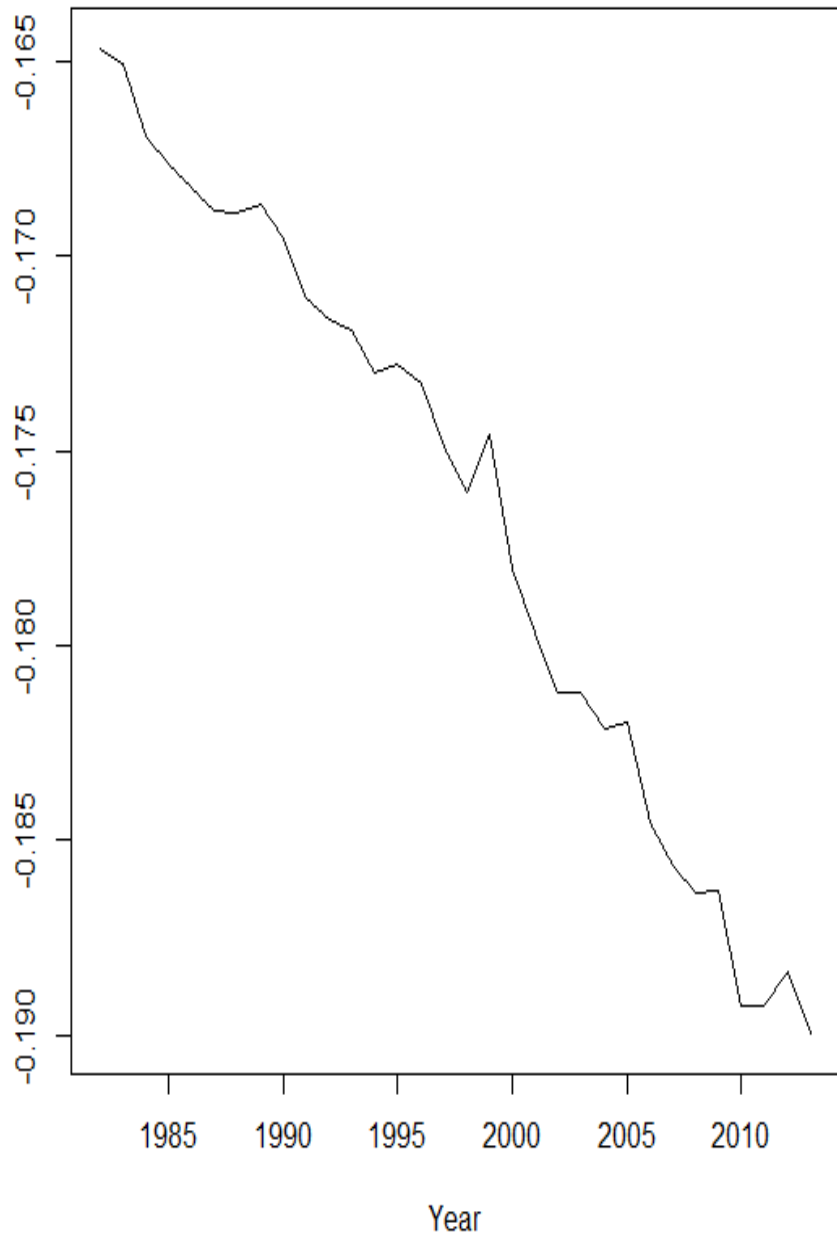


# Lee-Carter模型實例

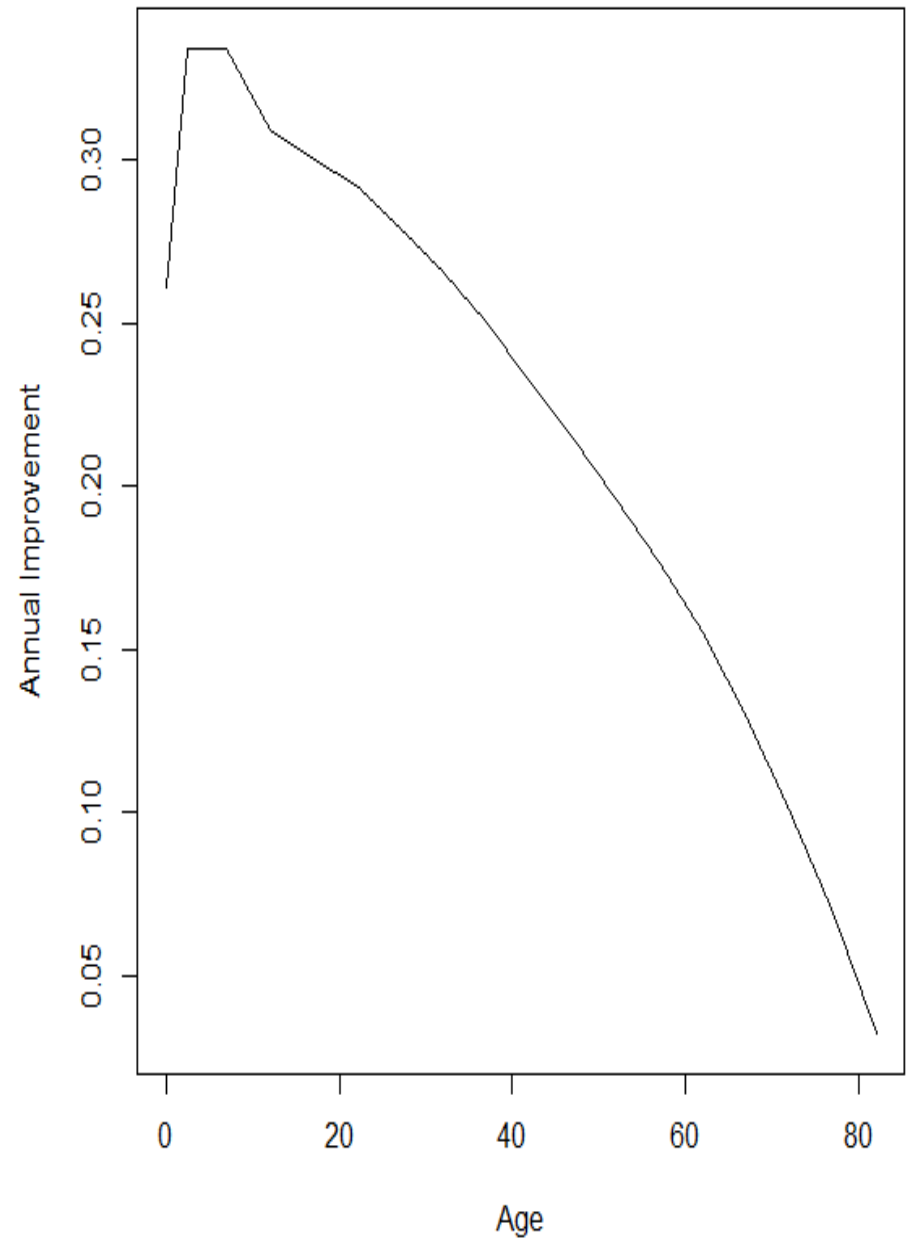
---

```
a1=matrix(scan("d:\\Taiwan(Female).txt"),ncol=18,byrow=T)
loga=log(a1)
b1=svd(loga)
b11=b1$d
b11[1]/sum(b11)
s1=b1$u[,1]
x1=c(0,2.5,7,12,17,22,27,32,37,42,47,52,57,62,67,72,77,82)
par(mfrow=c(1,2))
plot(1982:2013,s1,type="l",xlab="Year",main="kt in Lee-Carter
Model")
plot(x1,b1$v[,1],type="l",xlab="Age",main="Beta_x")
```

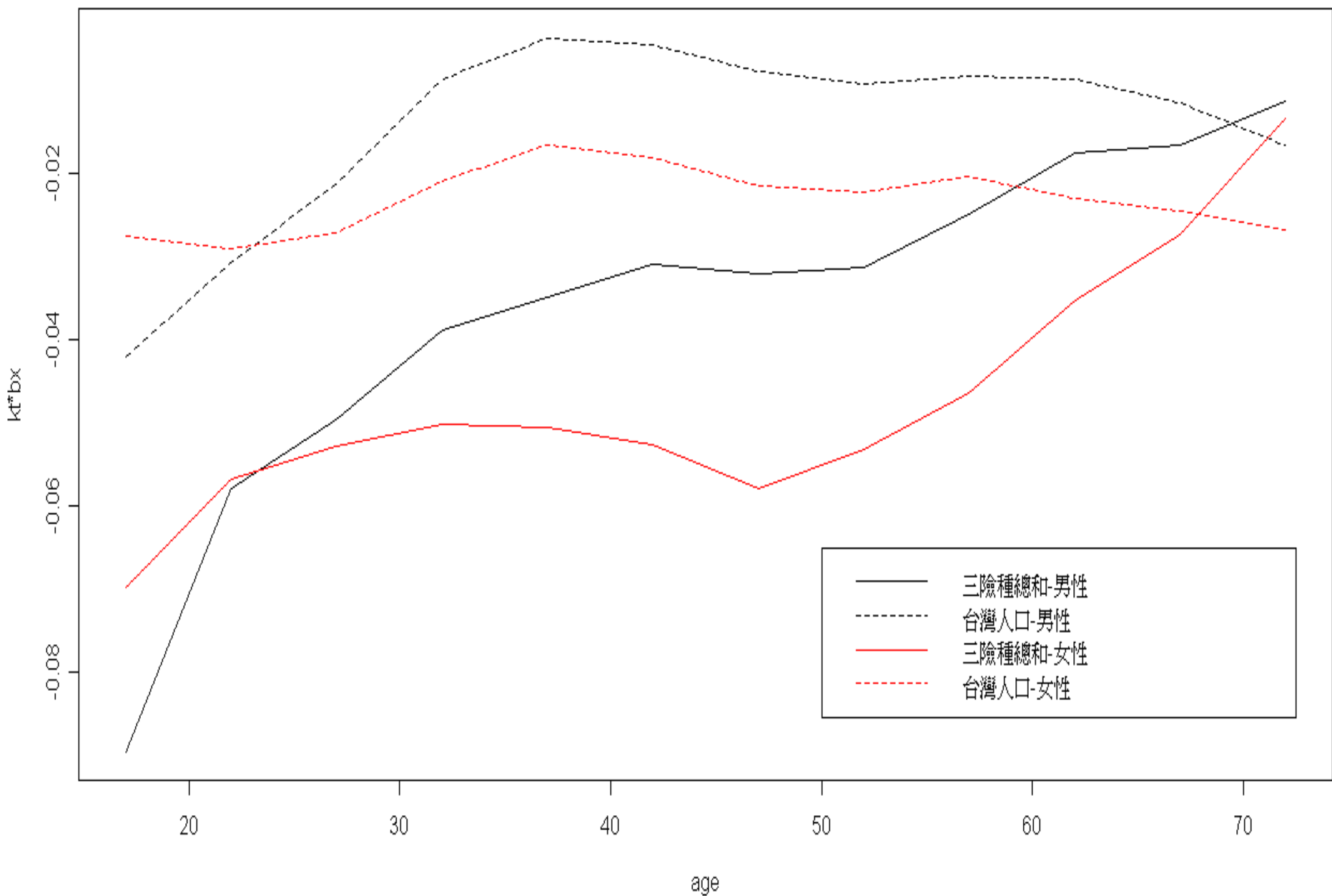
**kt in Lee-Carter Model**



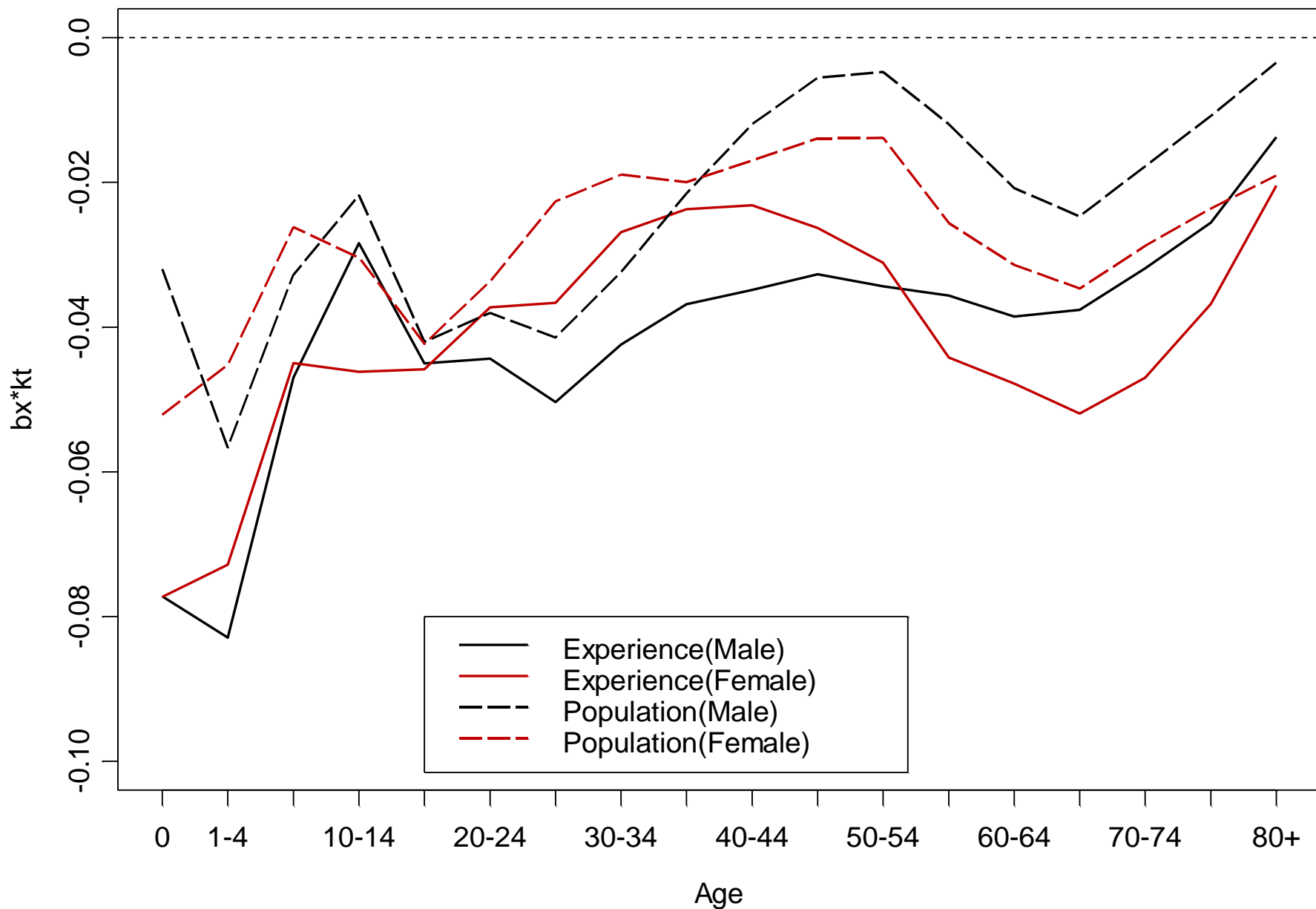
**bx in Lee-Carter Model**



# 壽險業死亡率年改善率 (LC模型、1982~2007年)



# 壽險業死亡率年改善率 (LC模型、2000~17年)



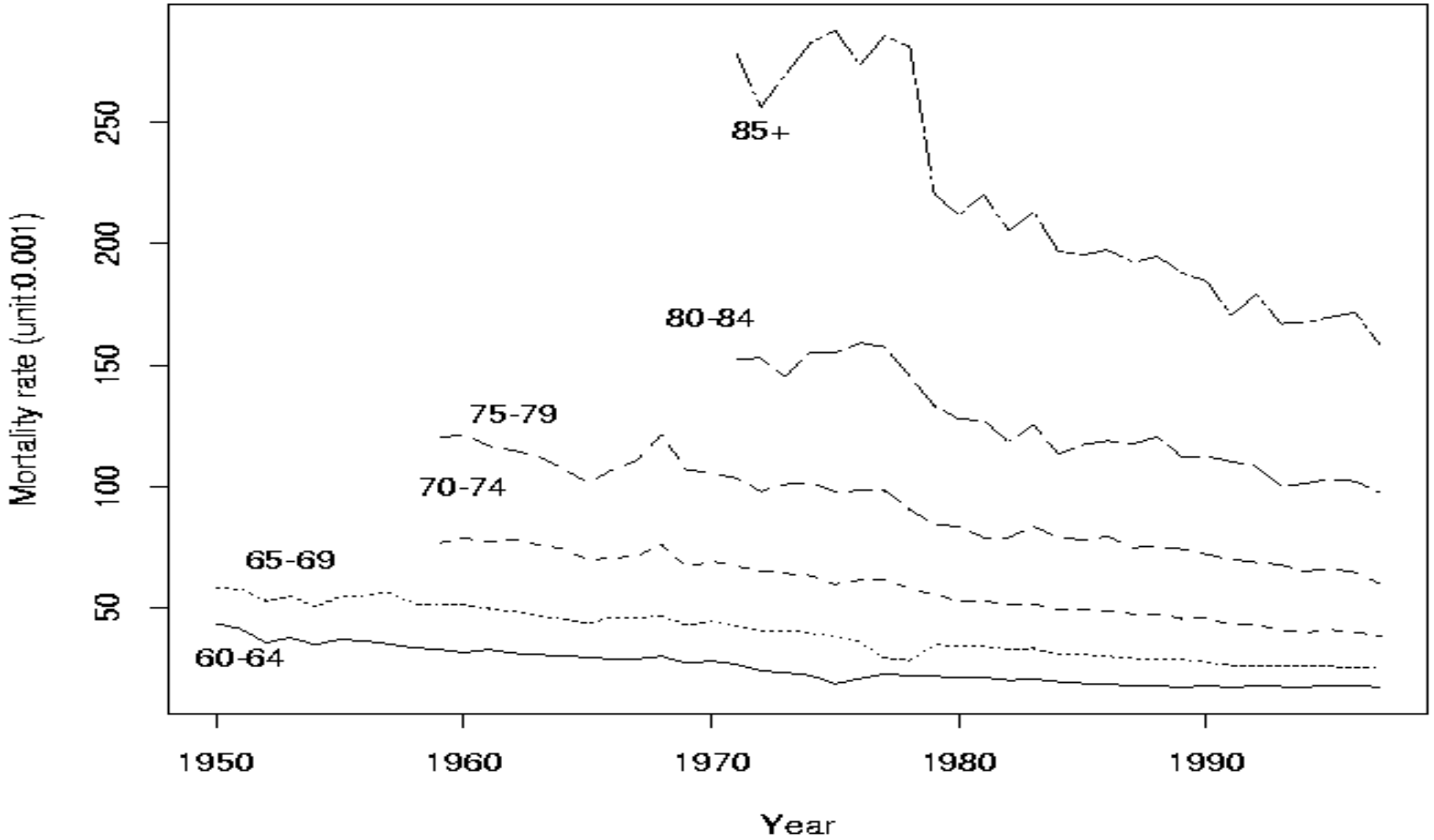


- 高齡死亡率資料的考量：

→ 過去缺乏台灣地區85歲以上高齡人口資料，近年來才有較完整的紀錄，例如：「台閩地區人口統計」中的最高年齡組在1971年為85+；1993年延至95+；1998年時至100+。

→ 與年輕的老年人比較，85歲以上老年人的資料品質較具爭議，而且驗證也相對困難。





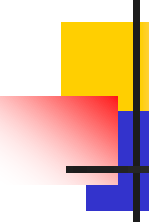
台灣地區男性死亡率(高齡不完整)



# SVD近似法(Approximation)求解

利用SVD近似法求解

1.  $\alpha_x$  為各年齡組的  $\ln(m_{x,t})$  在各年度的平均
2.  $\kappa_t$  為每年  $(\ln(m_{x,t}) - \alpha_x)$  在各年齡組的總和
3.  $\beta_x$  是  $(\ln(m_{x,t}) - \alpha_x)$  對  $\kappa_t$  配適無截距迴歸式之斜率，為使表示法唯一，加入  $\sum \kappa_t = 0$ 、 $\sum \beta_x = 1$  兩個限制式。
4. 以時間序列預測未來的參數值，再代回取指數得到年齡組死亡率預測值。



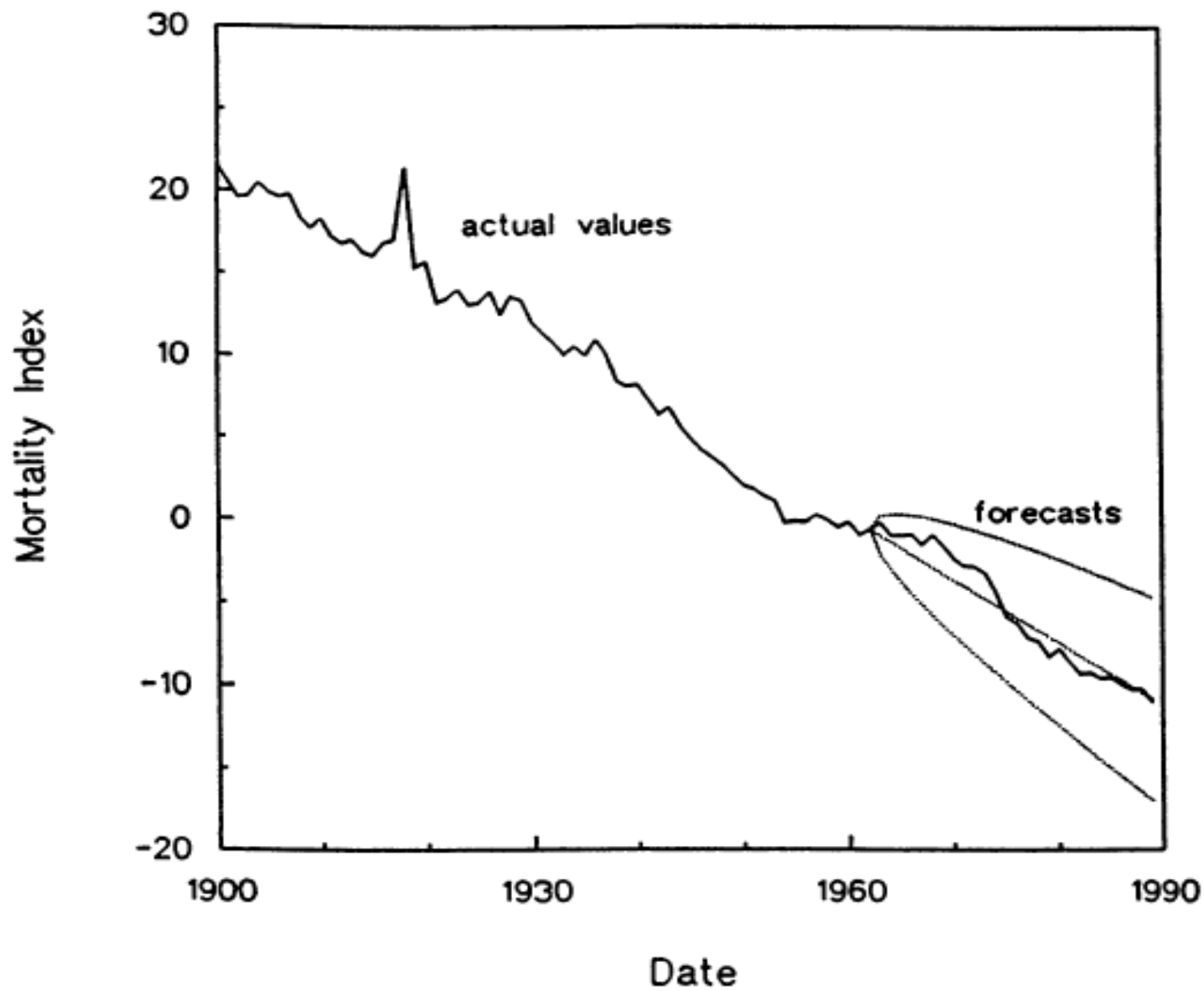
註： $\kappa_t$  為含位移的一階自我迴歸(First order autoregression with drift)的時間序列模型，通常假設：

$$\kappa_t = Z + \kappa_{t-1} + e_t$$

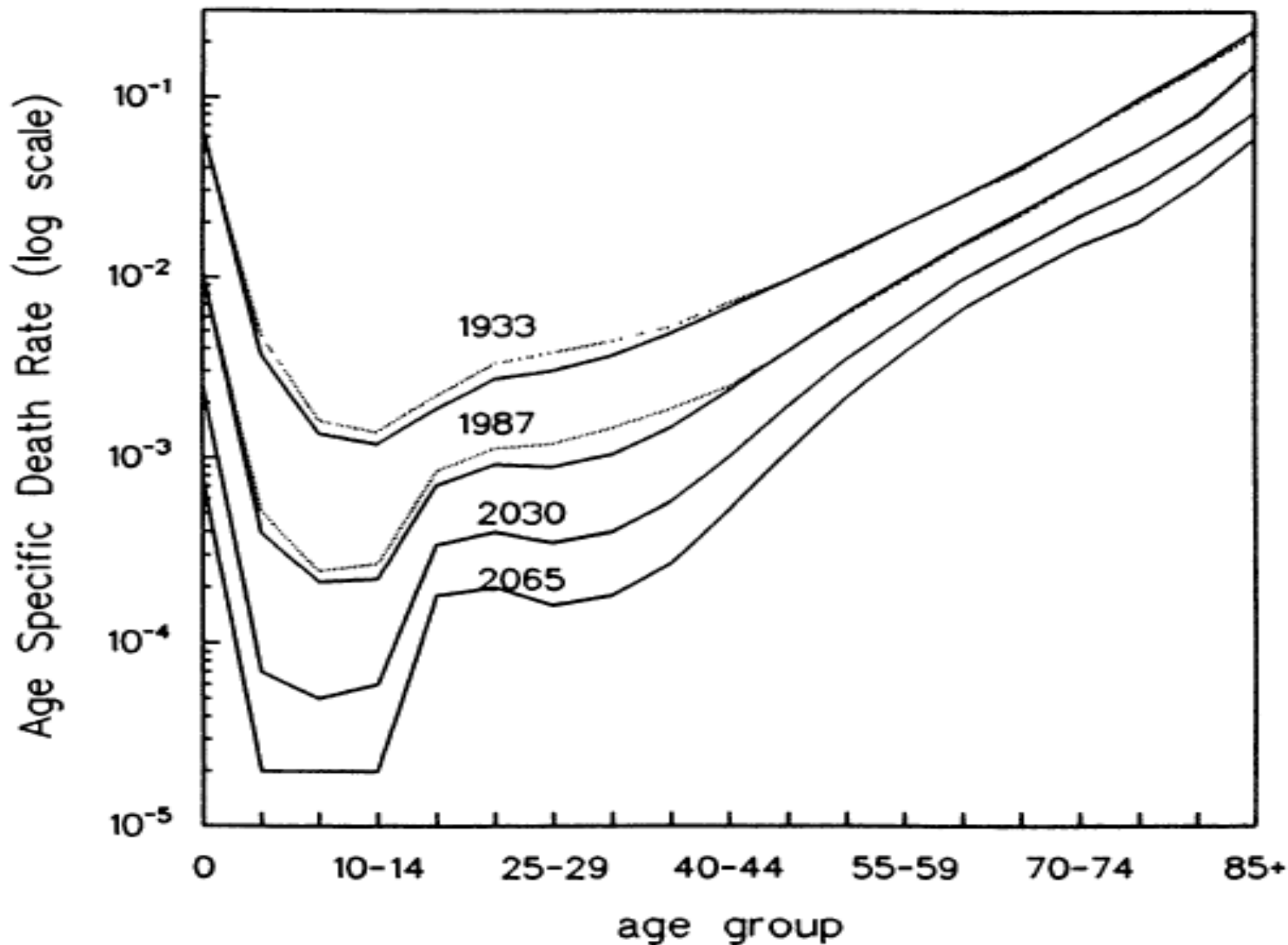
其中  $Z$  為平均位移常數， $e_t$  代表著隨機誤差項。

- Wilmoth (1993) 提出 WLS 及 MLE 兩種修正
  - 處理某些年齡特定死因為 0 的問題
  - 類似加權平均降低死亡率的變異數





*Figure 5. Within-Sample Forecast of Mortality Index, k Model Fitted on 1933–1962; Forecast for 1963–1989.*



*Figure 4. Comparison of Actual, Fitted, and Forecast Mortality by Age, for Sexes Combined, at Selected Dates. For 1933 and 1987, the last age group is 85+. For 2030 and 2065, it is 85–89 and, therefore, less than 85+. Actual schedule are dots. Fitted and forecasts are solid.*

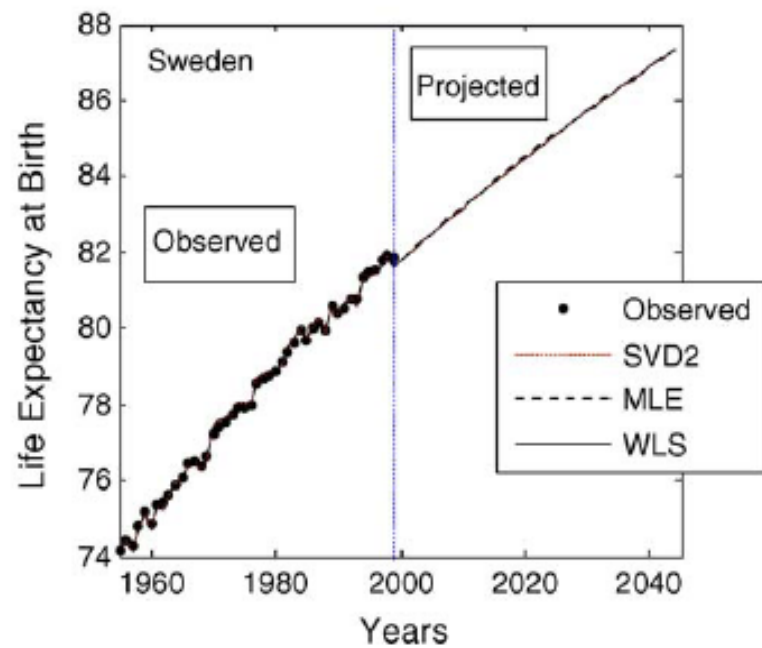
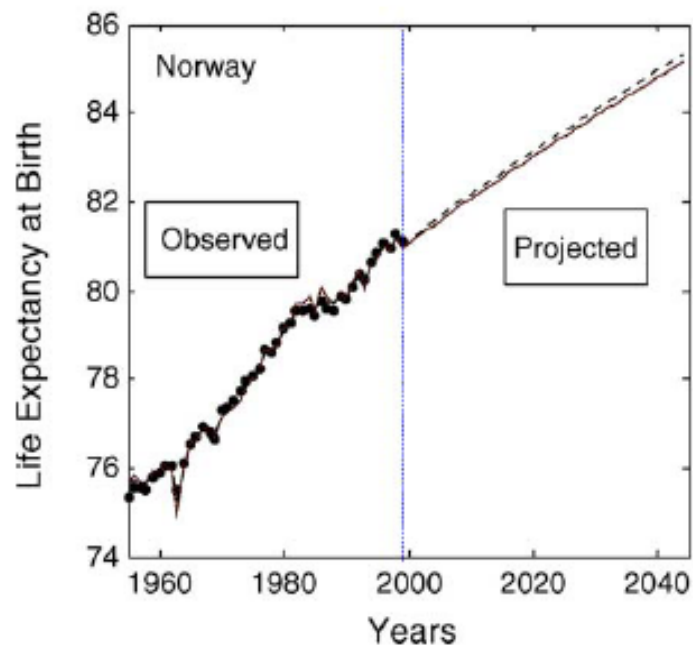
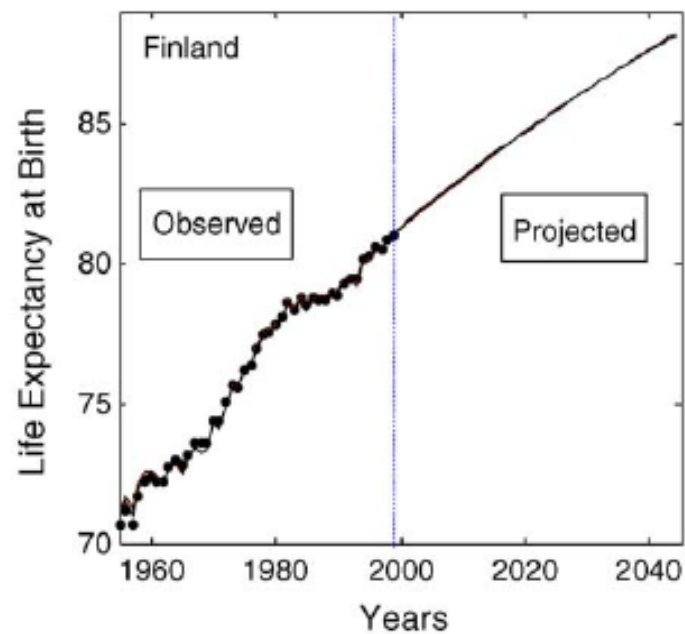
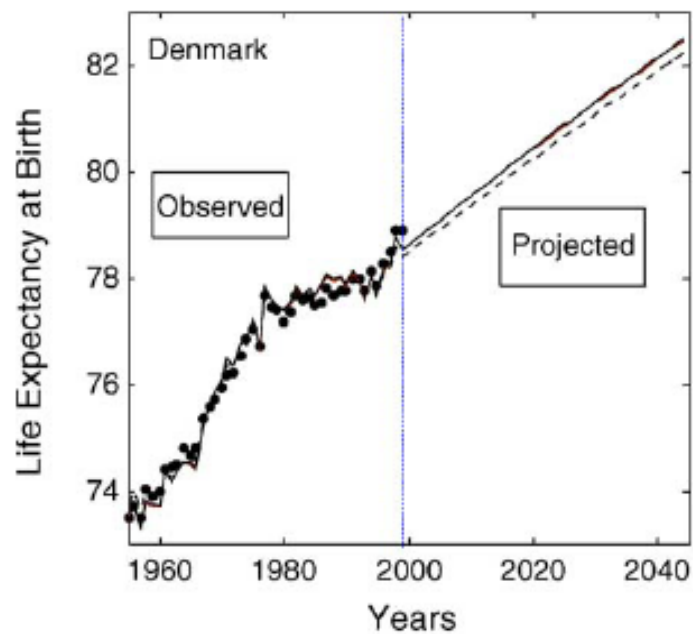
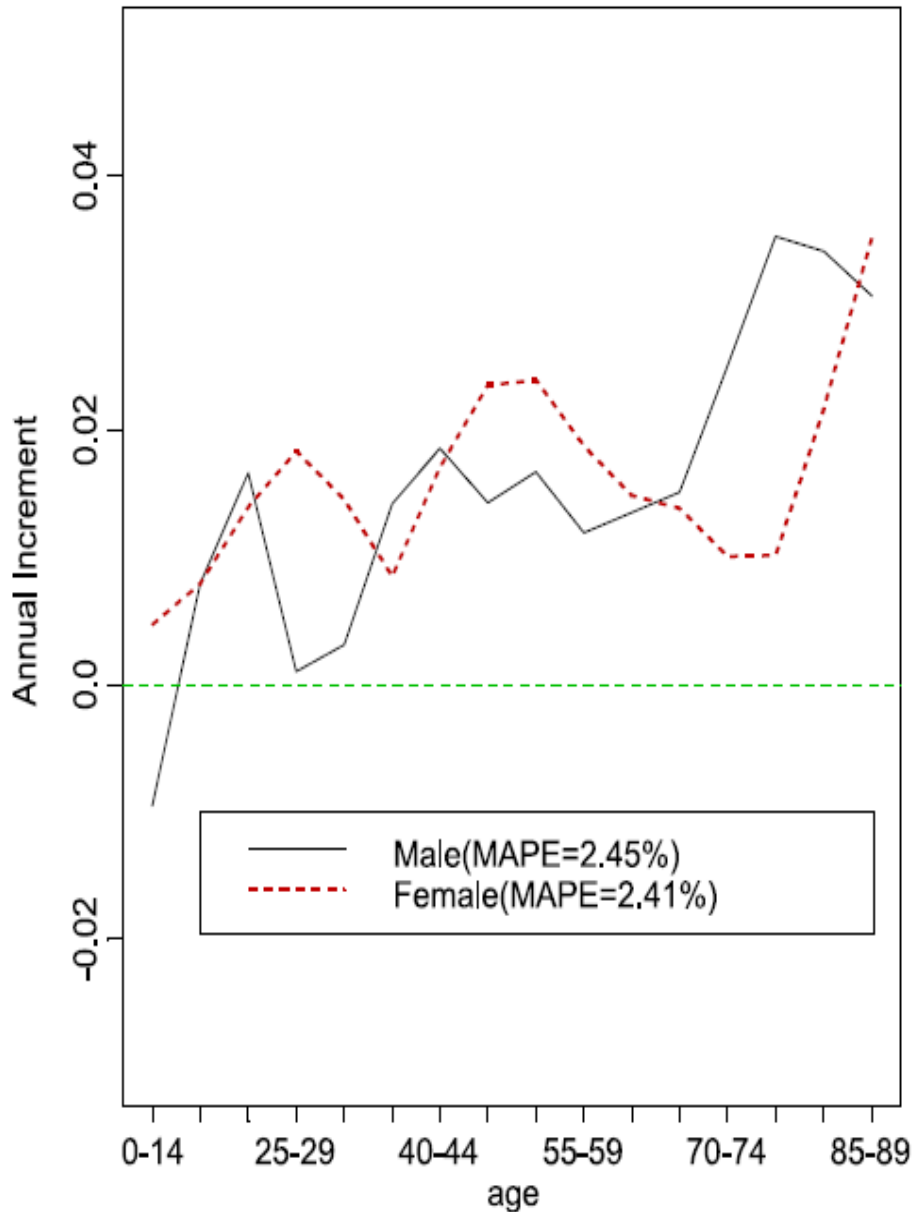


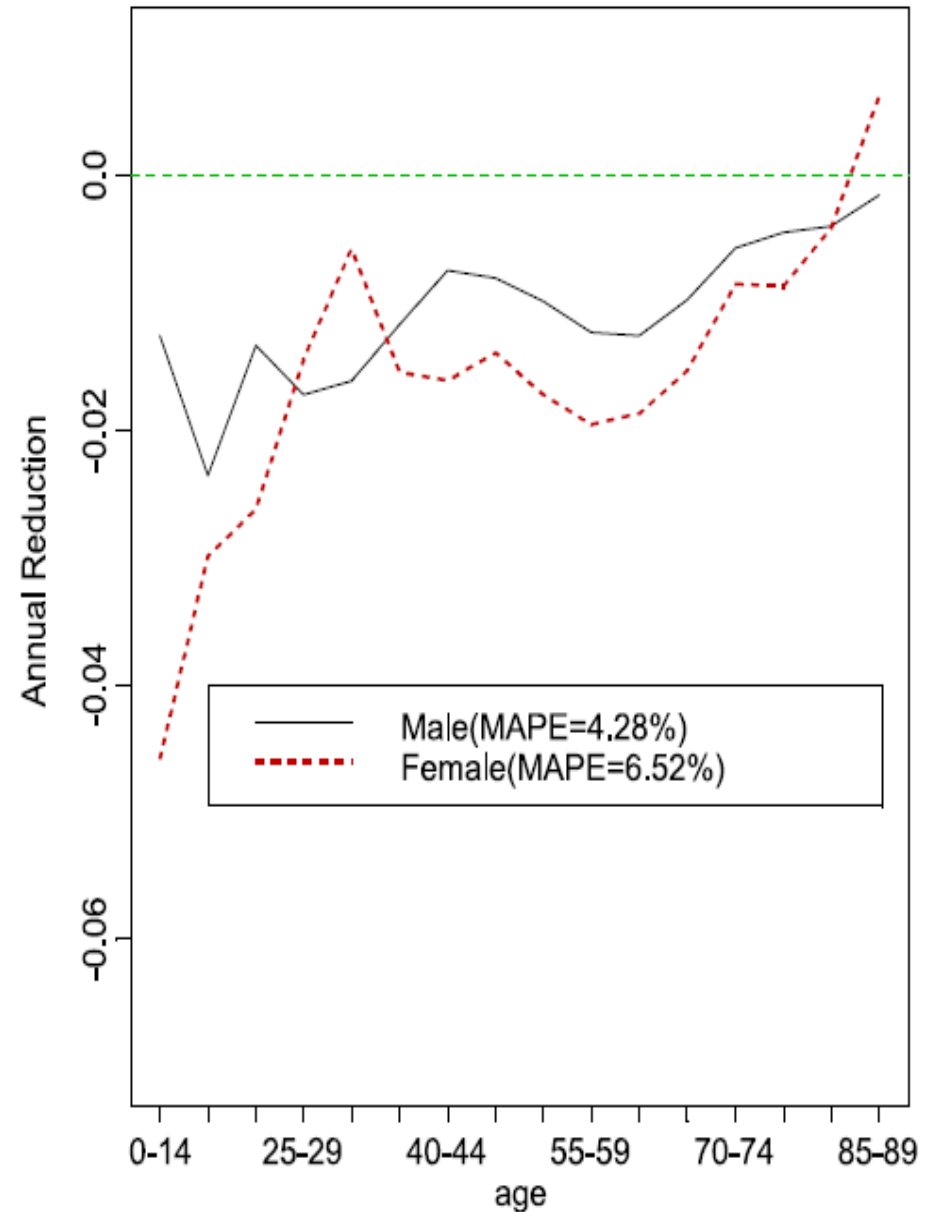
Fig. 2. Fitted (1955–1999) and forecasted (2000–2044) life expectancy at birth, Nordic countries, women.

# 癌症發生率、死亡率的Lee-Carter估計結果

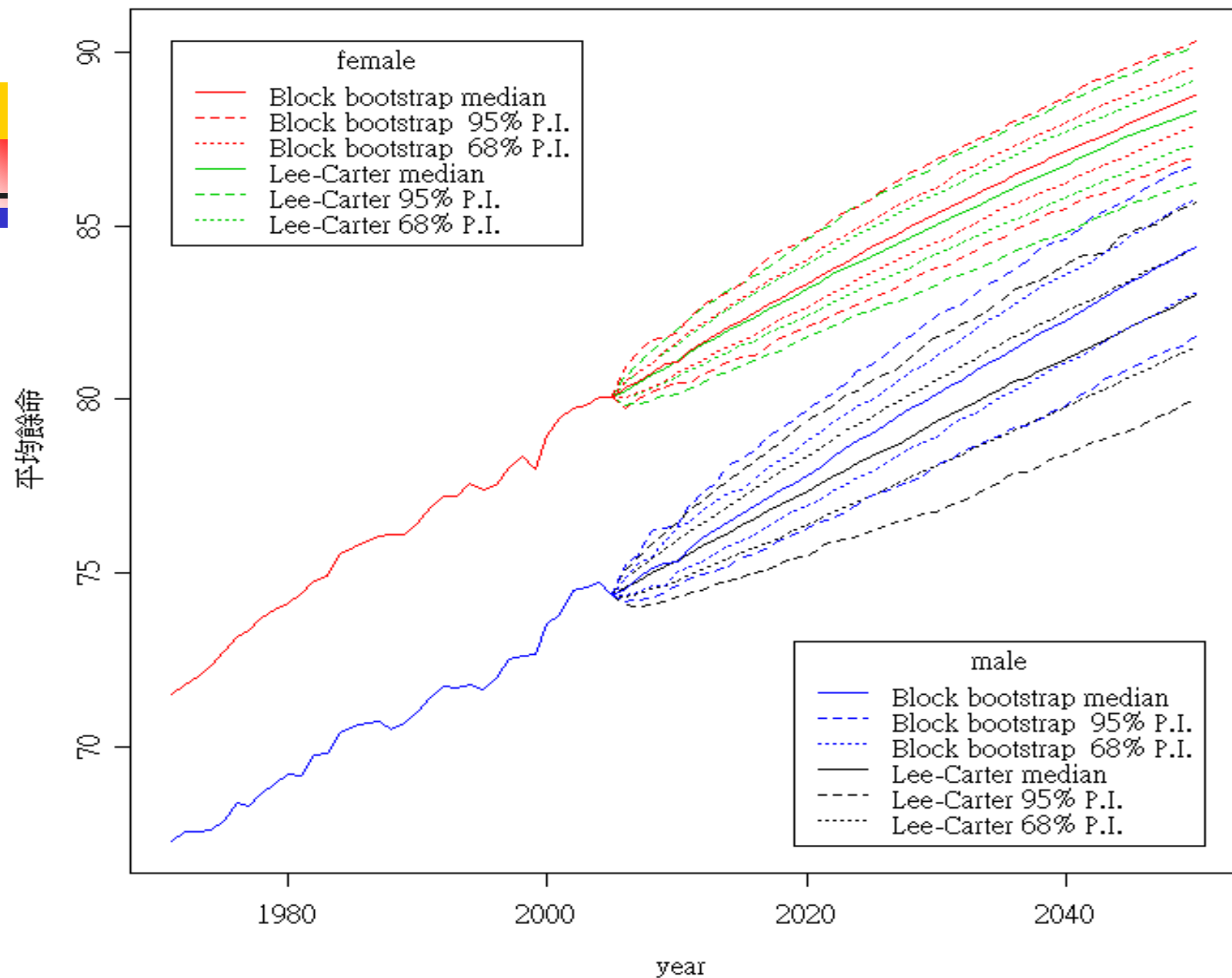
Incidence



Mortality



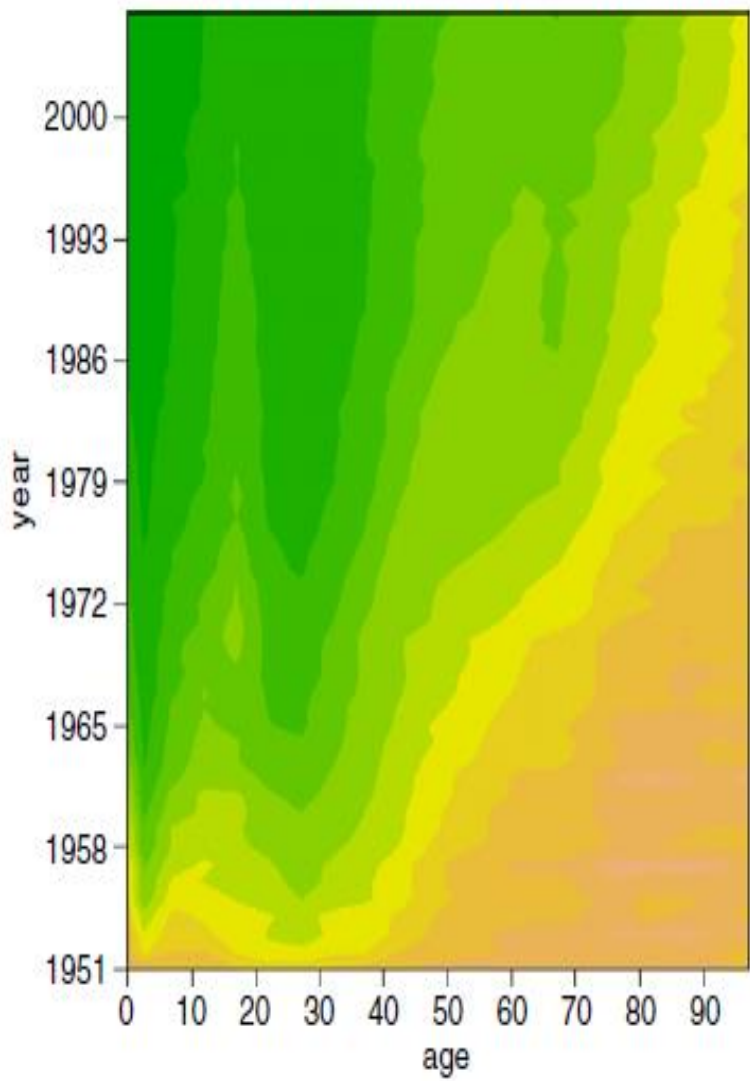
# TAIWAN 零歲平均餘命



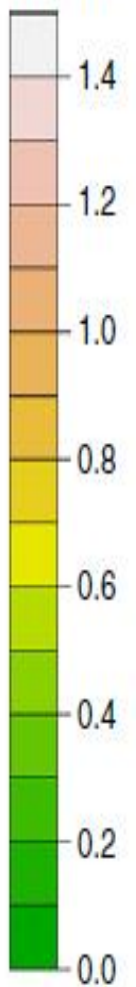
# 範例、各國年齡死亡率的變化趨勢

→ Human Mortality Database (日本、英國)

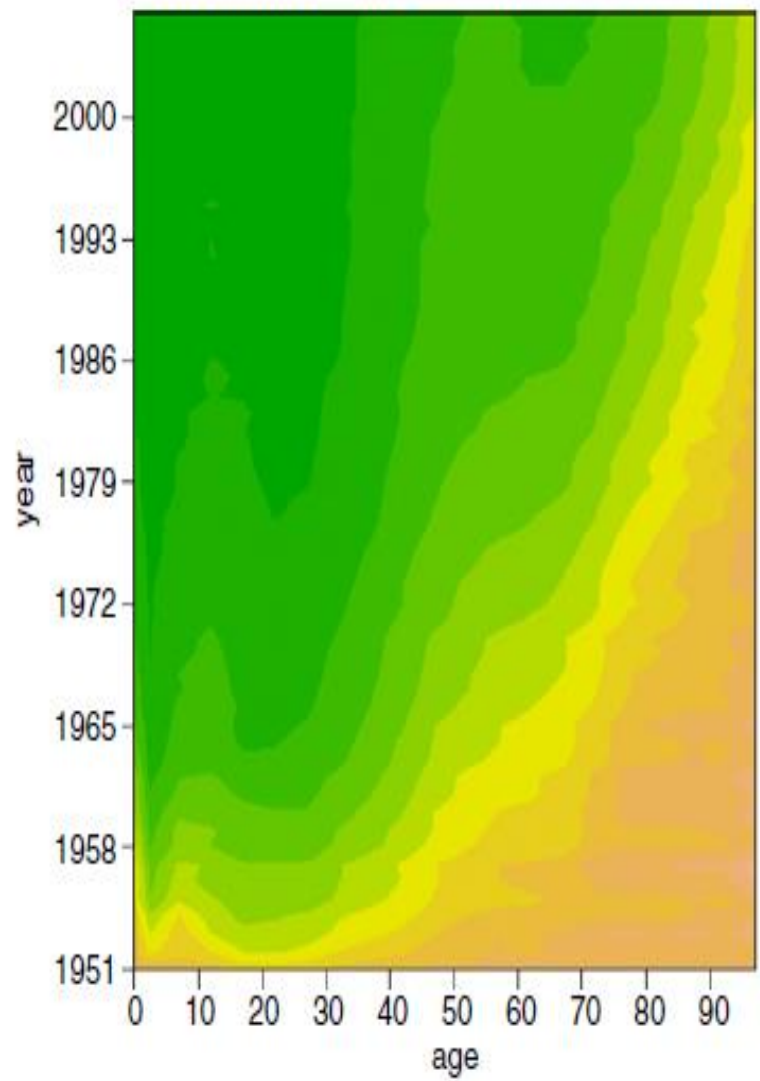
Reduction Factor (JP male 1951-2006)



Reduction



Reduction Factor (JP female 1951-2006)

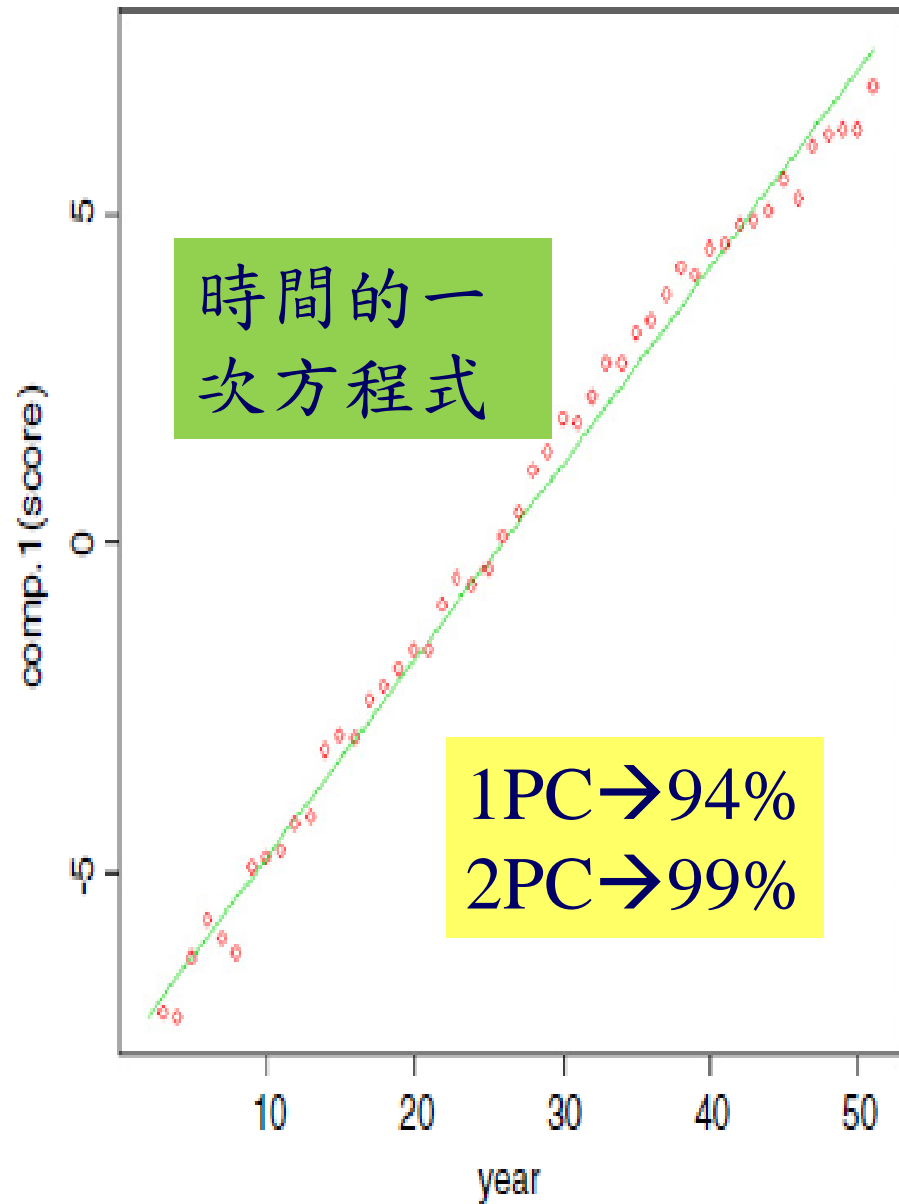


Reduction

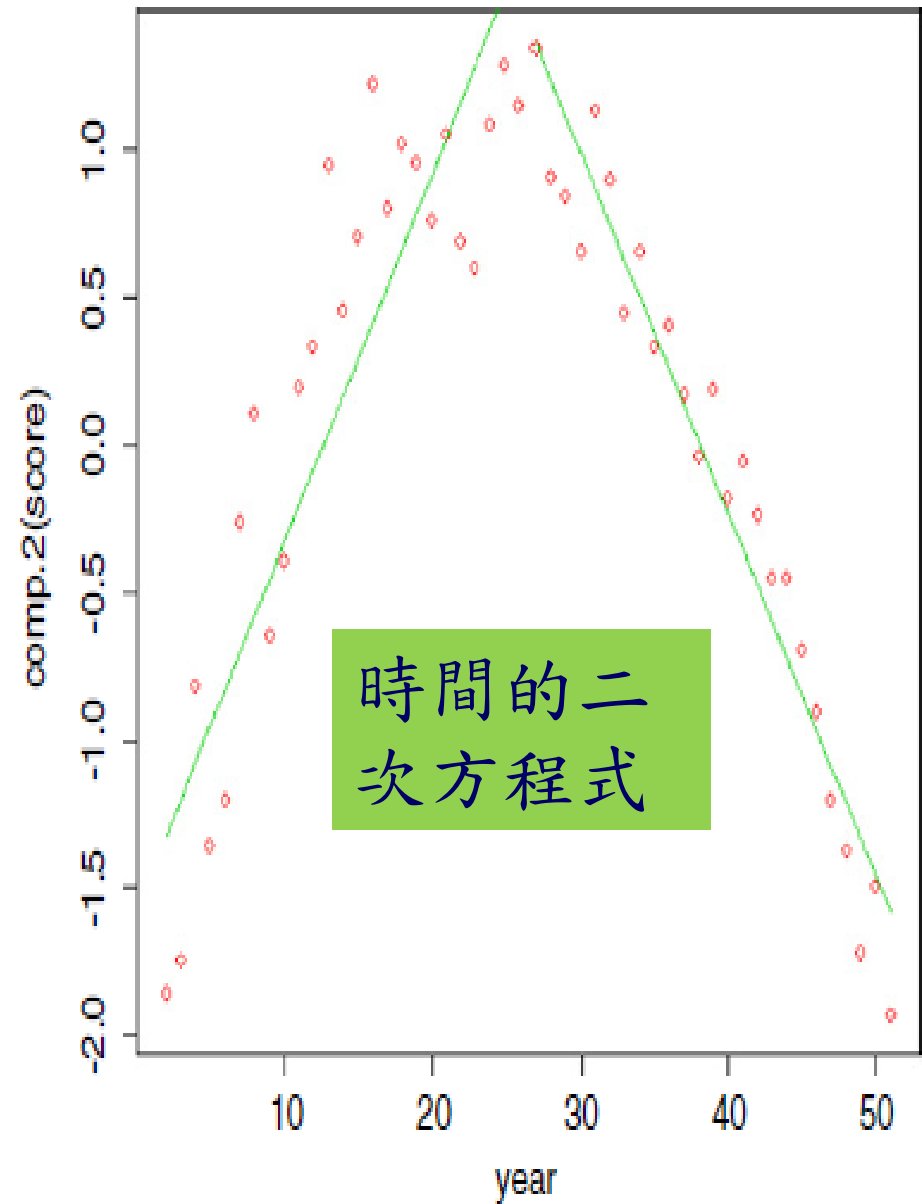


# 日本女性死亡率主成分分析結果(1951~2006年)

Japan (Female)

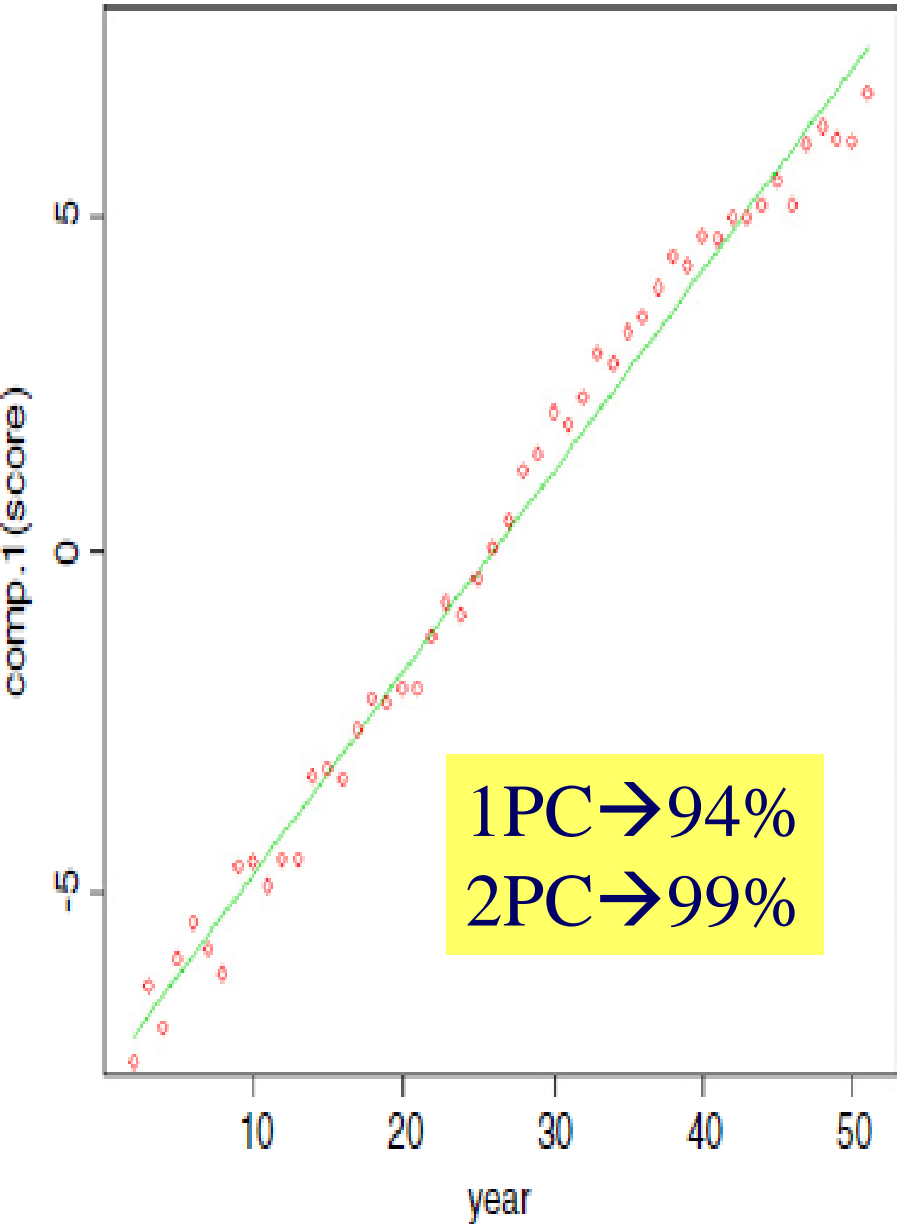


Japan (Female)

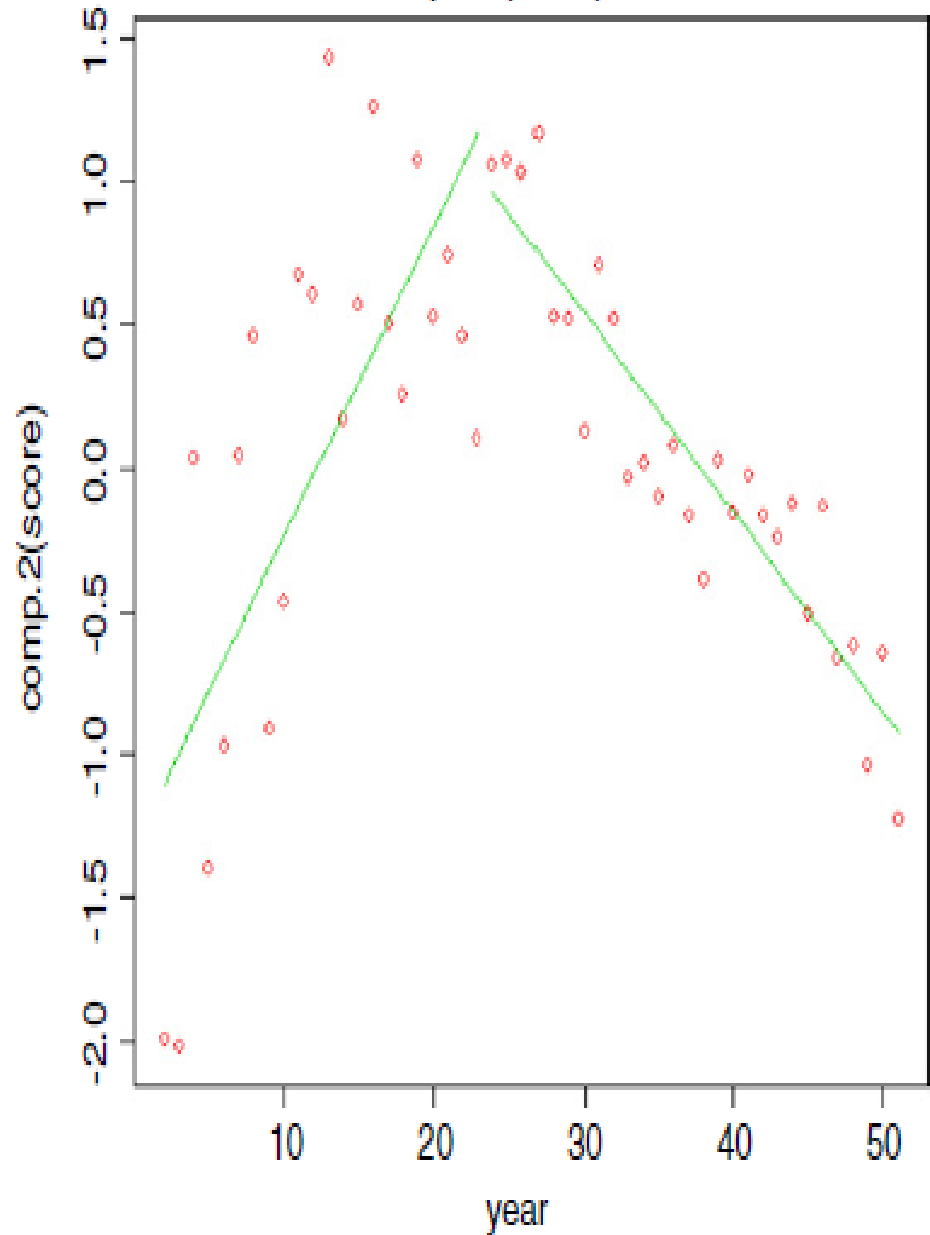


# 日本男性死亡率主成分分析結果(1951~2006年)

Japan (Male)



Japan (Male)





**Table C.3**

Parameter estimates of the first and second PCs for Japanese men.

	$a_i$	$b_i$	$R$ -squared	Adjusted $R$ -squared
First PC	-7.733756	0.298617	0.988	0.9878
Second PC-1	-1.314420	0.107650	0.5296	0.5072
Second PC-2	2.632501	-0.069705	0.8494	0.8436

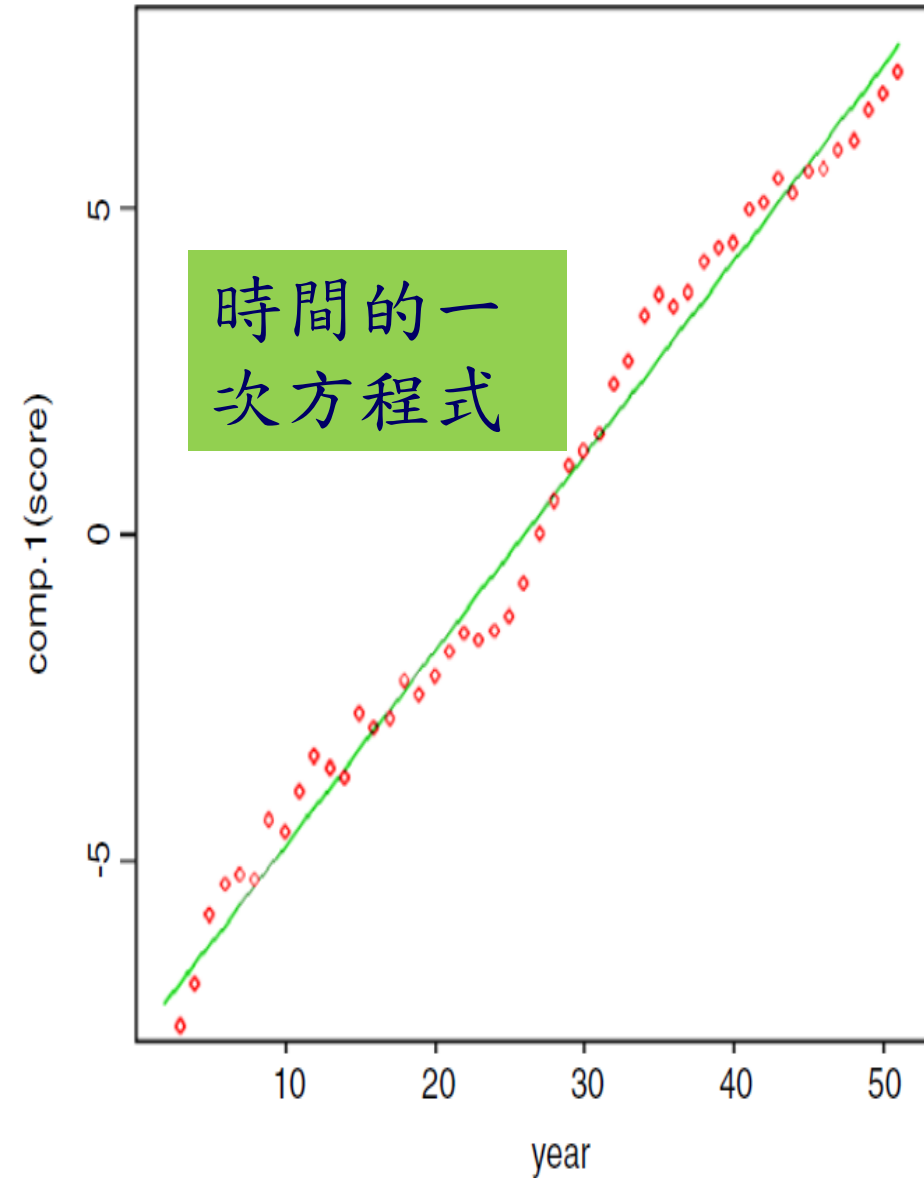
**Table C.4**

Parameter estimates of the first and second PCs for Japanese women.

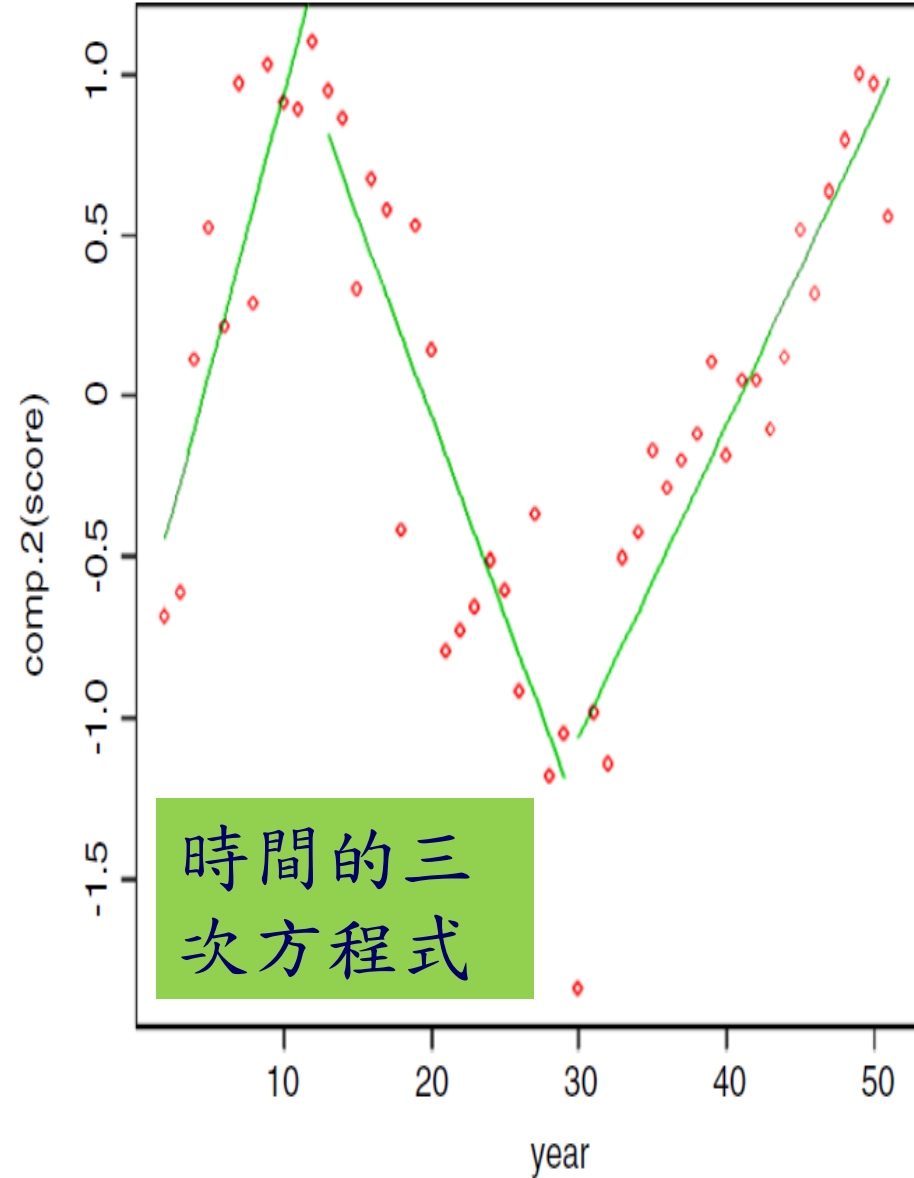
	$a_i$	$b_i$	$R$ -squared	Adjusted $R$ -squared
First PC	-7.739437	0.298702	0.9912	0.991
Second PC-1	-1.56803	0.12414	0.7988	0.7904
Second PC-2	4.654333	-0.122217	0.9444	0.9419

# 加拿大女性死亡率主成分分析結果(1951~2006年)

Canada (Female)

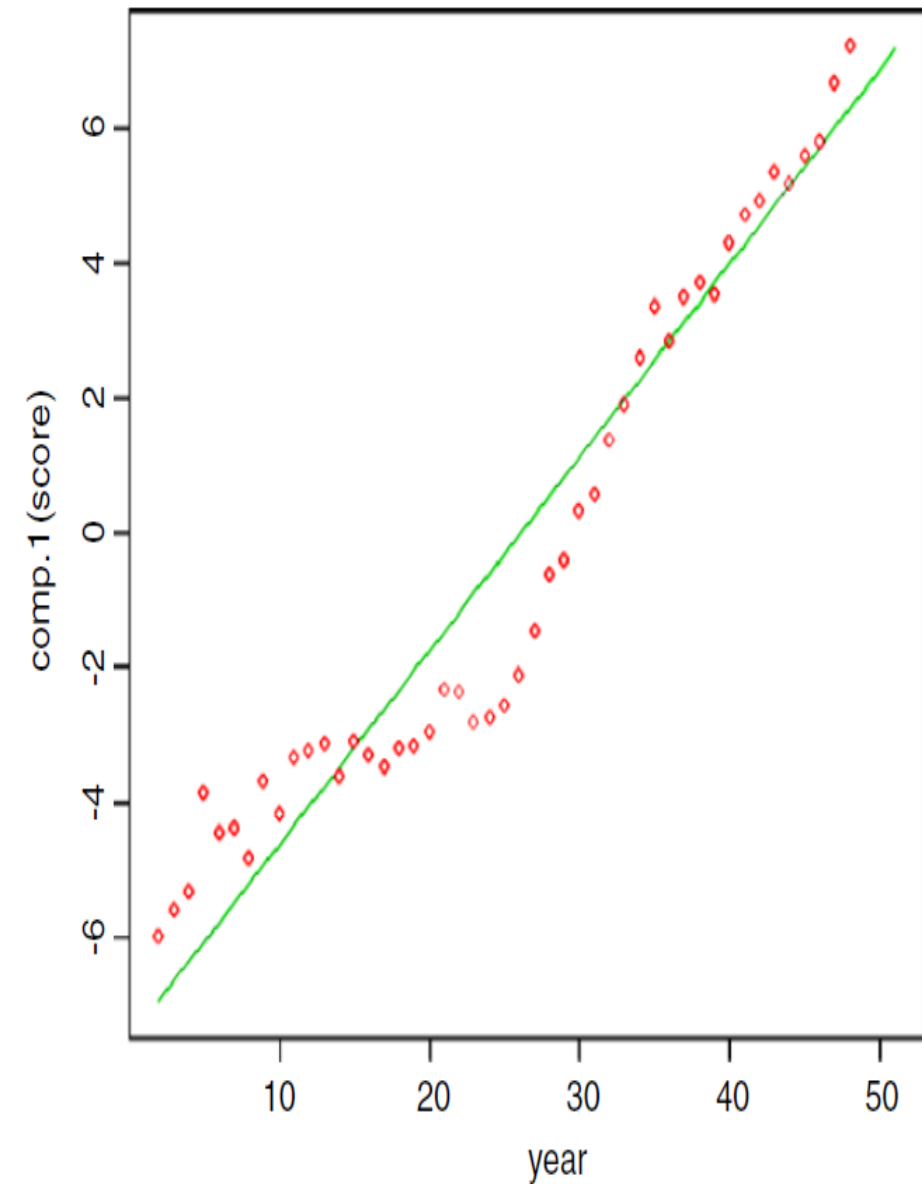


Canada (Female)

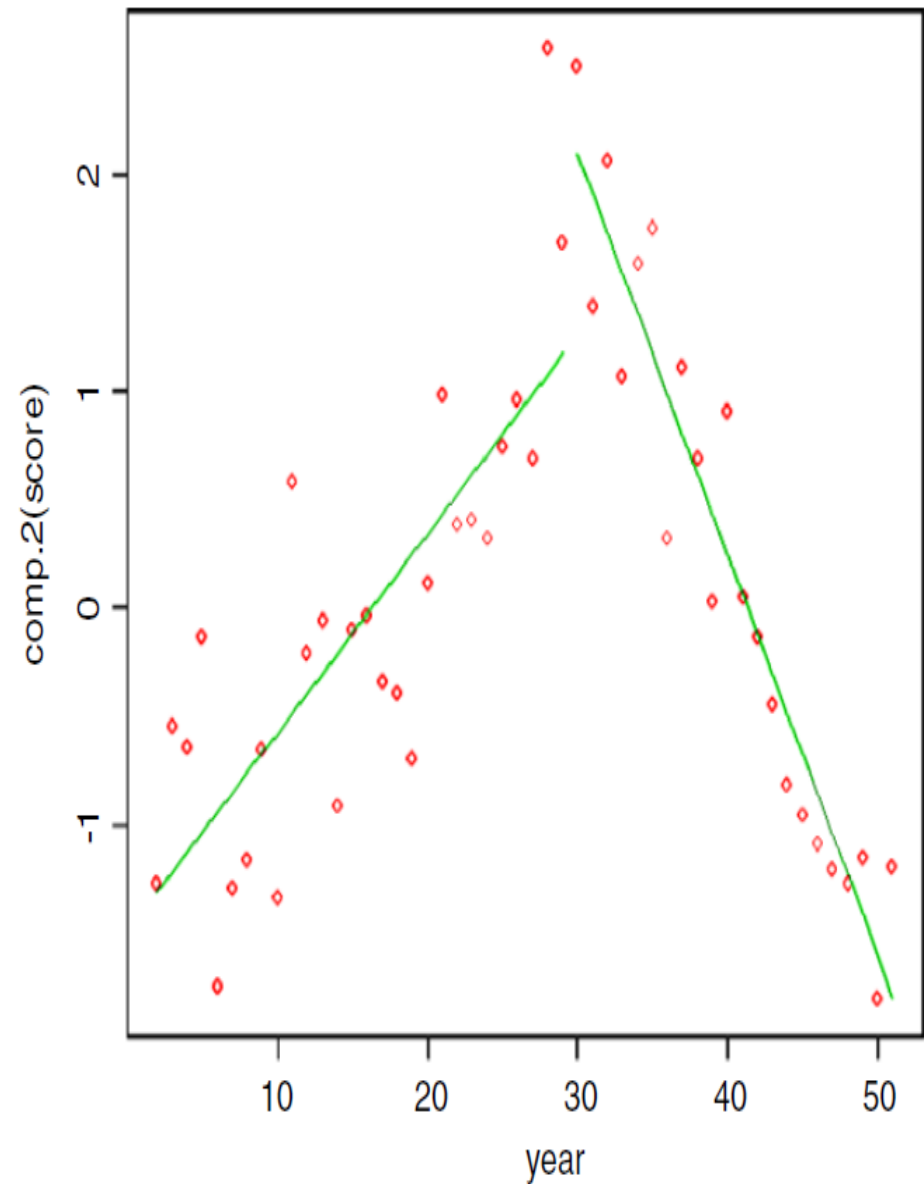


# 加拿大男性死亡率主成分分析結果(1951~2006年)

Canada (Male)



Canada (Male)



**Table C.7**

Parameter estimates of the first and second PCs for Canadian men.

	$a_i$	$b_i$	$R$ -squared	Adjusted $R$ -squared
First PC	-7.50734	0.28771	0.937	0.9357
Second PC-1	-1.49487	0.09193	0.6454	0.6322
Second PC-2	7.64473	-0.18498	0.9089	0.9044

**Table C.8**

Parameter estimates of the first and second PCs for Canadian women.

	$a_i$	$b_i$	$R$ -squared	Adjusted $R$ -squared
First PC	-7.744508	0.298879	0.9852	0.9849
Second PC-1	-0.78546	0.17244	0.8132	0.7945
Second PC-2	2.42407	-0.12435	0.776	0.7611
Second PC-3	-3.976459	0.097241	0.84	0.832

# 線性估計法

- 上述的參數模型中，大多藉由函數轉換使其成為線性模型，便於估計參數值。

→ 如果  $\underline{t} = \underline{t}(\underline{\phi})$  是未知參數向量  $\underline{\phi}$  的函數，如果參數求解是使適度性函數最小化：

$$F(\hat{\underline{\phi}}) = (\underline{u} - \underline{t}(\hat{\underline{\phi}}))' W (\underline{u} - \underline{t}(\hat{\underline{\phi}}))$$

微分後可得  $A'W(\underline{u} - A\hat{\underline{\phi}} - \underline{a}) = 0$  若  $\underline{t}(\underline{\phi}) = A\underline{\phi} + \underline{a}$

亦即

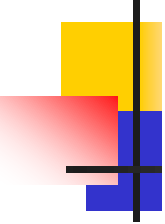
$$\begin{cases} \hat{\underline{\phi}} = (A'WA)^{-1} A'W(\underline{u} - \underline{a}) \\ \hat{\underline{y}} = A(A'WA)^{-1} A'W(\underline{u} - \underline{a}) + \underline{a} \end{cases}$$



# 非線性估計法

---

- 如果無法轉換成線性模型，參數估計將較為複雜，一般會透過遞迴(Iteration)求取參數估計值，在參數估計值變化低於某個預設門檻(Threshold)設為達到收斂。
- 但收斂值多半與起始點(Initial Value)有關，類似極值求取(Optimization)時遭遇的問題，即使理論上較佳，實用上的優勢未必顯著(e.g., Gompertz Law)。

- 
- 另一類型的死亡率參數公式則透過 $(q_x/p_x)$  比值，例如：

$$\frac{q_x}{p_x} = A + Hx + BC^x = f(x);$$

Heligman and Pollard(1980)或是  $q_x = \frac{f(x)}{1 + f(x)}$ 。

推廣  $f(x)$  的類型可得

$$f(x) = GM_{\alpha}^{r,s}(x) = \sum_{i=1}^r \alpha_i x^i + \exp\left\{ \sum_{i=r+1}^{r+s} \alpha_i x^{i-r-1} \right\}$$

⇒ 相當於  $f(x) = \text{多項式} + \exp(\text{多項式})$

→ 這類方法尚無較統一的參數選擇  $r$  與  $s$ 。



- 縮減因子(Reduction Factor ; RF )

→ 1924年時英國的Institute of Actuaries已經採取與時間有關的死亡率預測模型：

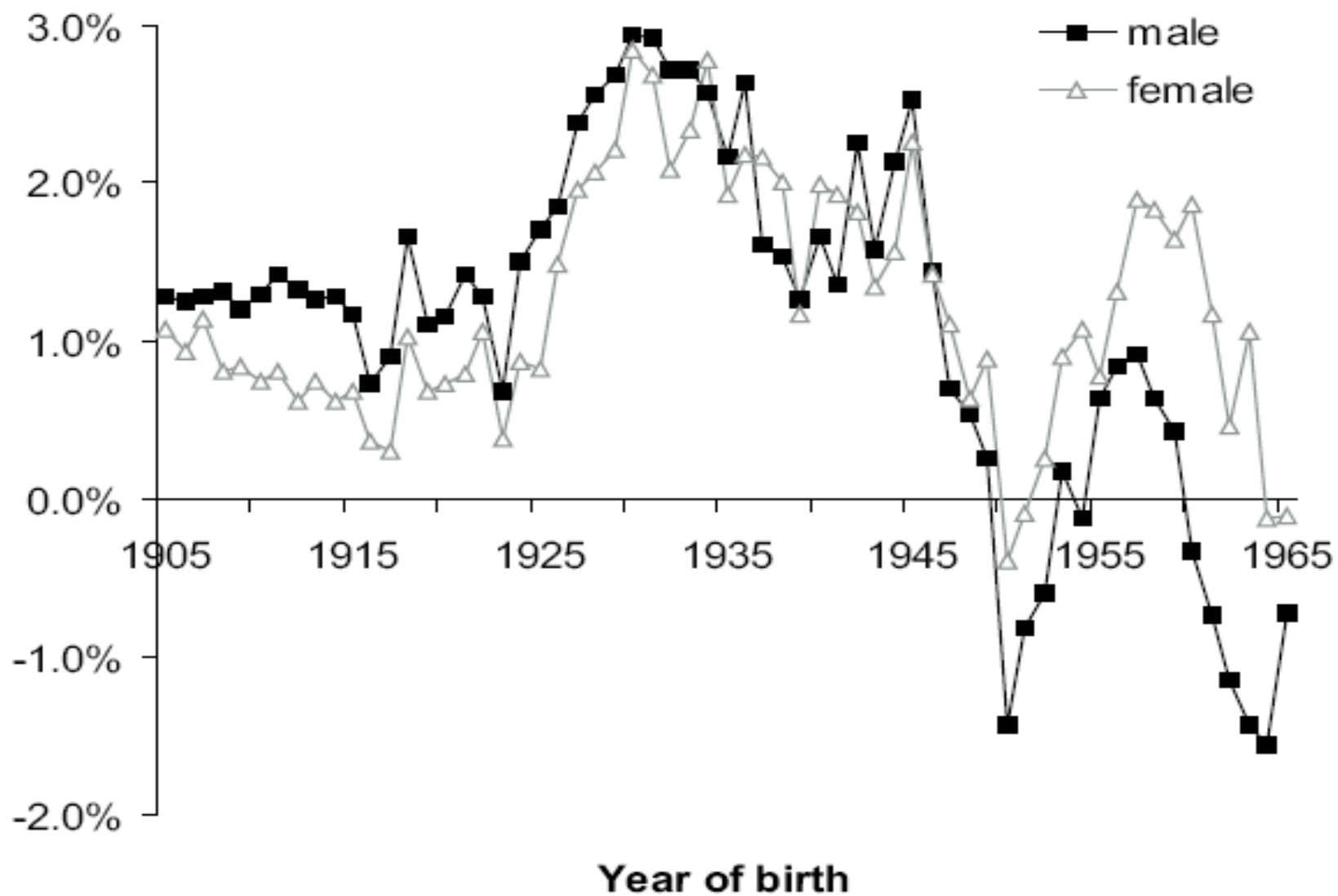
$$q_x(t) = a(x) + b(x)c(x)^t$$

這與美國SOA模型很類似。

→ 之後英國連續死亡率調查局(Continuous Mortality Investigation Bureau)繼續提出死亡率改善模型(1999)：

$$\frac{q_{x,t}}{q_{x,0}} = RF(x,t) = \alpha(x) + [1 - \alpha(x)][1 - f(x)]^{\frac{t}{20}}$$

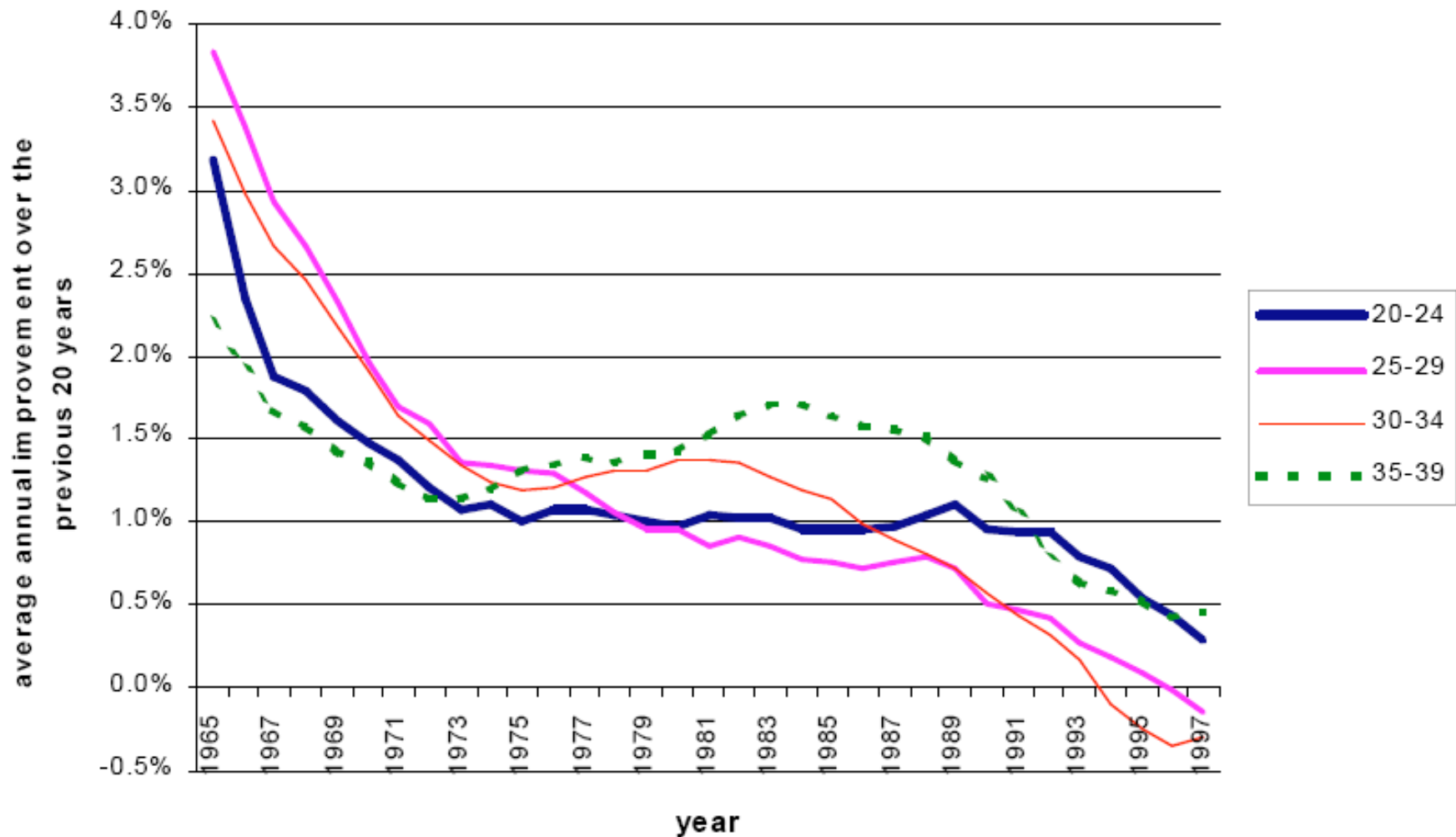


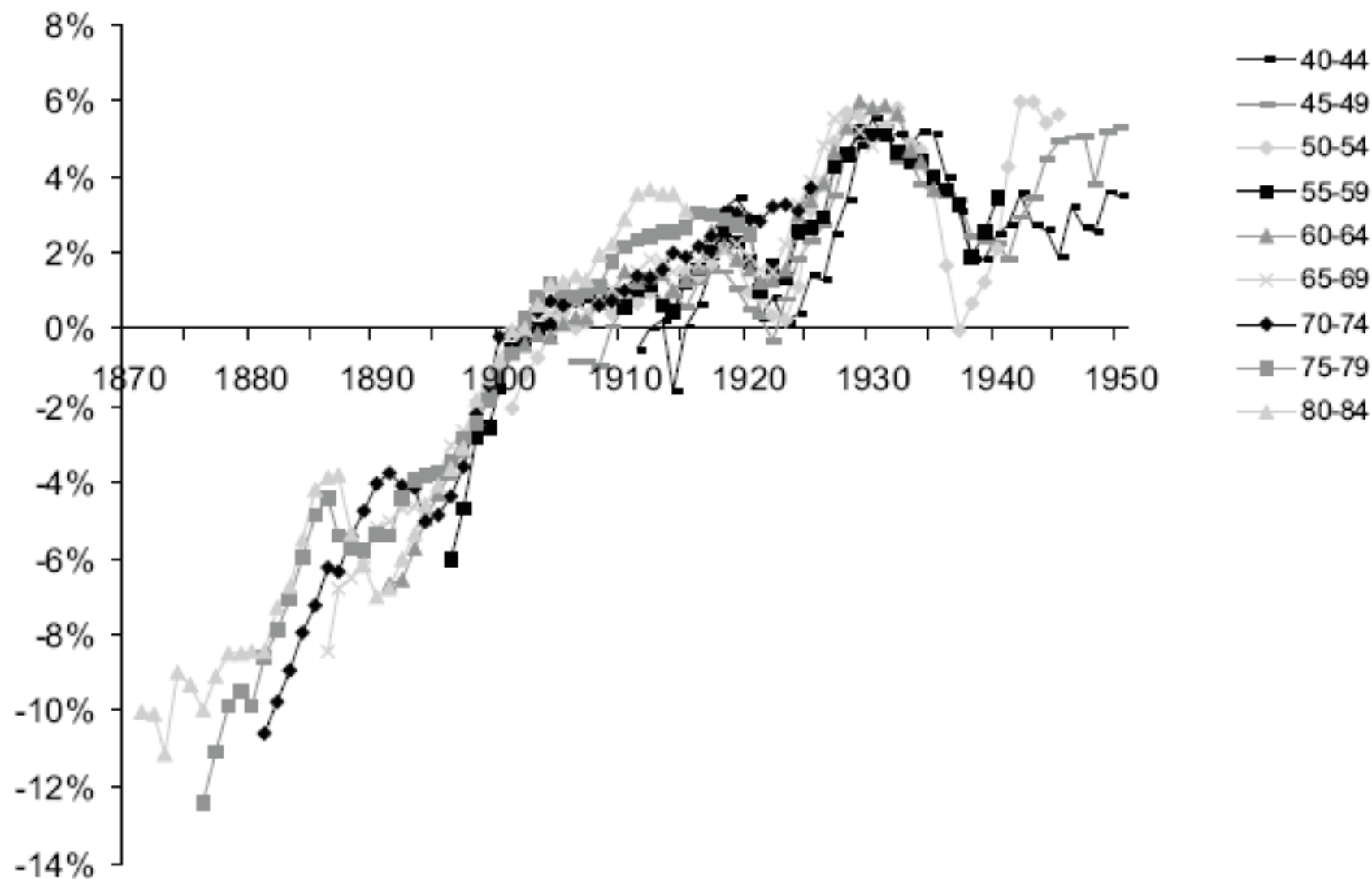


Data source: GAD

Figure 2. Average rate of mortality improvement by year of birth, smoothed using five year rolling averages, and gender over the period 1961-2002 for the population of England and Wales

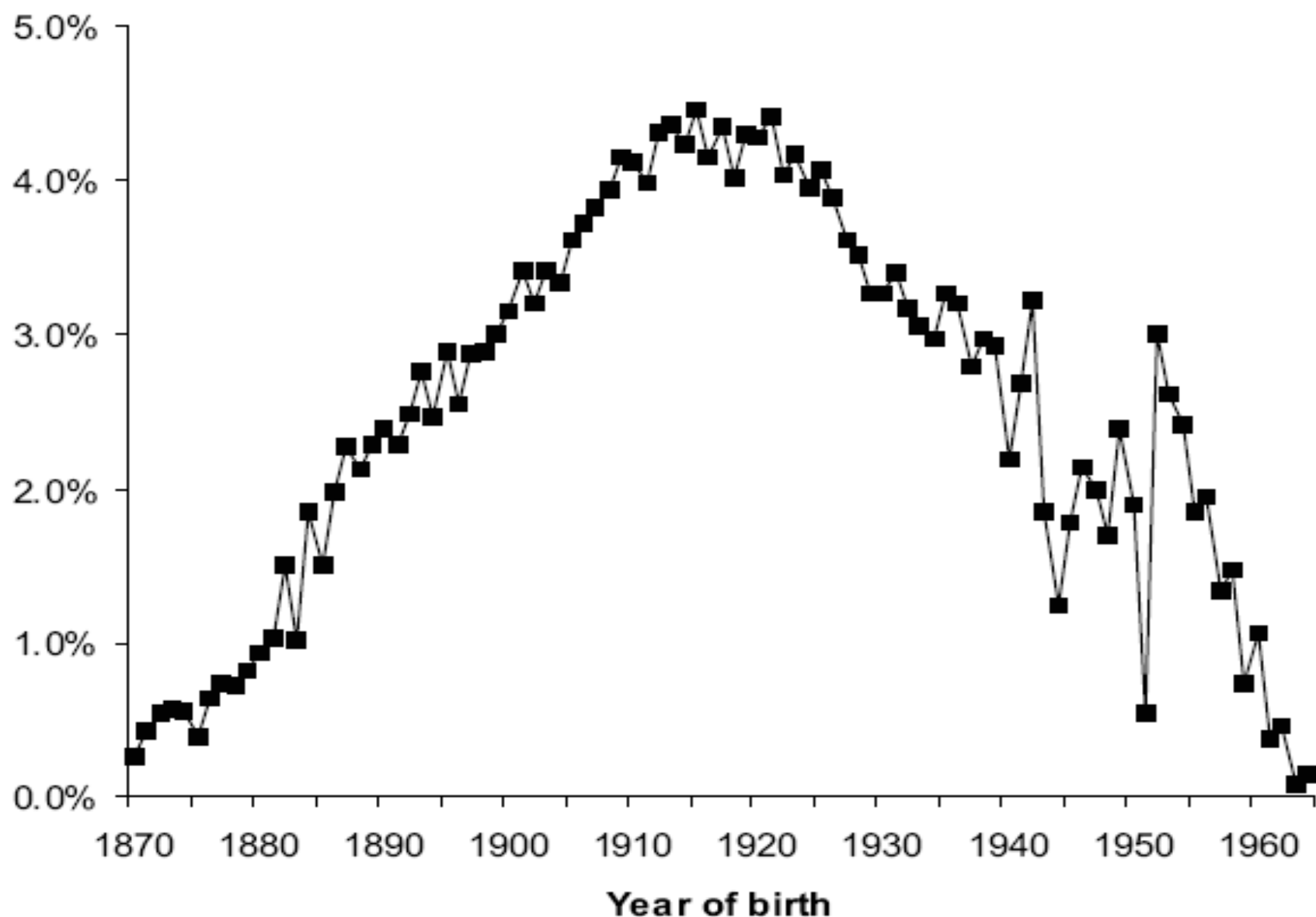
### Mortality improvement for men in their 20s and 30s over the period 1946 to 1997 - England & Wales population





Data source: O.N.S. (2001)

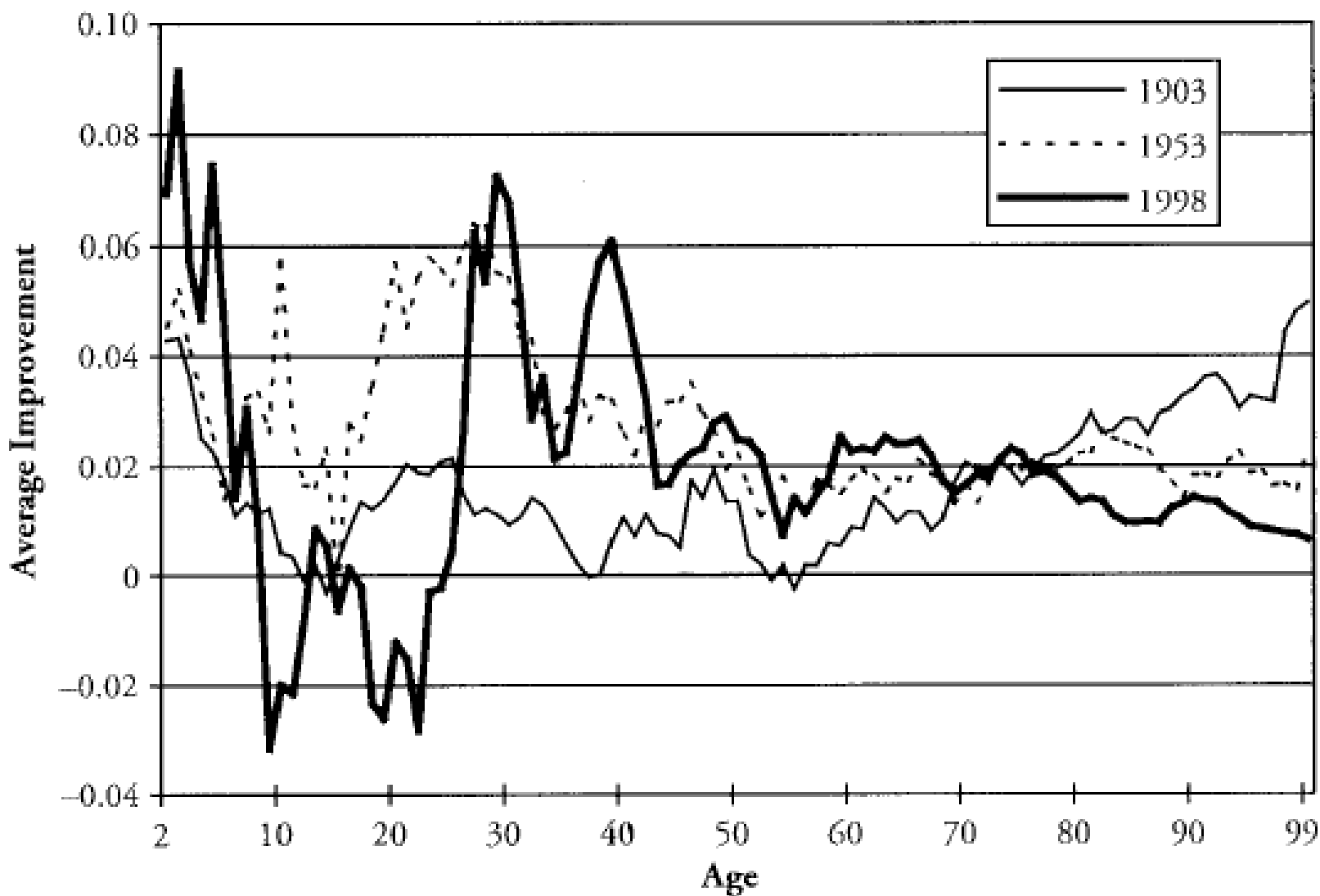
Figure 9. Average annual improvement in central mortality rates from lung cancer by year of birth and five year age group for males in England and Wales



Data source: [www.mortality.org](http://www.mortality.org)

Figure 22. Average annual rate of mortality improvement in the Japanese population by year of birth (nine-year rolling averages) for females

Figure 1. Five-Year Moving Average of the Improvement in Mortality at Ages 2–99 for Sweden in 1903, 1953, and 1998



The mortality improvement model adopted by the CMI is

$$RF(x, t) = \alpha(x) + [1 - \alpha(x)][1 - f(x)]^{t/20},$$

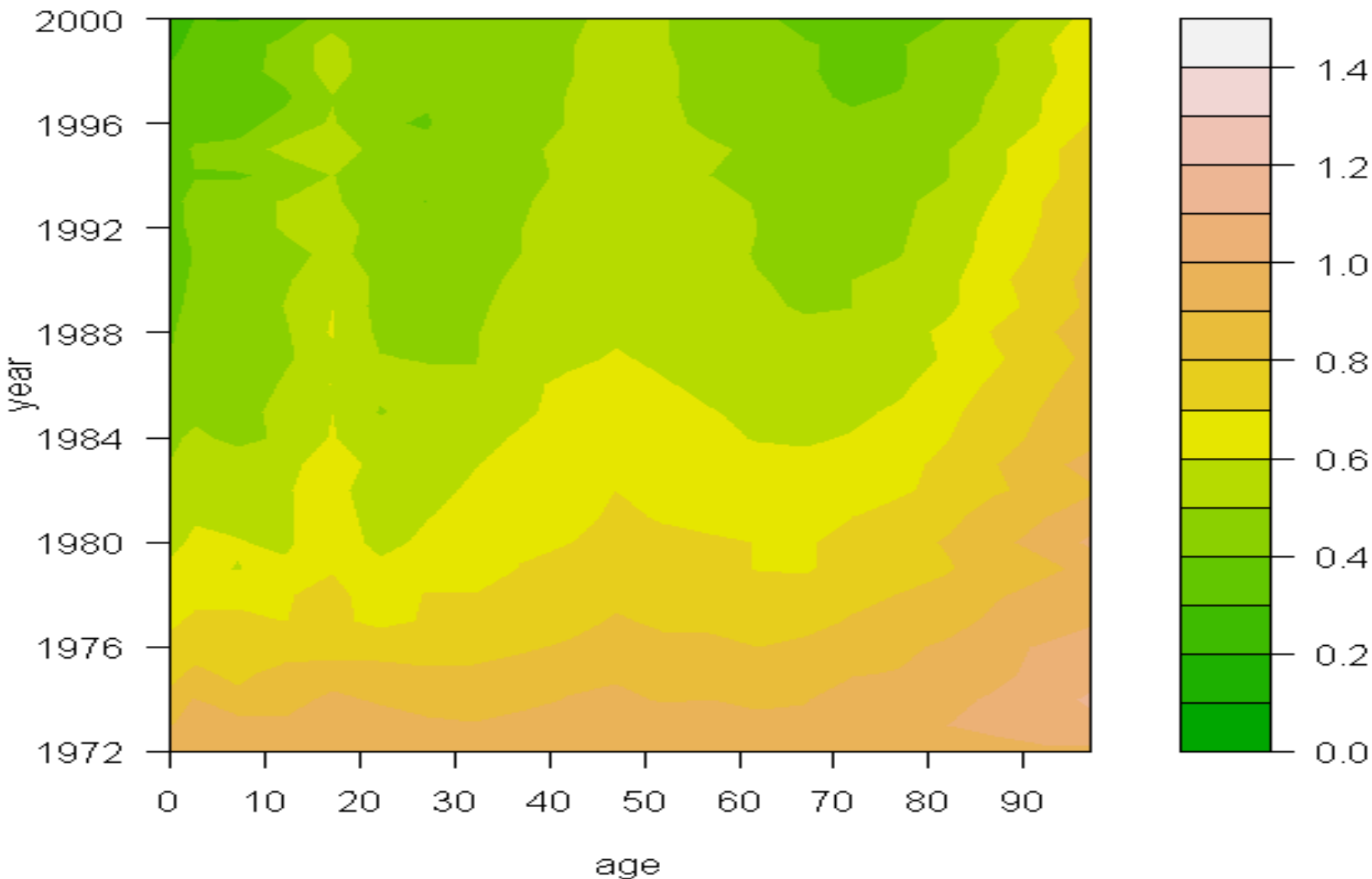
where

$$\alpha(x) = \begin{cases} c, & x < 60, \\ 1 + \frac{(1 - c)(x - 110)}{50}, & 60 \leq x \leq 110, \\ 1, & x > 110, \end{cases}$$
$$f(x) = \begin{cases} p, & x < 60, \\ \frac{[(110 - x)p + (x - 60)q]}{50}, & 60 \leq x \leq 110, \\ q, & x > 110, \end{cases}$$

with  $c = 0.13$ ,  $p = 0.55$  and  $q = 0.29$ .

Redcution Factor (JP female1972~2000)

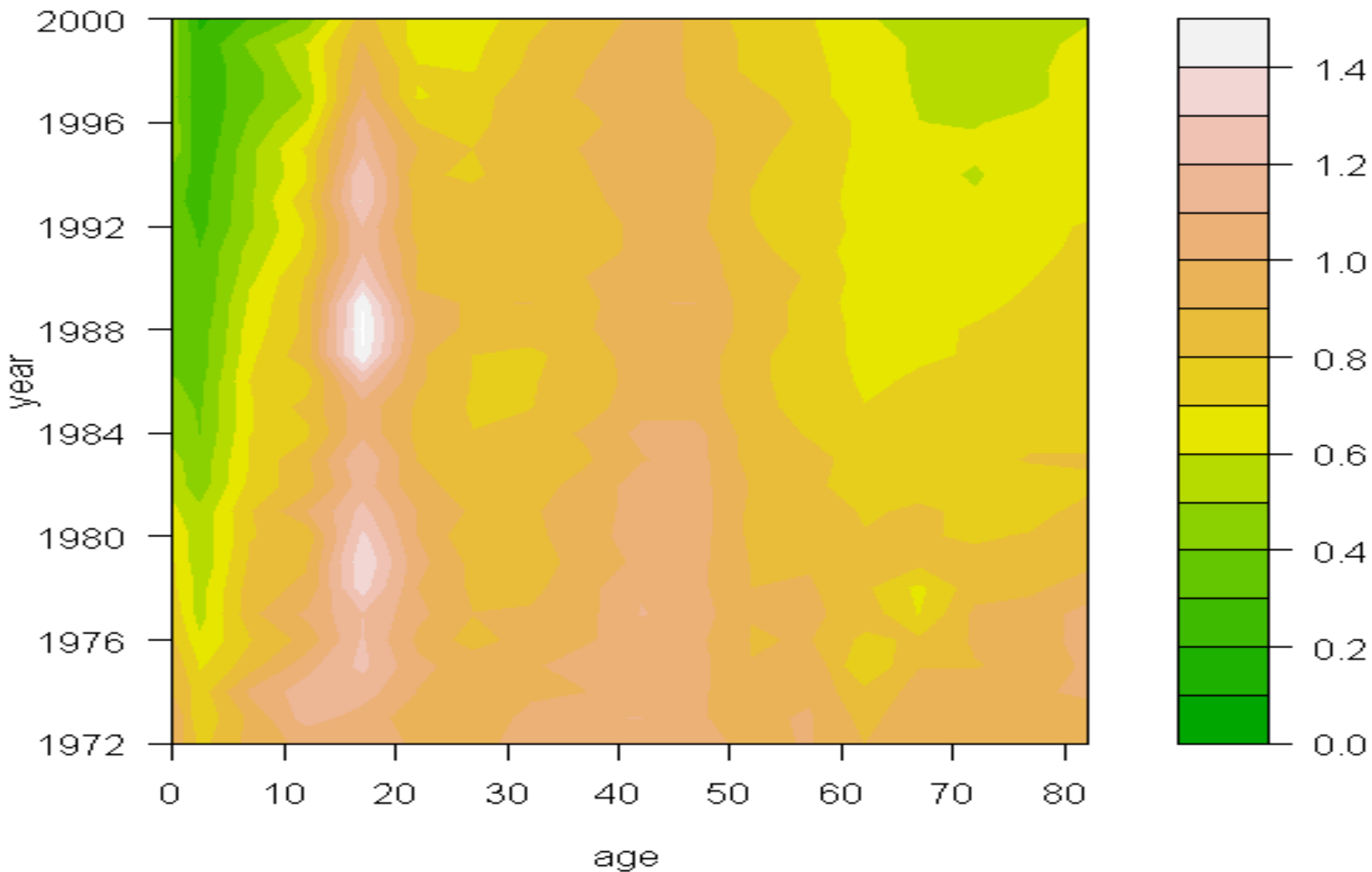
Reduction



日本女性1970~2000年死亡率下降趨勢

Redcution Factor (Taiwan male1972~2000)

Reduction



臺灣男性1970~2000年死亡率下降趨勢



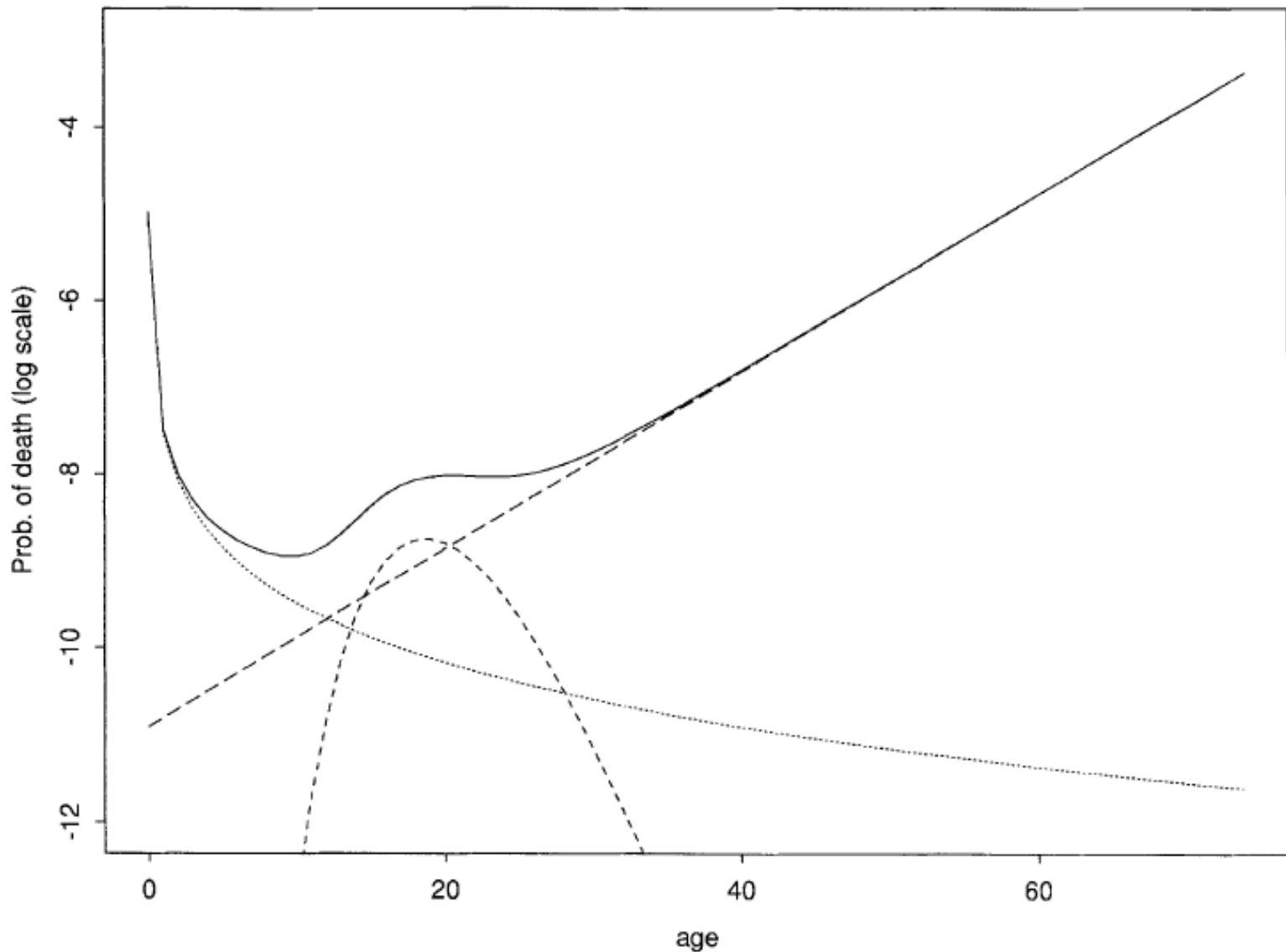


## 參數綜合法

- Heligman and Pollard(1980)首先提出用幾個參數的合成去估計澳洲的死亡率曲線：

$$\frac{q_x}{p_x} = A^{(x+B)^c} + D \cdot \exp\{-E(\log x - \log F)^2\} + GH^x$$

- $A^{(x+B)^c}$  代表幼兒時期的遞減死亡率；
- $D \cdot \exp\{-E(\log x - \log F)^2\}$  代表二十歲左右因意外死亡的隆起；
- $GH^x$  代表成年時期死亡率指數遞增。



Heligman-Pollard 模型中各年齡組的死亡曲線  
 (A,B,C,D,E,F,G,H) =  
 (.000544,.0170,.101,.000158,10.72,18.67,.0000183,1.11)



## 參數綜合法(HP)的缺點

---

- 雖然HP模型的詮釋先當合理，其合成曲線也頗為符合實際死亡率曲線，但因為需要同時找出8個參數的估計值，加上估計的求解必須透過非線性(Non-linear)，不能保證求出的解是局部最佳(Local minimum)或是整體最佳(Global minimum)。
- 聯合國於1990年代提供使用HP模型的估計軟體(MORTPAK)，但實證上使用不頻繁。



## 參數綜合法(續)

- Carriere(1992)延伸HP參數綜合法，提出以下的參數公式：

$$s(x) = \sum_{k=1}^m \psi_k s_k(x)$$

其中  $\psi_k$  ( $k = 1, \dots, m$ ) 的解釋為死亡於第  $k$  個死因的機率， $s_k(x)$  為對應於第  $k$  個死因的生存函數。Carriere的生存函數選擇有 Gompertz及Inverse Gompertz (以上為成年時期死亡率)、Weibull(幼兒時期死亡率)、Inverse Weibull(青少年時期死亡率)。

(1) Gompertz :  $s(x) = \exp\left\{e^{-\frac{m}{\sigma}} - e^{-\frac{(x-m)}{\sigma}}\right\}$

(2) Inverse Gompertz :  $s(x) = (1 - \exp\{-e^{-\frac{(x-m)}{\sigma}}\}) / (1 - \exp\{-e^{-\frac{m}{\sigma}}\})$

(3) Weibull :  $s(x) = \exp\left\{-\left(\frac{x}{m}\right)^{\frac{m}{\sigma}}\right\}$

(4) Inverse Weibull :  $s(x) = 1 - \exp\left\{-\left(\frac{x}{m}\right)^{-\frac{m}{\sigma}}\right\}$

其中各參數模型的  $m$  代表位置 (Location Parameter)， $\sigma$  代表曲線散佈程度 (Dispersion Parameter)，均具有人口學與統計學的實質意義。Carriere 證明他的模型與一般的多重脫退模型 (Multiple Decrement Model) 是一樣的。



# APC模型

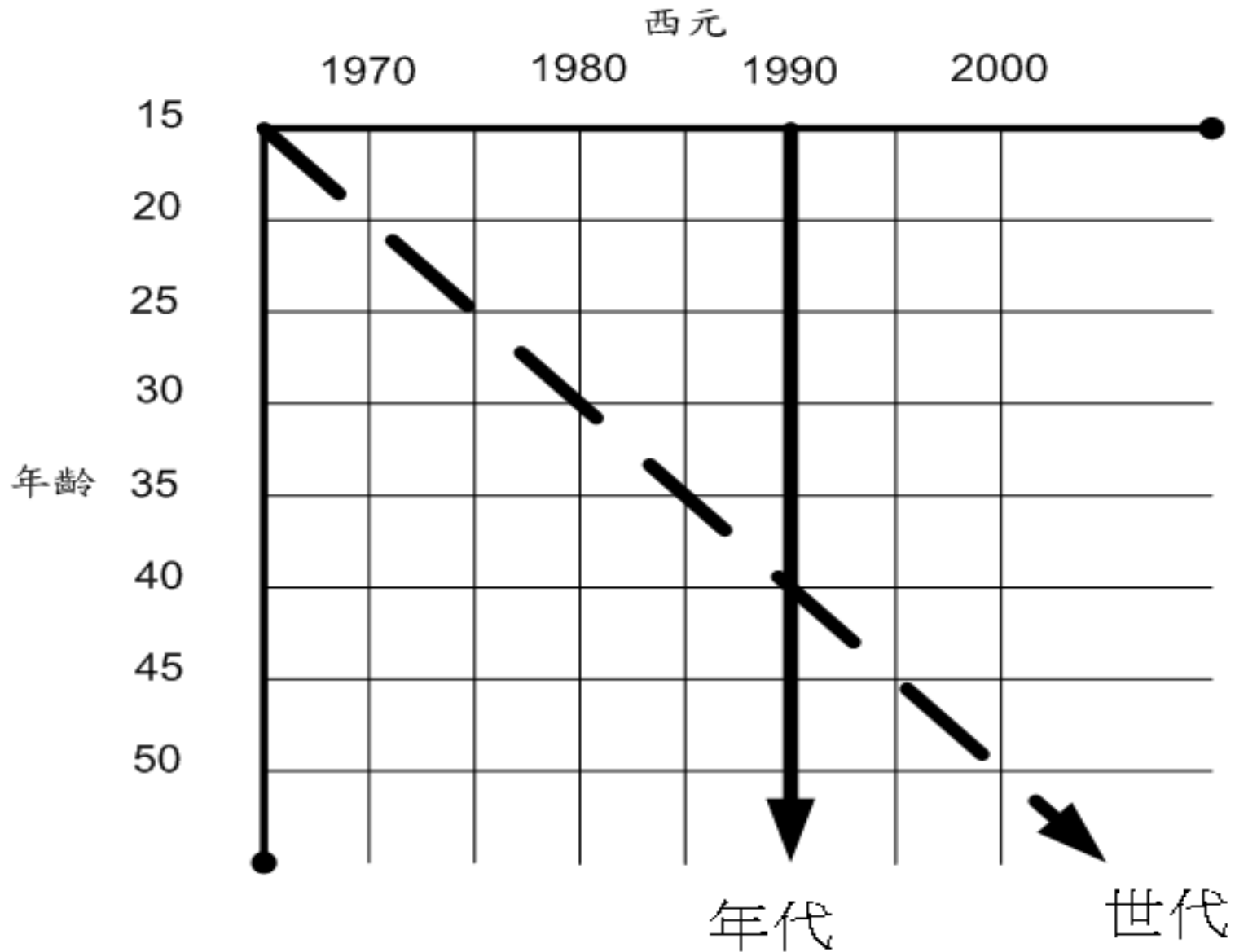
- Age-Period-Cohort模型源自於流行病學

假設第*i*個年齡在第*j*個年代的死亡人數  
( $O_{ij}$ ) 為卜瓦松分配，相對應的總人數  
( $N_{ij}$ ) 為非隨機之固定值

$$Y_{ij} = \log\left(\frac{O_{ij}}{N_{ij}}\right) = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ij}$$

$$i = 1, \dots, a; j = 1, \dots, p; k = 1, \dots, a + p - 1$$

# 年代與世代的差異



年齡		年代				世代	
		1961 (1)	1966 (2)	1971 (3)	1976 (4)		
0	(1)	1265	782	544	415	1976	(7)
5	(2)	120	92	72	57	1971	(6)
10	(3)	86	69	57	54	1966	(5)
15	(4)	142	127	108	132		
世代		1946 (1)	1951 (2)	1956 (3)	1961 (4)		

年代 - 年齡 = 世代





# APC模型的問題

---

- 年代 - 年齡 = 世代
  - 僅有二階差異唯一
  - 三者呈現共線性關係
  - 參數估計值將有無限多組
  - 甄別問題 (Identification Problem)

- 世代重疊 (Overlap)

1981-1985年：55-59歲死亡 → 1921-1930年出生  
60-64歲死亡 → 1916-1925年出生

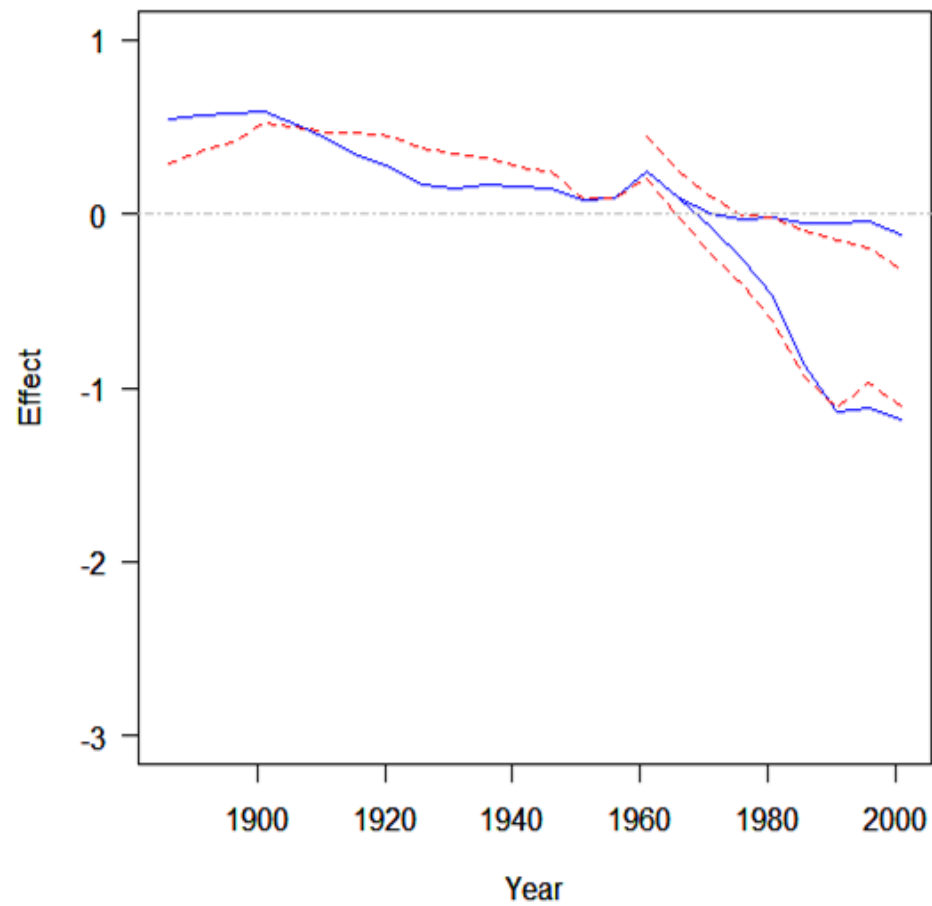
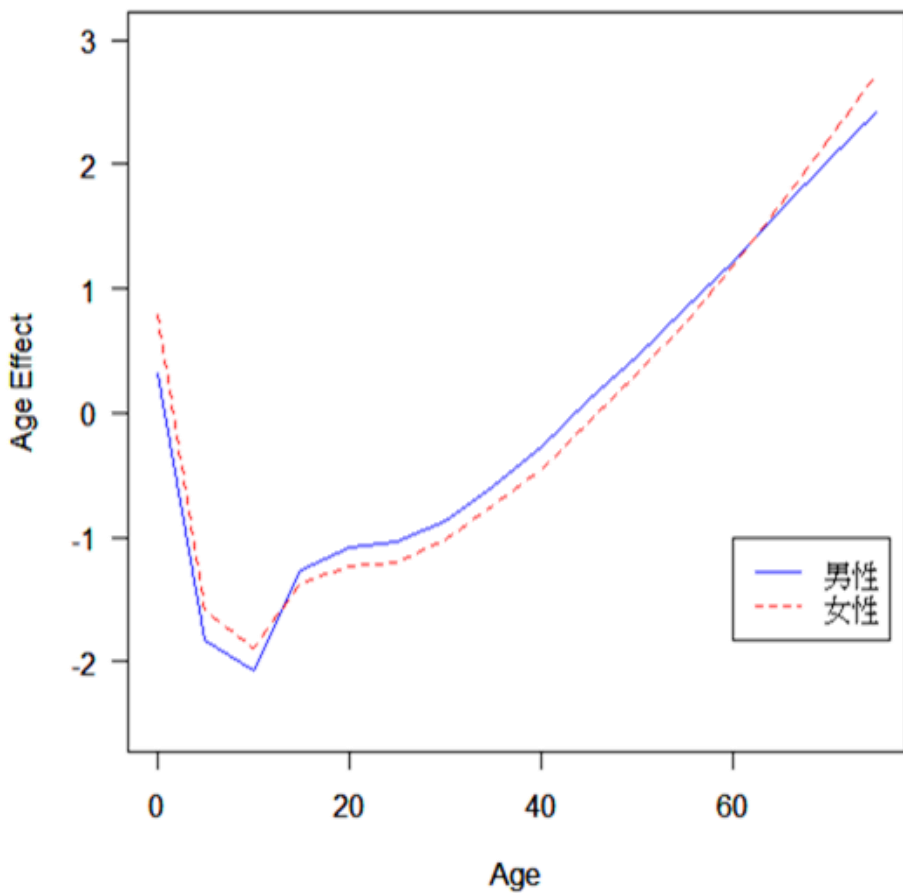


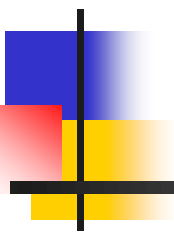
# 估計APC模型參數的方法

---

- 解決共線性
  - 增加限制式
  - 懲罰函數法
  - 無母數方法
  - 序列法

# APC模型 (本質估計量)





# 參數模型的比較實例： 臺灣婚姻別死亡率變化趨勢

---

Lee-Carter模型

VS.

Age-Period-Cohort模型



## 臺灣地區婚姻死亡率相關紀錄

年度 (西元年)	人口數 (五齡組)	死亡人數 (五齡組)
1973-1974	15~50+	無
1975-1991	15~50+	15~85+
1992-1993	15~50+	15~95+
1994-1997	15~100+	15~95+
1998-2008	15~100+	15~100+

資料來源：內政部統計處



# 資料年度與分析方法

---

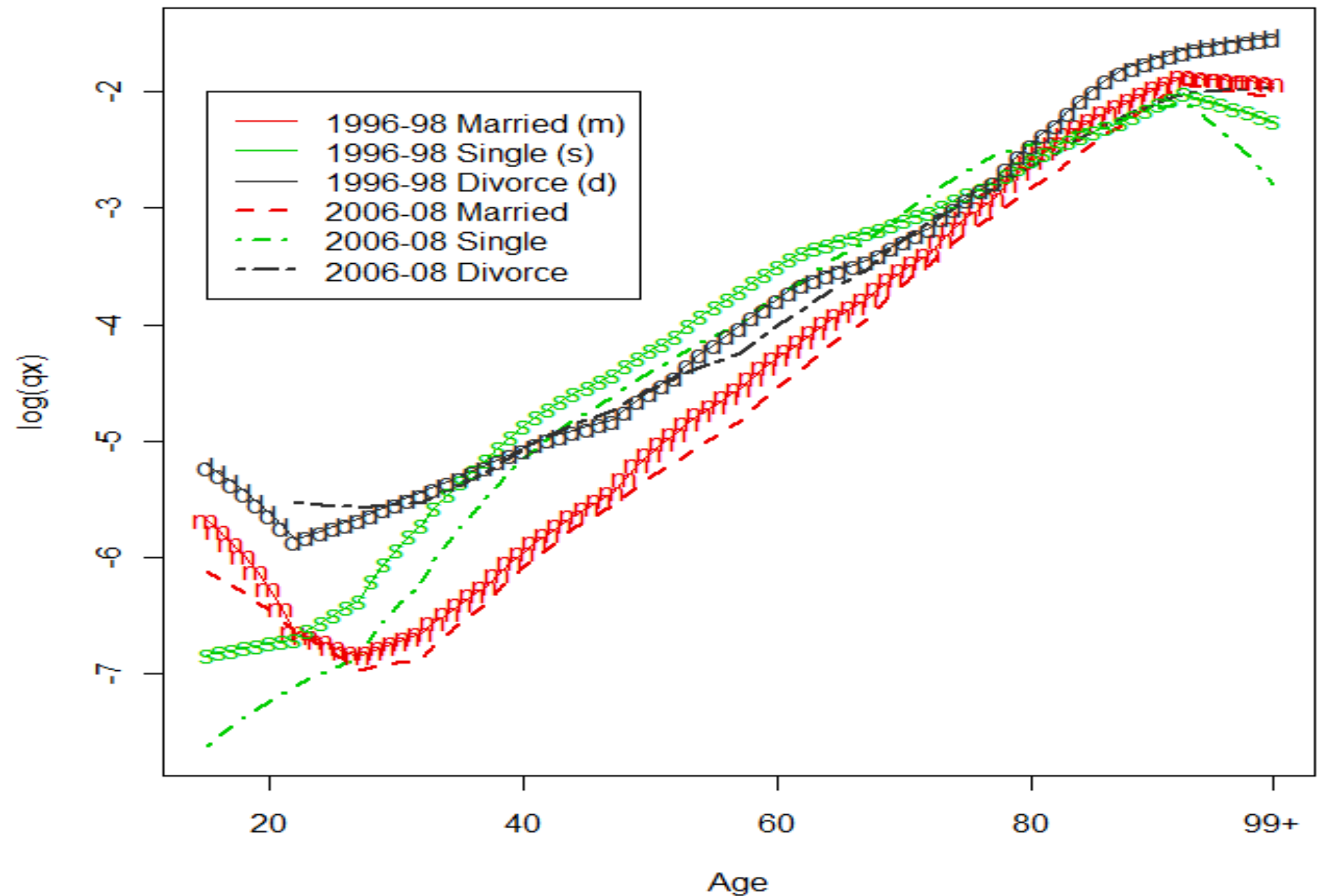
- 因為1994年之後才有較完整紀錄，分成兩部分討論：
  - 比較1996-1998年、2006-2008年的死亡率，以編算國民生命表的想法，比較三種婚姻狀況（已婚、未婚、離婚/喪偶）趨勢變化；
  - 較長時間(1994-2008年)較完整年齡層(15歲以上)資料，以Lee-Carter、APC模型估算不婚姻相關死亡率的改善趨勢。



## 十年間的變化比較

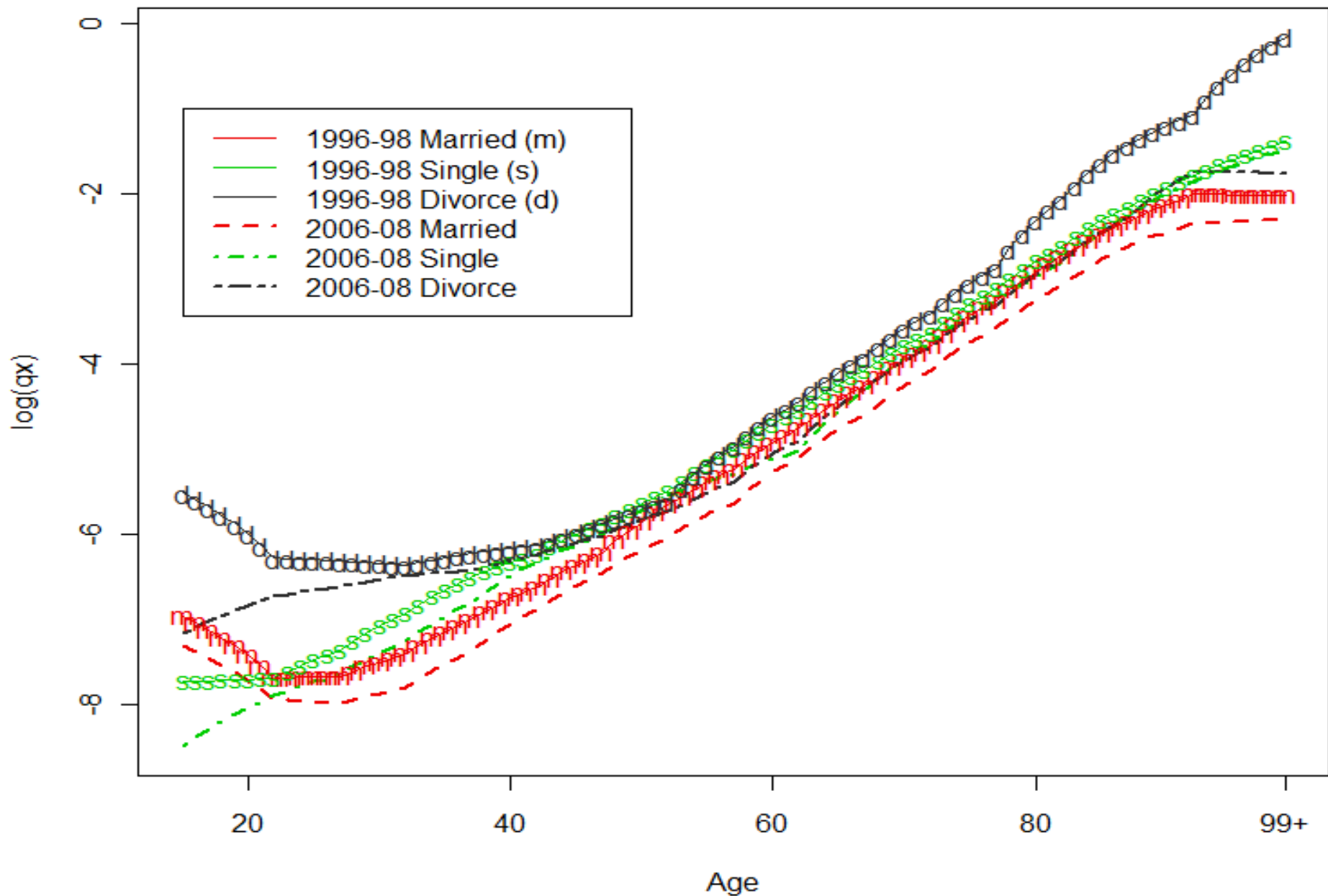
---

- 國民生命表每十年編算一次，以連續三年的死亡記錄為基礎。因為各年齡的人數及隨機誤差，死亡率曲線不見得變化和緩，因此有時需考慮修勻(Graduation)，調整死亡率使其符合某種共識(例如：平滑)。
- 可使用的修勻方法包括Moving Weighted Average、Whittaker、Spline、Kernel四種方法，詳細討論可參考先前講義。

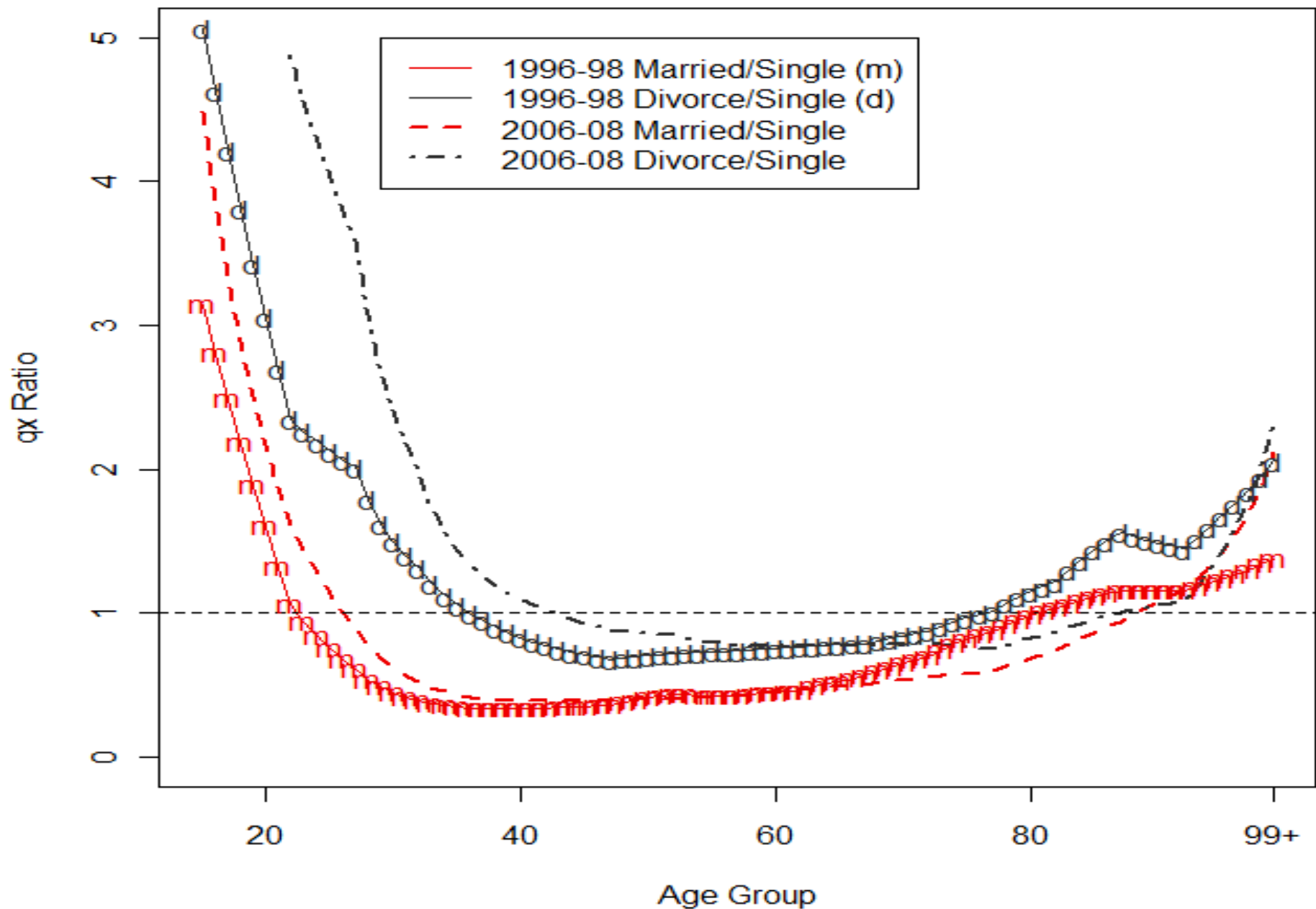


1996-98年及2006-08年男性婚姻狀況死亡率

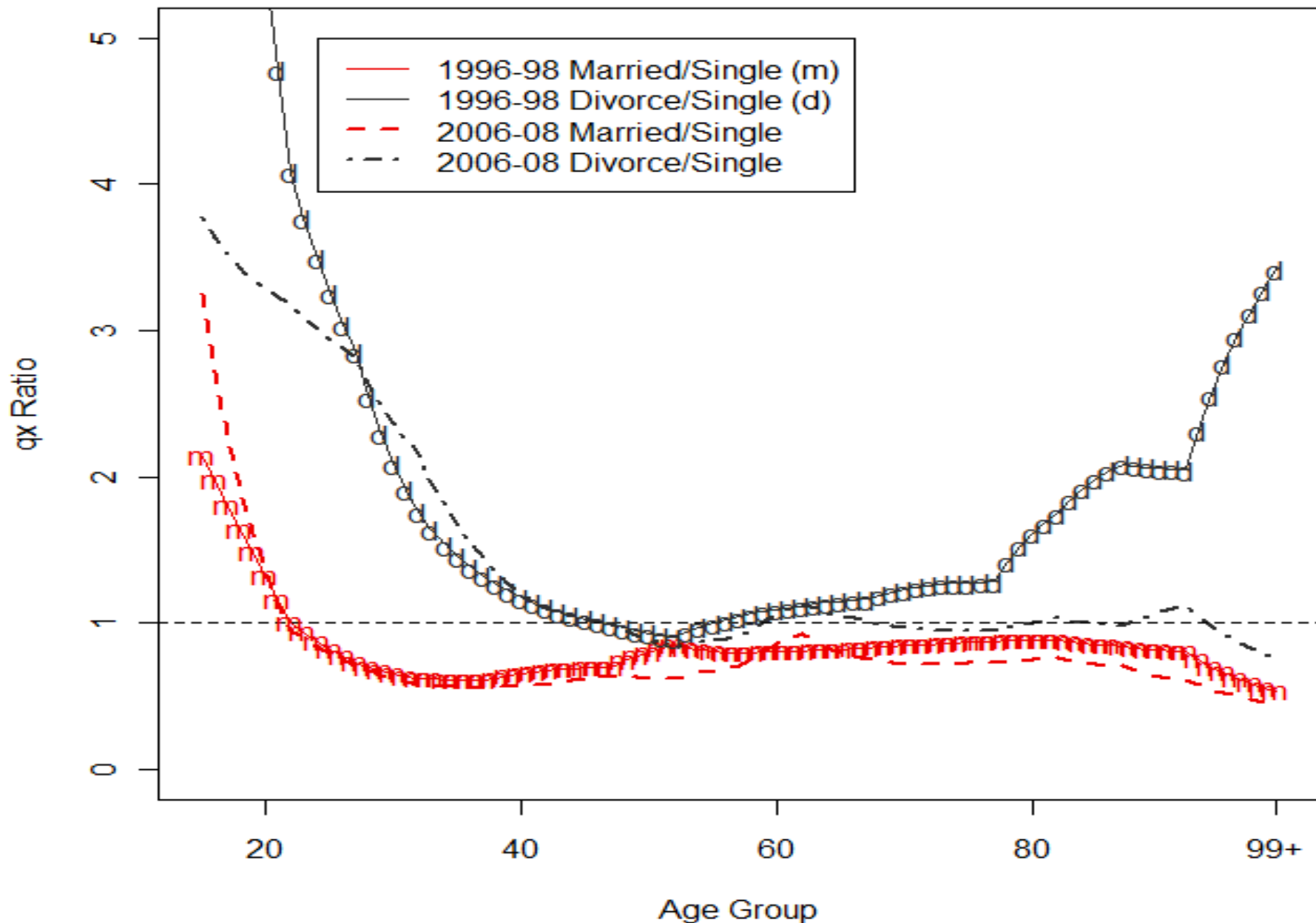




1996-98年及2006-08年女性婚姻狀況死亡率

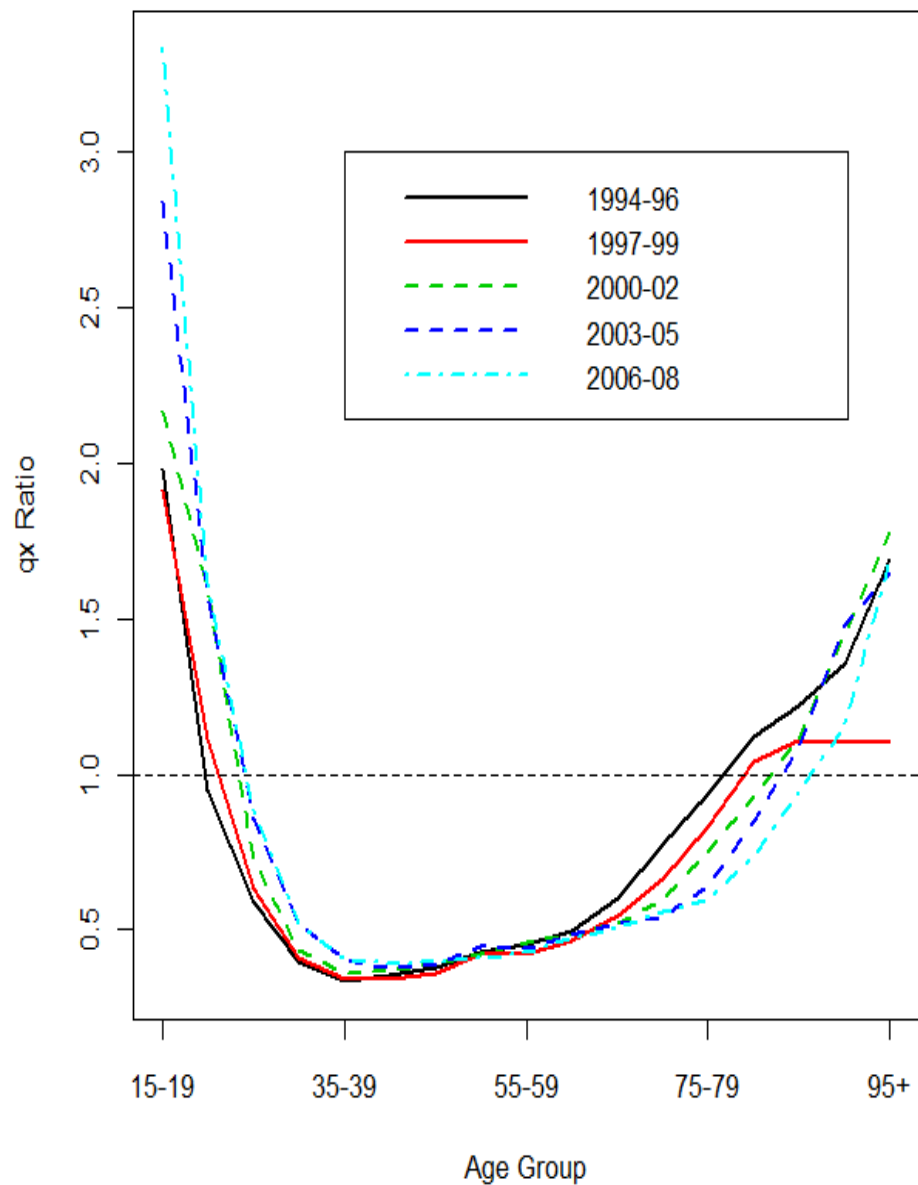


1996-98年及2006-08年男性婚姻死亡率比值(未婚分母)

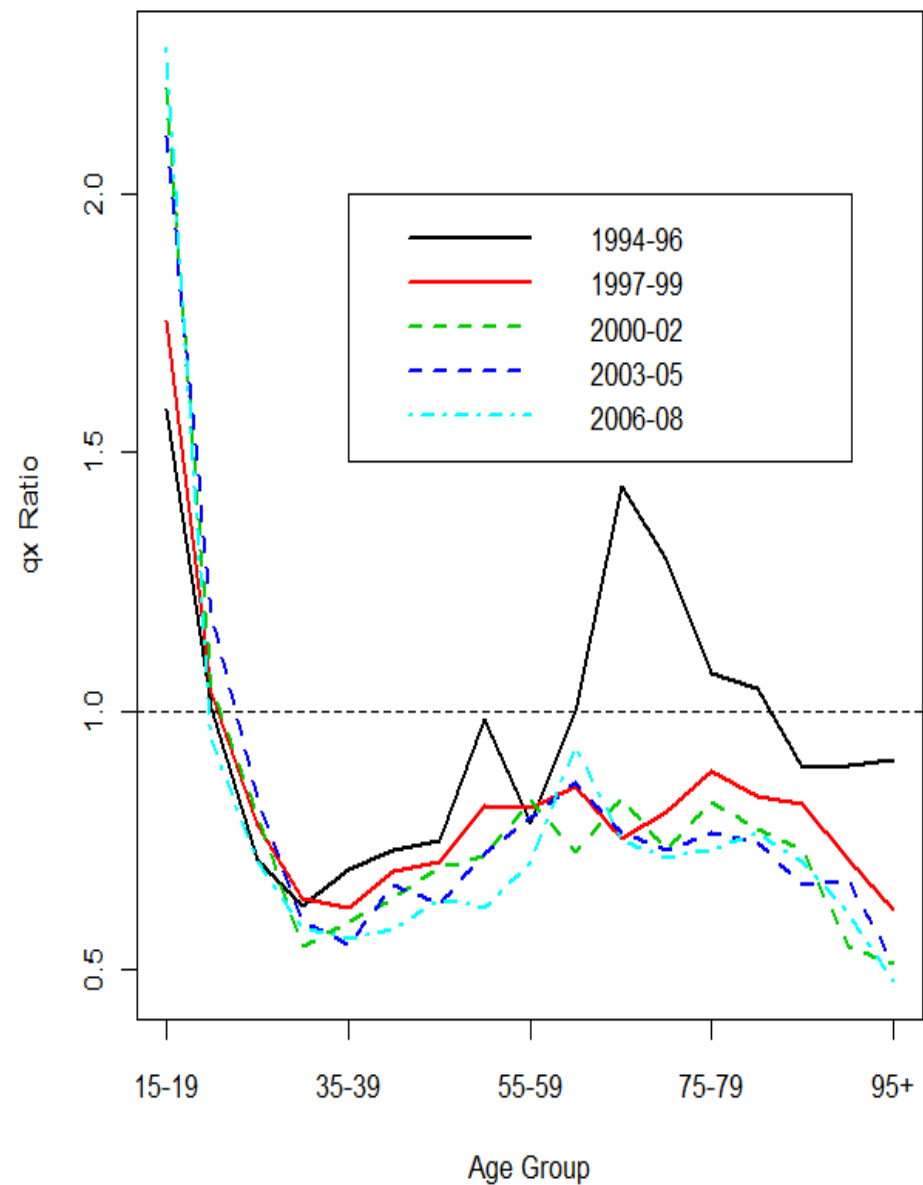


1996-98年及2006-08年女性婚姻死亡率比值(未婚分母)

Male

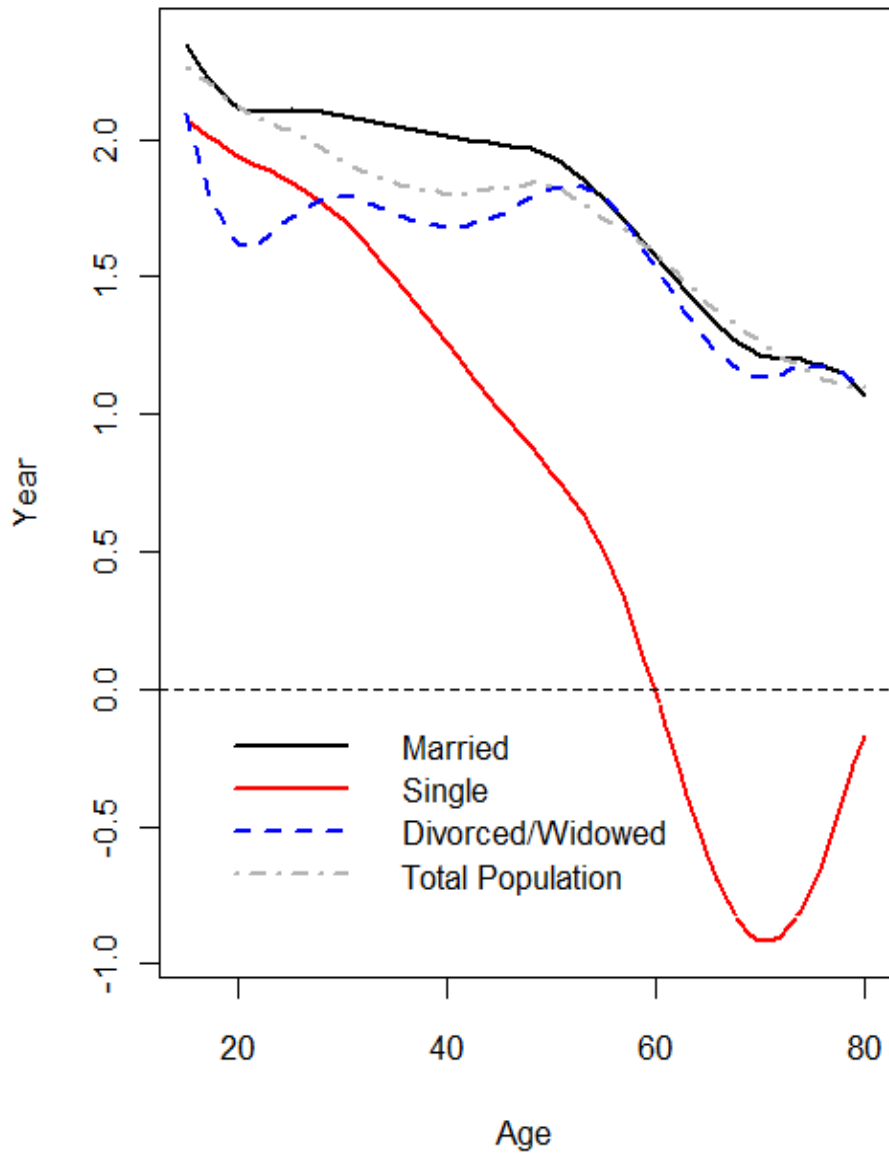


Female

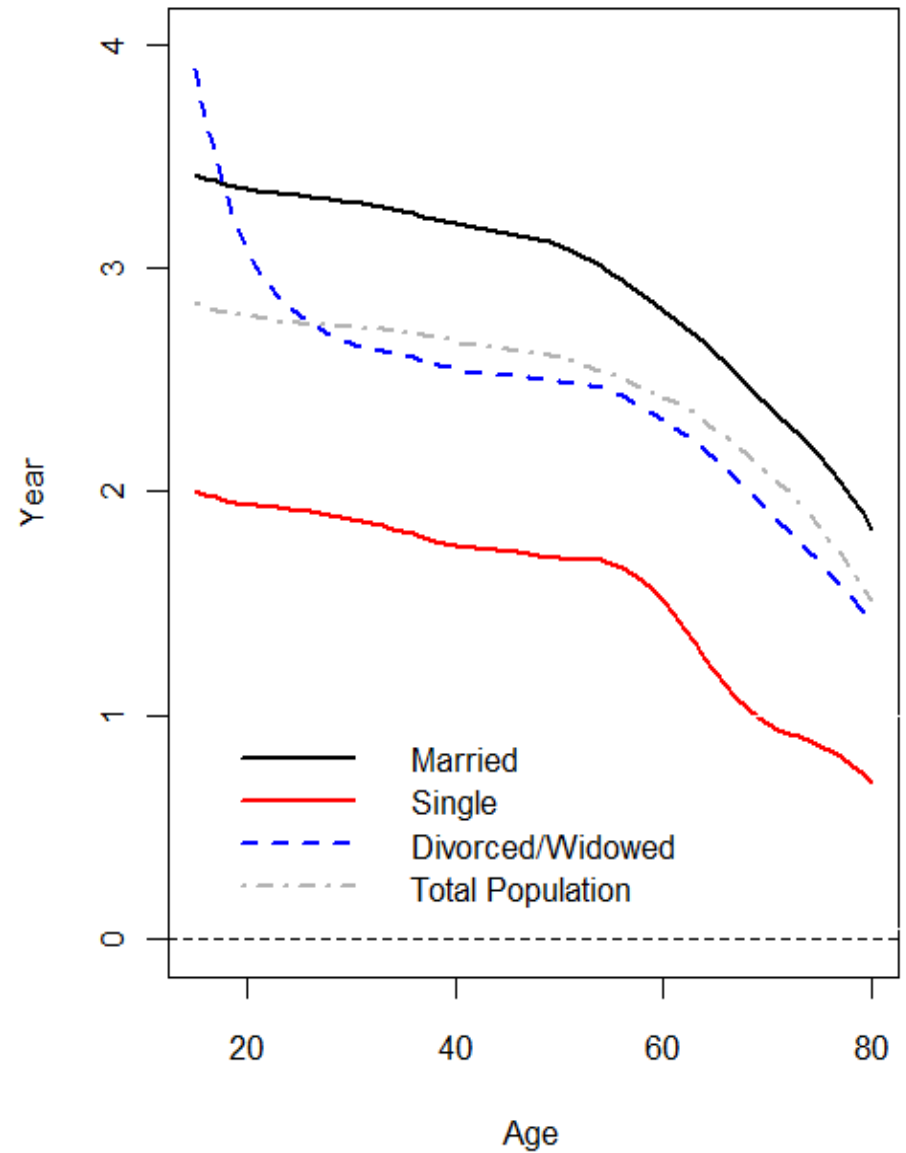


臺灣男女性已婚與未婚者死亡率比值(未婚為分母)

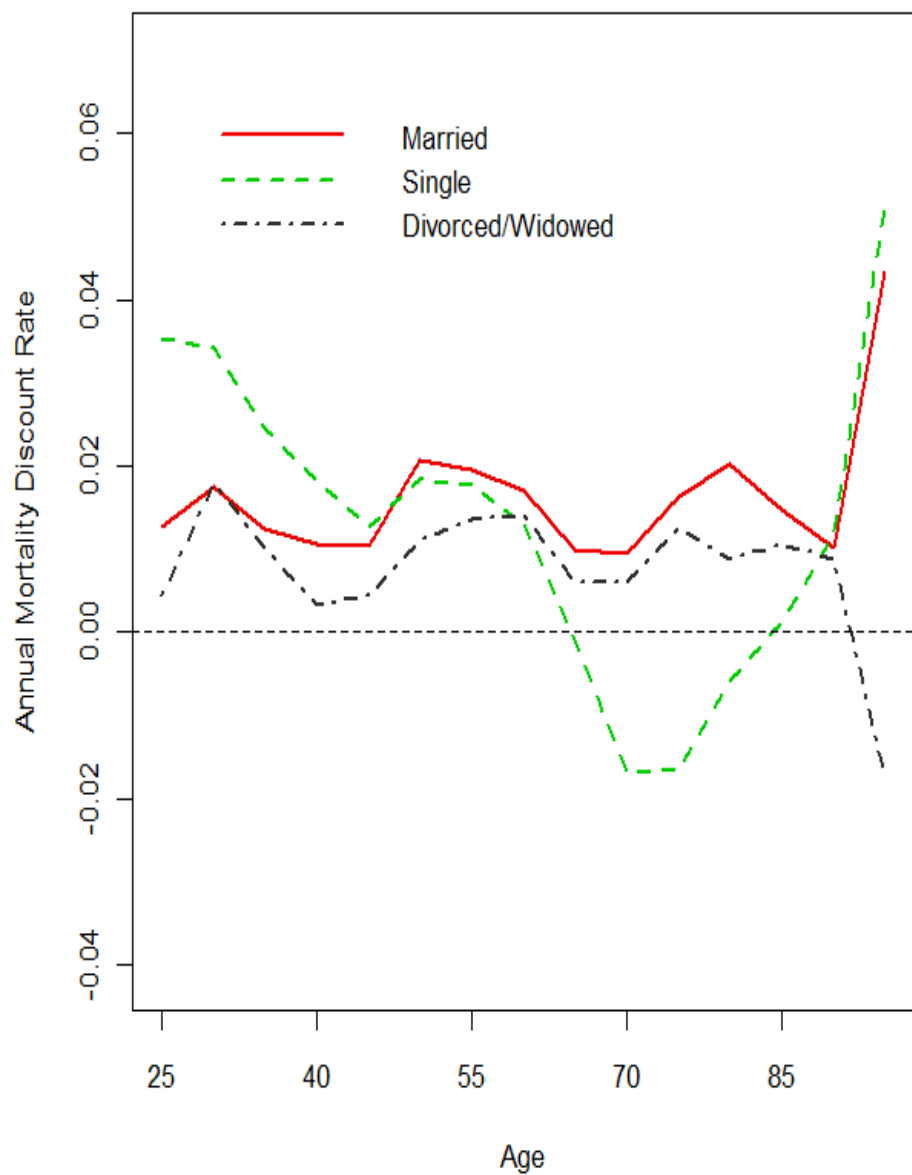
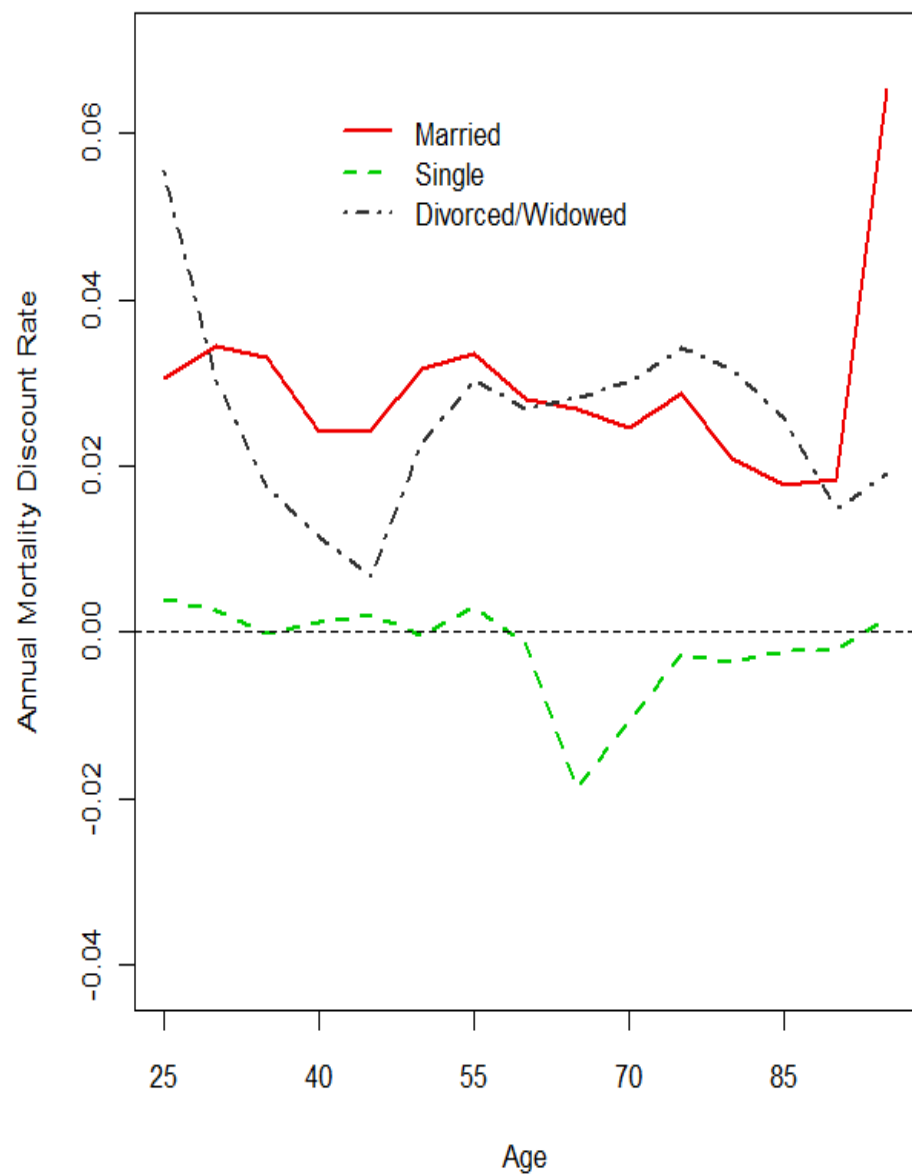
Male



Female



1996-98年至2006-08年的平均餘命上升幅度

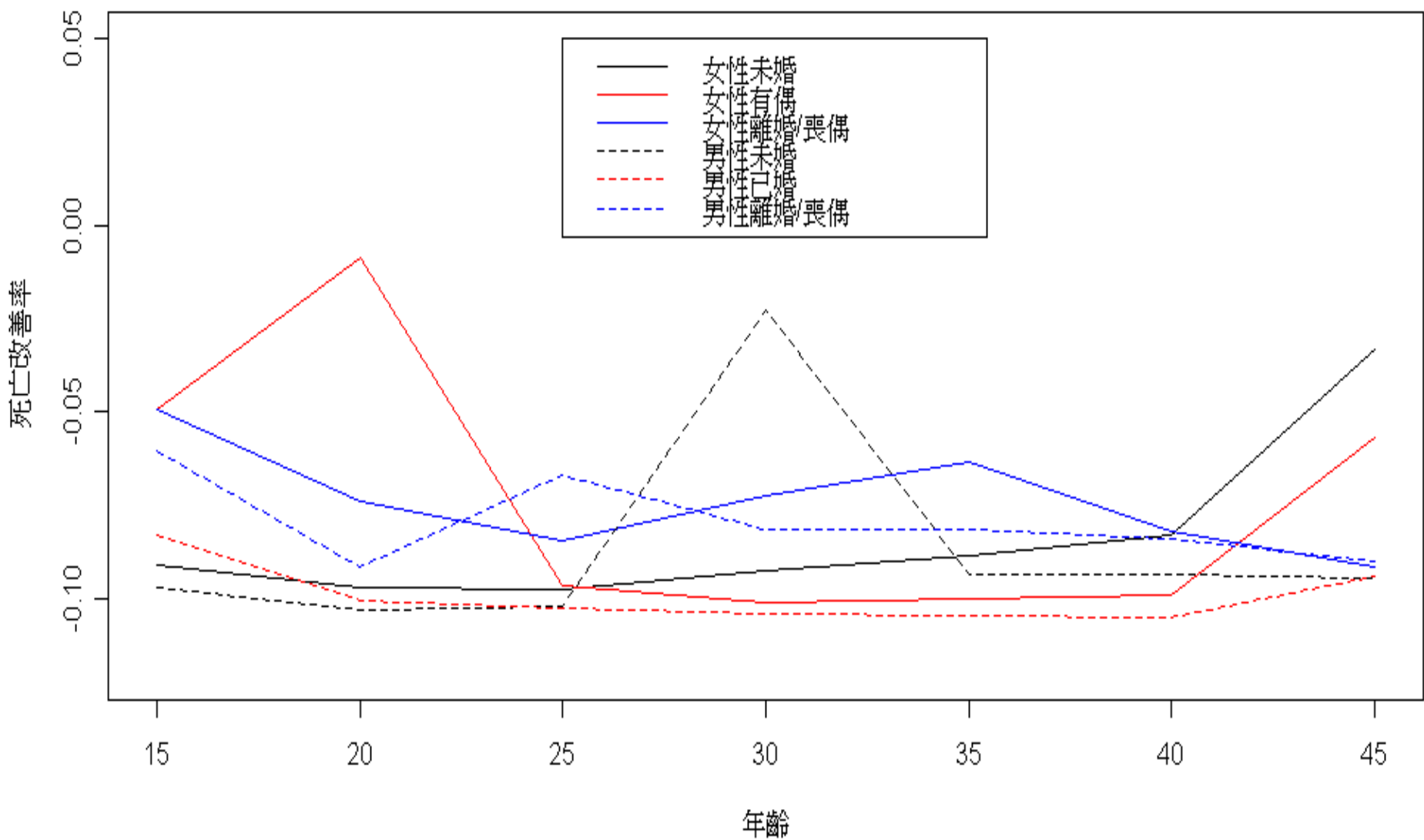
**Male****Female**

臺灣婚姻狀況的死亡率改善幅度(1994-2008, LC Model)



## 三種婚姻狀況在APC模型中的變化趨勢

	已婚	未婚	離婚/喪偶
男性	-0.0196	-0.0251	-0.0762
女性	-0.0697	-0.0510	-0.0798



臺灣婚姻狀況的死亡率變化趨勢(Lee-Carter模型)





# 婚姻的優勢

- 根據 Trowbridge(1994)的解釋，已婚者具有較低死亡率的原因有：
  - 自我選擇：婚姻也是一種自我選擇，強壯 (Robust)的男女性較容易結婚；
  - 責任感：不會選擇危險性較高的活動；
  - 互相照顧：生活起居及病痛煩惱時互相照顧，也有安定心理的作用；
  - 社會壓力：未婚者遭受有形及無形的壓力。