修勻學(Graduation) – Moving Weighted Average

授課教師：余清祥教授

課程日期：2024年9月25日

資料下載：

http://csyue.nccu.edu.tw

修勻的方法與模型

- 修勻(Graduation)大略可分成兩類：

— 不涉及模型的資料分析，例如：內插法。

— 假設生命函數(如：q_x、μ_x)服從某種模型，模型可為參數(Parametric)或半參數(Semi-parametric)

- 早期的修勻法大多屬於第一類，接下來先介紹的早期發展的兩種修勻法：MWA及Whittaker。

Moving Weighted Average (MWA)

- Moving Weighted Average (移動加權平均)

→因為變異數與樣本數成反比，變異數愈小也代表估計值與期望值愈接近。如果死亡率與年齡有接近線性的關係，幾個相鄰年齡死亡率的平均值，應該會更接近真實的死亡率(較小的變異數)。例如： $\frac{\hat{q}_{x-1}+\hat{q}_x+\hat{q}_{x+1}}{3} \leftrightarrow \hat{q}_x$

$$Var\left(\frac{\hat{q}_{x-1}+\hat{q}_x+\hat{q}_{x+1}}{3}\right) \simeq Var(\hat{q}_x)/3$$

MWA的發展歷史

- 保險界使用移動加權平均(MWA)，最早可追溯至19世紀，又稱為線性合成法(Linear Compound Formula)，也經常用於時間數列(Time Series)。

→因為計算容易，只需要考量各年齡死亡率的加權平均，加上樣本數較大時有不錯的效果，至今仍有不少國家仍然使用MWA編算生命表。

MWA的符號

u_x：x 歲原始估計值(大寫U_x為隨機變數，通常代表死亡率)

v_x：x 歲的修勻值(大寫V_x為 隨機變數)

t_x ：x 歲理論值(真實的死亡率)

e_x：x 歲死亡率的隨機誤差

d_x：x 歲死亡人數

l_x ：x 歲總人數(或記為P_x)

q_x：x 歲死亡率

死亡率的基本假設

- 處理的問題時，通常假設

  觀察值＝模型＋誤差；

  Observation = Model + Error.

以死亡率的角度而言，

$$u_x = t_x + e_x$$

為求計算方便，大多假設誤差與年齡無關，例如： $e_x \sim N(0,\sigma^2)$ ，而且各年齡間的誤差也互相獨立。

- 但這些誤差假設與實務有不少差異。

→例如：一群年齡相同的人，在一年之內會死亡的機會大約相同，因此可假設死亡人數服從二項分配：

$$D_x \sim B(P_x, q_x) \Rightarrow Var(\hat{q}_x) = \frac{q_x(1-q_x)}{P_x}$$

因為各年齡的人數不見得相同，死亡率也有差異，誤差的變異數會隨著年齡改變。

→同理，相鄰年齡的死亡率一般都有關聯。

範例一、死亡率隨年齡線性上升

- 假設死亡率隨年齡而直線上升，也就是

$$q_x = \alpha + \beta x$$

以電腦模擬測試MWA的穩定性，若各年齡有一萬人，參數滿足α=0.01、β=0.001，年齡50歲至70歲，死亡人數為二項分配隨機變數。

可預期各年齡死亡率觀察值滿足

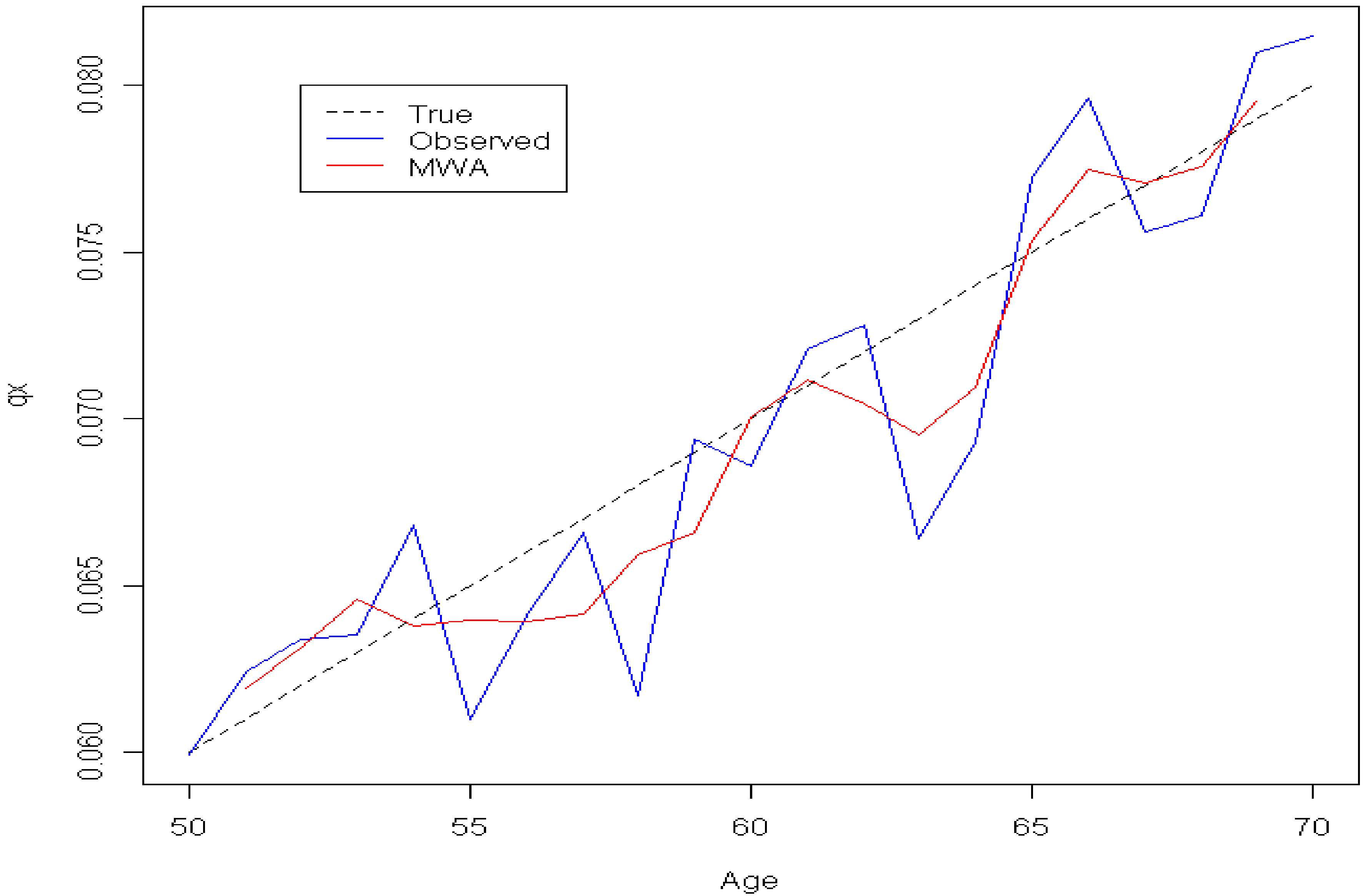
$$E(\hat{q}_x) = E(\frac{d_x}{n_x}) = \alpha + \beta x = q_x$$

$$Var(\hat{q}_x) = \frac{q_x(1-q_x)}{10000}$$

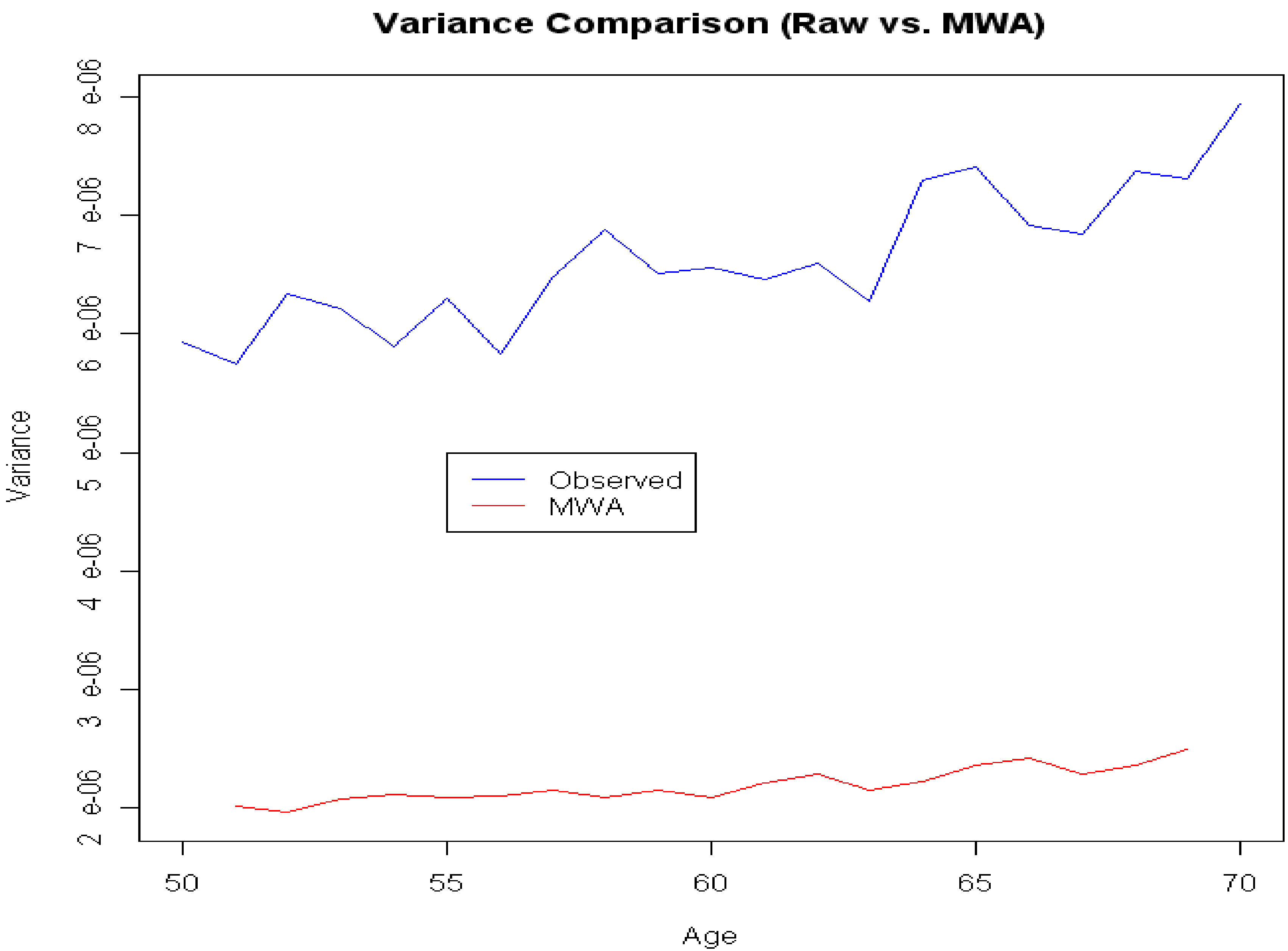
- 如果死亡率觀察值能反映真正的死亡率，在本範例中為隨年齡而直線上升，則MWA修勻後的死亡率將更平滑、更接近理論值。

→如果 $v_x = (u_{x-1} + u_x + u_{x+1})/3$ ，平均而言，各年齡的死亡率修勻值將更接近理論值，可能震盪幅度也會小一些。

→以電腦模擬驗證，重複1000次模擬，死亡率修勻值的變異數確實小於觀察值的變異數，其差額與MWA的項數有關。



理論、觀察、MWA修匀死亡率比較(範例一)



觀察、MWA修勻死亡率的變異數比較(1000次模擬)

MWA的基本公式

- MWA的修勻公式一般表示為

$$v_x = \sum_{r=-\infty}^{\infty} a_r u_{x+r}$$

通常又會假設以下兩個條件：

(1) 有限性(Boundedness)：$a_r = 0$，如果 $r > n$ 或 $r < -n$，其中$2n+1$為範圍(Range)。

(2)對稱性(Symmetry)： $a_r = a_{-r}$ ，$r = 1, 2, 3, ..., n$。

MWA的基本公式(續)

- 代入上述條件，MWA的公式可改為

$$v_x = \sum_{r=-n}^{n} a_r u_{x+r}$$

- MWA兩個常見的要求：

(1)還原性(Reproduce)：$E(V_x) = t_x$；

(2)縮小變異(Reducing Variations)：

$$\begin{cases} Var(V_x) \le Var(U_x) \\ Var(\Delta^z V_x) \le Var(\Delta^z U_x) \end{cases}$$

還原性

- 如果真實值為 k 次多項式，則還原性要求

$$\sum_{r=-n}^{n} a_r = 1 \qquad (1)$$

$$\sum_{r=-n}^{n} r^j a_r = 0,\ j = 1, 2, \ldots, k \qquad (2)$$

但(1)式中的奇數次方程式必然成立，上述的限制式其實只包括[k/2]+1 個方程式。

- 如果 t_x 與年齡 x 無關，或是 $t_x = c_0$，代入可得

$$t_x = \sum_{r=-n}^{n} a_r t_{x+r} \Rightarrow \sum_{r=-n}^{n} a_r c_0 = c_0 \Rightarrow \sum_{r=-n}^{n} a_r = 1$$

也就是 $a_0 + 2a_1 + \cdots + 2a_n = 1$

- 如果 $t_x = x$（與年齡呈現線性關係），

$$t_x = \sum_{r=-n}^{n} a_r t_{x+r} \Rightarrow x = \sum_{r=-n}^{n} a_r (x+r)$$

可得 $\sum_{r=-n}^{n} a_r = 1, \ \sum_{r=-n}^{n} r a_r = 0$

- 如果 $t_x = x^2$，可得 $\sum_{-n}^{n} r a_r = \sum_{-n}^{n} r^2 a_r = 0$

縮小變異

- 縮小變異通常透過下列數值獲得：

$$R_z^2=\frac{Var(\Delta^z V_x)}{Var(\Delta^z U_x)}=\frac{1}{\binom{2z}{z}}\sum_{-\infty}^{\infty}(\Delta^z a_r)^2$$

其中 $z=0$為特例，數值的計算為：

$$R_0^2=\frac{Var(V_x)}{Var(U_x)}$$

註：上述的推導需假設不同年齡的死亡率不相關；通常 z 的選擇為 2 或 3。

- 顧名思義，修勻後的死亡率應該會比較平滑，或是修勻值的變異數較小，即 $R_0^2=\dfrac{Var(V_x)}{Var(U_x)}\le 1$

→計算 $Var(V_x)=E[(V_x-t_x)^2]=E[(\sum_{-n}^{n}a_rE_{x+r})^2]$

$$(\text{誤差不相關})=\sum_{-n}^{n}a_r^2 Var(E_{x+r})=\sigma^2\sum_{-n}^{n}a_r^2$$

因此要求

$$R_0^2=\frac{Var(V_x)}{Var(U_x)}=\frac{\sigma^2\sum_{-n}^{n}a_r^2}{\sigma^2}=\sum_{-n}^{n}a_r^2\le 1$$

→若誤差與樣本數有關，$R_0^2=\dfrac{Var(V_x)}{Var(U_x)}=\sum_{-n}^{n}a_r^2\left(\dfrac{n_x}{n_{x+r}}\right)$

- 修勻後各年齡間的死亡率會比較平滑，若死亡率為年齡的線性函數，對修勻後的死亡率微分(差分)兩次應該為0。定義： $R_z^2 = \frac{Var(\Delta^z V_x)}{Var(\Delta^z U_x)}$

→令N為不同年齡的個數，則

$$\sum_x Var(\Delta^z V_x) = N\sigma^2 \times \sum_{-n}^{n} (\Delta^z a_r)^2$$

$$\sum_x Var(\Delta^z U_x) = \sum_x Var(\sum_{j=0}^{z} \binom{z}{j} (-1)^{z-j} U_{x+j})$$

$$= N\sigma^2 \times \binom{2z}{z}$$

MWA係數的一般式

- MWA的係數可由下列函數表示：

$$a_r = [(n+1)^2 - r^2]\cdots[(n+1)^2 - r^2][h + k(n^2 - r^2)]$$

其中參數 h 及 k 由下列方程式求解

$$\begin{pmatrix} S_{n,z} & S_{n-1,z+1} \\ S_{n,z+1} & S_{n-1,z+2} \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix} = \begin{pmatrix} 1 \\ (n+z+1)^2 \end{pmatrix}$$

且

$$S_{n,m} \equiv (m!)^2 \begin{pmatrix} 2n+2m+1 \\ 2m+1 \end{pmatrix}$$

還原三次多項式的 a_r

	a_0	a_1	a_2	a_3	a_4
n=2, z=0	0.4857	0.3429	–0.0857	—	—
n=2, z=3	0.5594	0.2937	–0.0734	—	—
n=4, z=0	0.2554	0.2338	0.1688	0.0606	–0.0909
n=4, z=3	0.3311	0.2666	0.1185	–0.0099	–0.0407

MWA的優點在於計算方便

- MWA的係數通常可透過公式簡化。

→歷史上較為知名的有Spencer修勻公式，較為常用的有15點及21點公式兩種：

15點公式： $v_x = \dfrac{1}{320}[4]^2[5]\{1+6[3]-3[5]\}u_x$

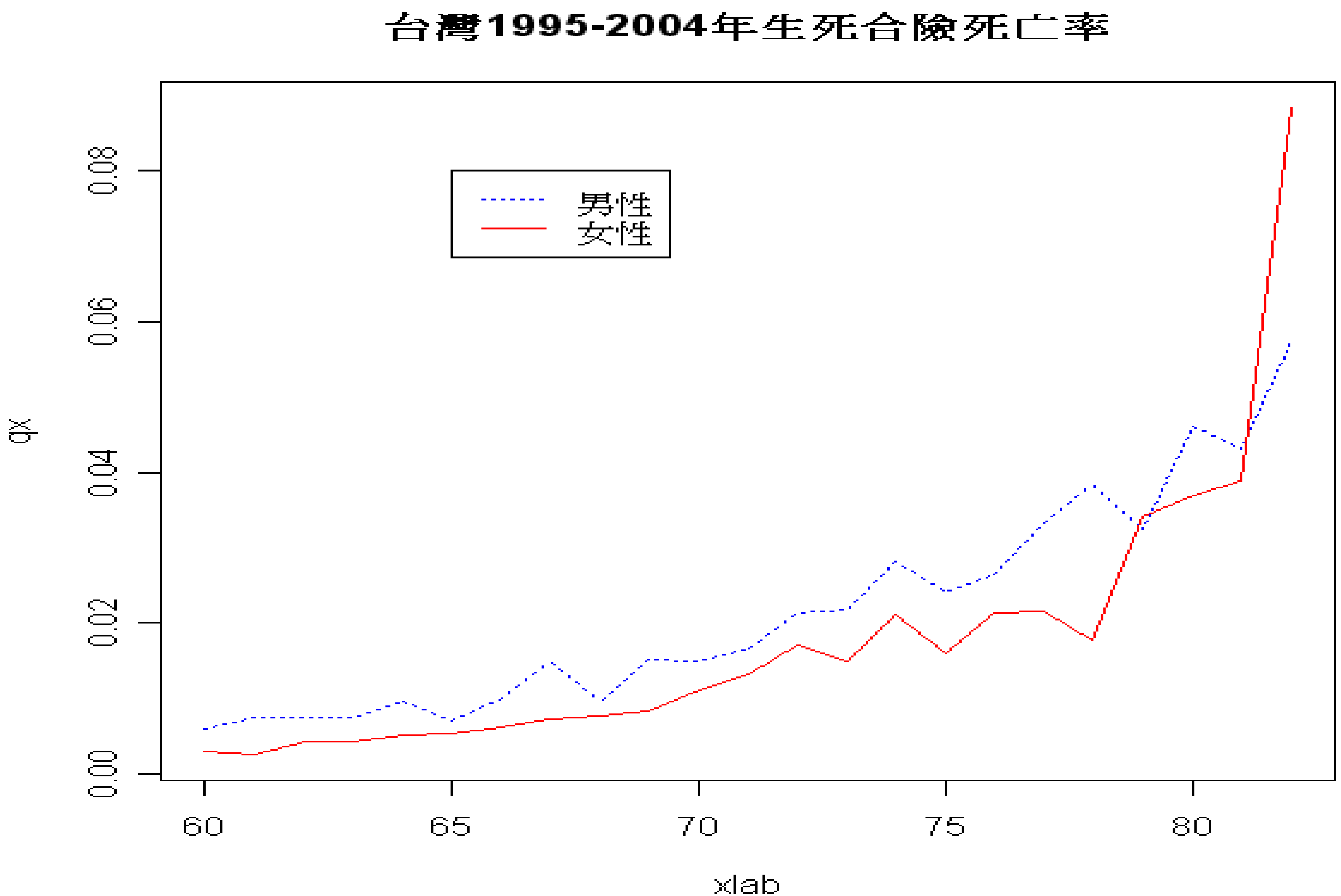
21點公式： $v_x = \dfrac{1}{350}[5]^2[7]\{1+[3]+[5]-[7]\}u_x$

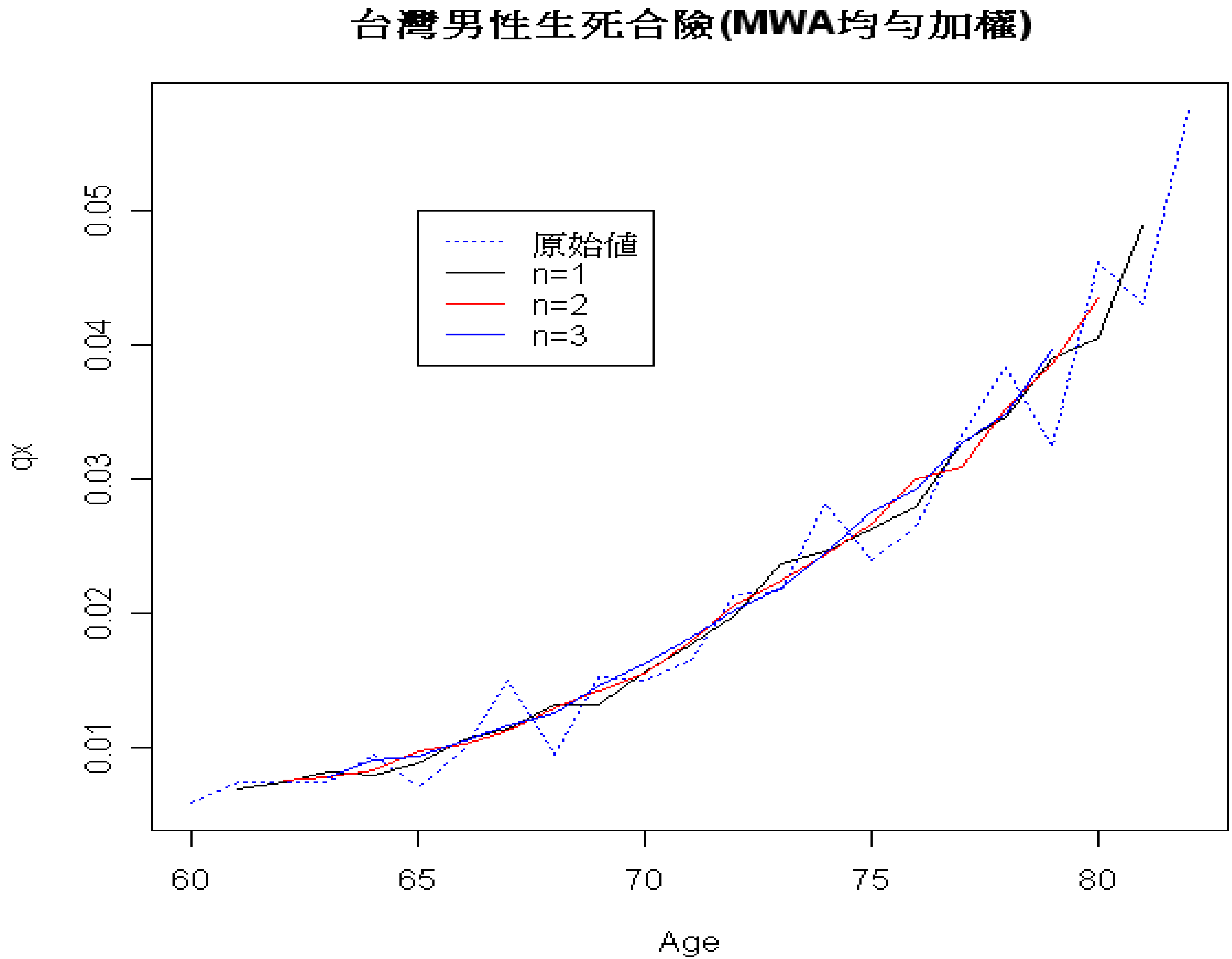
其中[ ]為移轉運算子(Shift Operator)，滿足

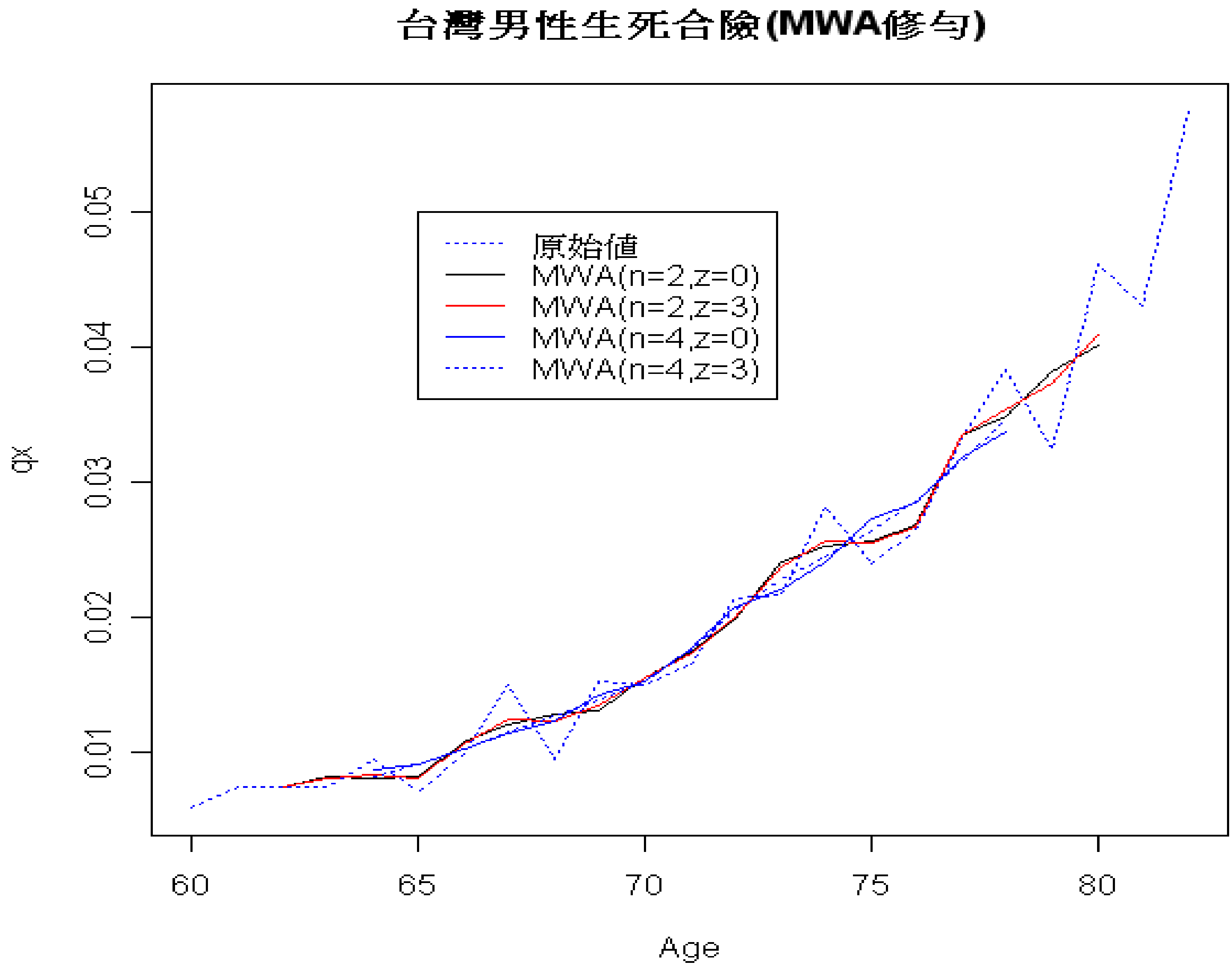
$$[3]u_x = u_{x-1} + u_x + u_{x+1}$$

$$[4]u_x = u_{x-3/2} + u_{x-1/2} + u_{x+1/2} + u_{x+3/2}$$

範例二：台灣1995-2004年生死合險

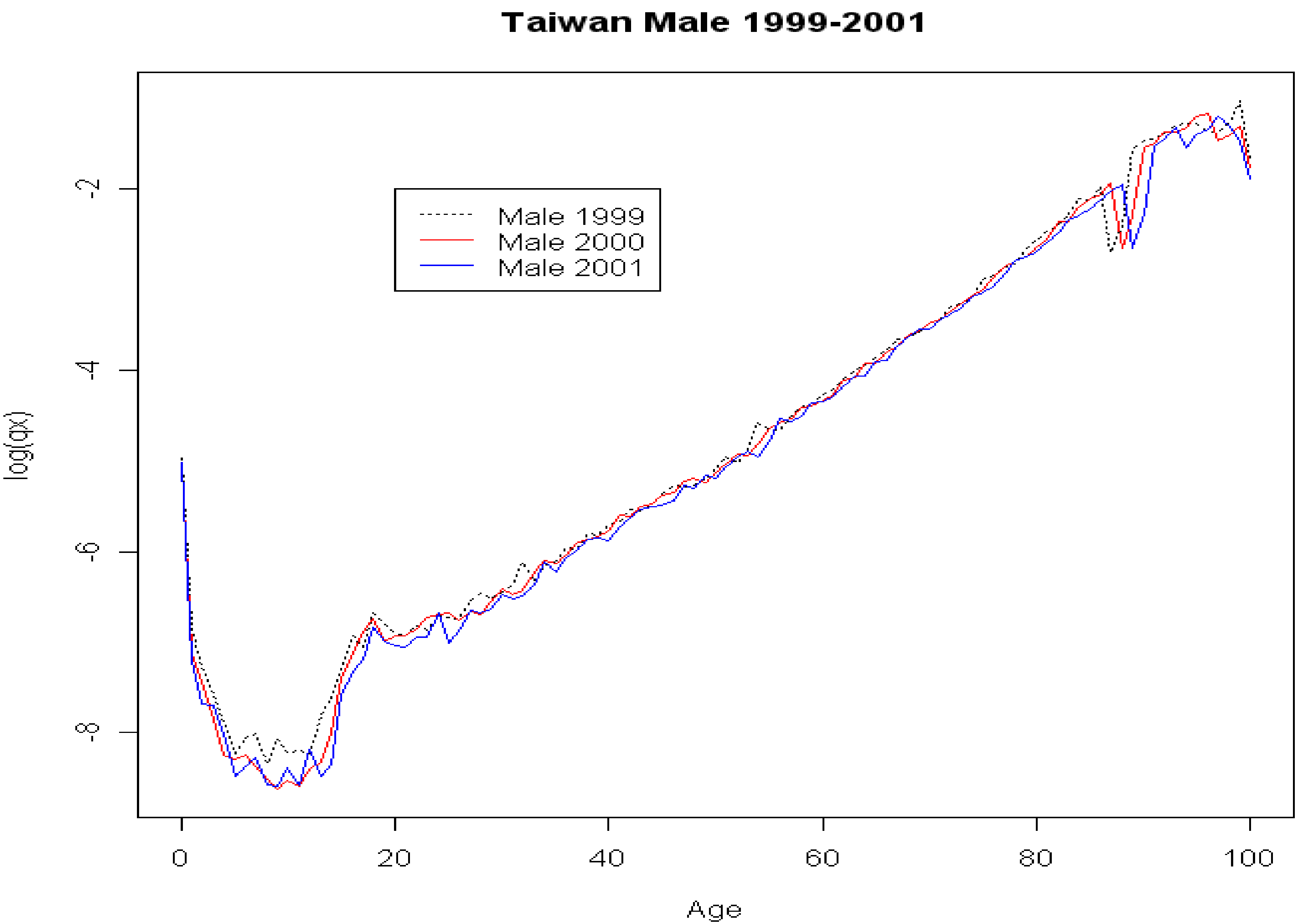



台灣男性生死合險(MWA均勻加權)
原始值
n=1
n=2
n=3
qx
0.01
0.02
0.03
0.04
0.05
60
65
70
75
80
Age



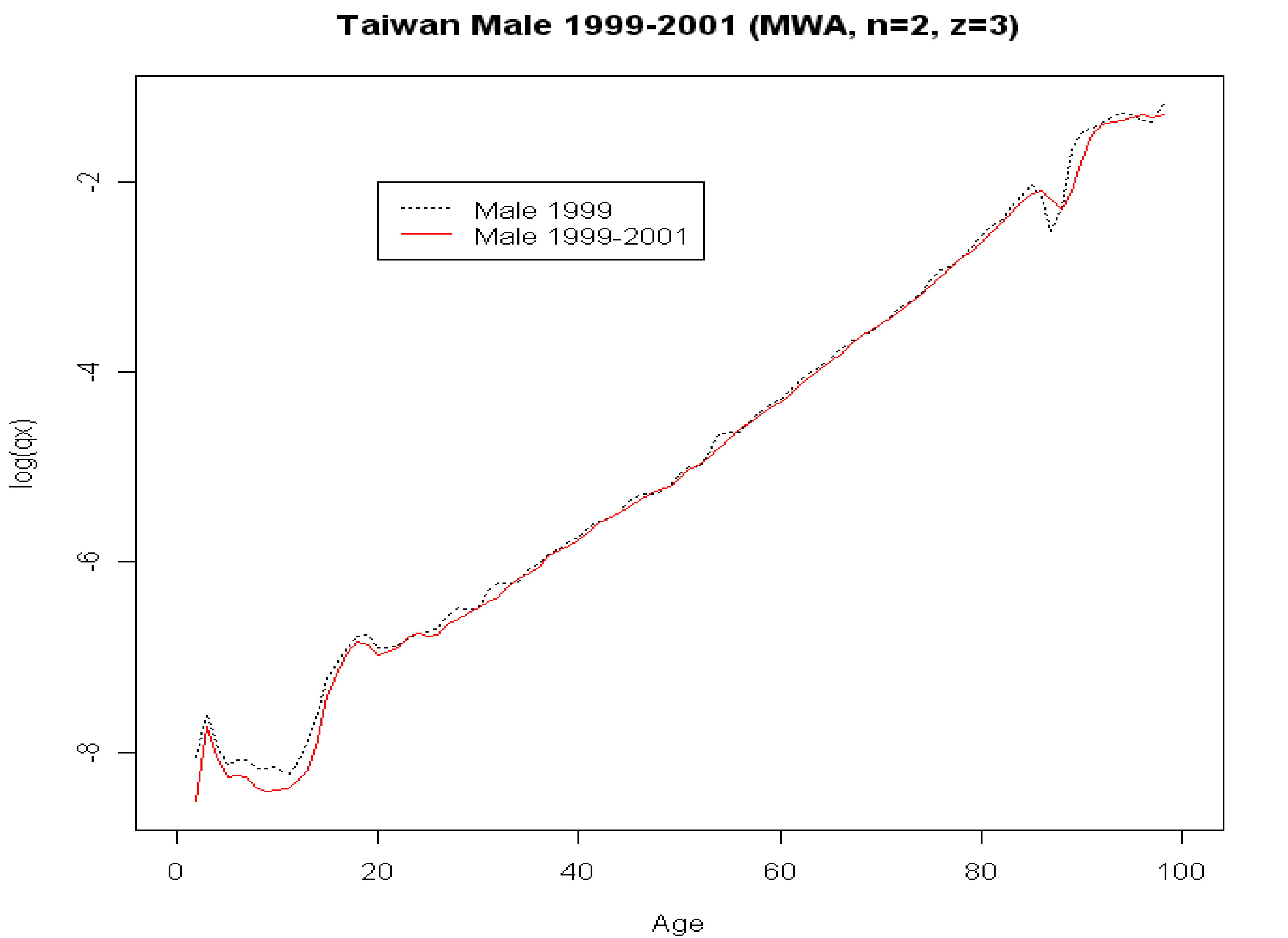
台灣男性生死合險(MWA修勻)
原始值
MWA(n=2,z=0)
MWA(n=2,z=3)
MWA(n=4,z=0)
MWA(n=4,z=3)
qx
Age
0.01
0.02
0.03
0.04
0.05
60
65
70
75
80

範例三：台灣1999-2001年男性生命表



- 單一年度的生命表通常因為死亡率變動大，通常會合併幾個年度的資料再考慮修勻，以第九回台灣地區完全生命表為例，即合併了1999-2001年的人口數、死亡人數編算而得，修勻後的死亡率較單一年度平滑。

→實證上，合併的年數與各地方人口數有關，較常見的為3年或5年，人口數少者也有合併7年、甚至9年，像美國人口數較多（超過三億），近年來每年公佈單一年齡死亡率的生命表。



Taiwan Male 1999-2001 (MWA, n=2, z=3)
Male 1999
Male 1999-2001
log(qx)
-2
-4
-6
-8
0
20
40
60
80
100
Age

- 註：MWA的參數中，對死亡曲線較有影響者是 n 值，較大的範圍通常代表較為平滑。然而，MWA會犧牲邊緣值，也就是死亡率兩端各有 n 個年齡無法修勻，而且接近邊緣值的死亡率也較不平滑。

→犧牲邊緣值的問題，可只取單邊的修勻公式，雖然會損失部分的平滑性。較常見的修勻公式為Greville’s 三次九項，我國內政部統計處過去編算生命表使用這個公式。

Greville三次九項公式

$$q_1 = \frac{1}{14586}\left(9449q_1^{'} + 9800q_2^{'} + 980q_3^{'} - 5880q_4^{'} - 4410q_5^{'} + 1512q_6^{'} + 4060q_7^{'} + 1000q_8^{'} - 1925q_9^{'}\right)$$

$$q_2 = \frac{1}{58344}\left(13475q_1^{'} + 23096q_2^{'} + 20090q_3^{'} + 8820q_4^{'} - 1470q_5^{'} - 5040q_6^{'} - 2702q_7^{'} + 700q_8^{'} + 1375q_9^{'}\right)$$

$$q_3 = \frac{1}{29172}\left(385q_1^{'} + 5470q_2^{'} + 11464q_3^{'} + 11340q_4^{'} + 5040q_5^{'} - 1860q_6^{'} - 3760q_7^{'} - 772q_8^{'} + 1595q_9^{'}\right)$$

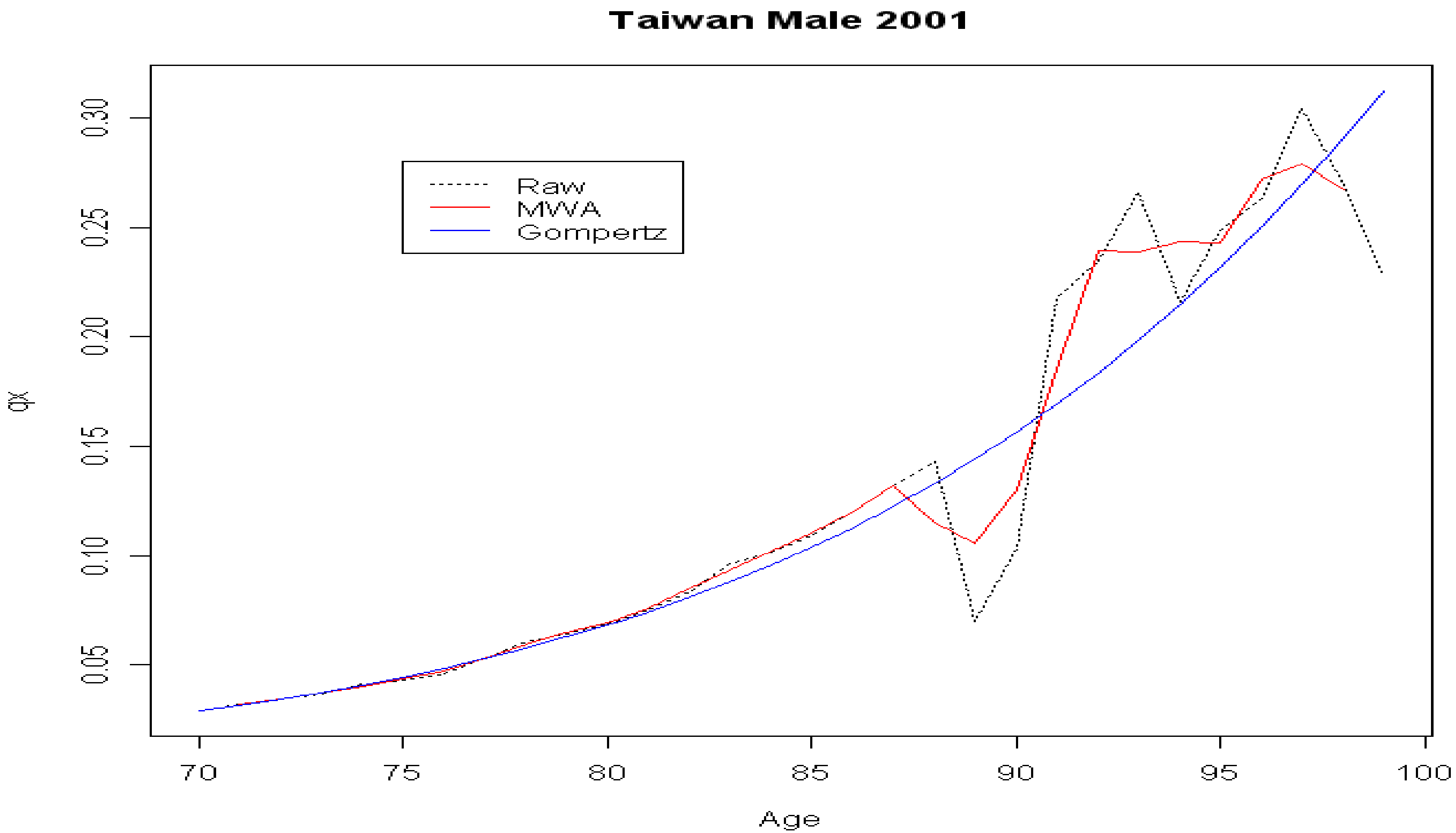
$$q_4 = \frac{1}{19448}\left(-1155q_1^{'} + 1260q_2^{'} + 5670q_3^{'} + 7736q_4^{'} + 5670q_5^{'} + 1620q_6^{'} - 930q_7^{'} - 720q_8^{'} + 297q_9^{'}\right)$$

五歲以上、七十歲以下

$$q_x = \frac{1}{2431}(-99q_{x-4}^{'} - 24q_{x-3}^{'} + 288q_{x-2}^{'} + 648q_{x-1}^{'} + 805q_x^{'} + 648q_{x+1}^{'} + 288q_{x+2}^{'} - 24q_{x+3}^{'} - 99q_{x+4}^{'})$$

註：0至1歲的死亡率分開處理。

- 除了邊緣值外，MWA的修勻為各年齡死亡率的加權平均，權數 a_r 與各年齡人數無關，通常在高齡死亡率修勻時容易產生較大震盪。 $v_x = (u_{x-1} + u_x + u_{x+1})/3$



- 上述MWA的修勻公式多半假設固定變異數、各年齡間的死亡率不相關求得，雖然可以解除這兩個假設求得修勻係數，但係數將與變異數、樣本數有關，反而喪失MWA計算上的優勢，實證上很少採用。

→如果考量樣本數的加權，建議可使用下一單元介紹的Whittaker修勻。

→如果兼顧五齡組死亡率轉化成單齡死亡率，可採用Spline或是Kernel修勻。

MWA的使用建議

- 當各年齡樣本數接近、死亡率沒有劇烈震盪，因為MWA提供穩定修勻值，可以斟酌使用。

→建議使用修勻方法前，先評估各年齡死亡率、樣本數，再決定是否要套用MWA或其他方法。

→特別注意當資料有離群值、遺漏值，或需要由五齡組死亡率轉成單齡死亡率。

→MWA結果可與其他修勻結果比較，作為檢驗死亡資料的參考。