人口統計(Demography)

授課教師：余清祥教授

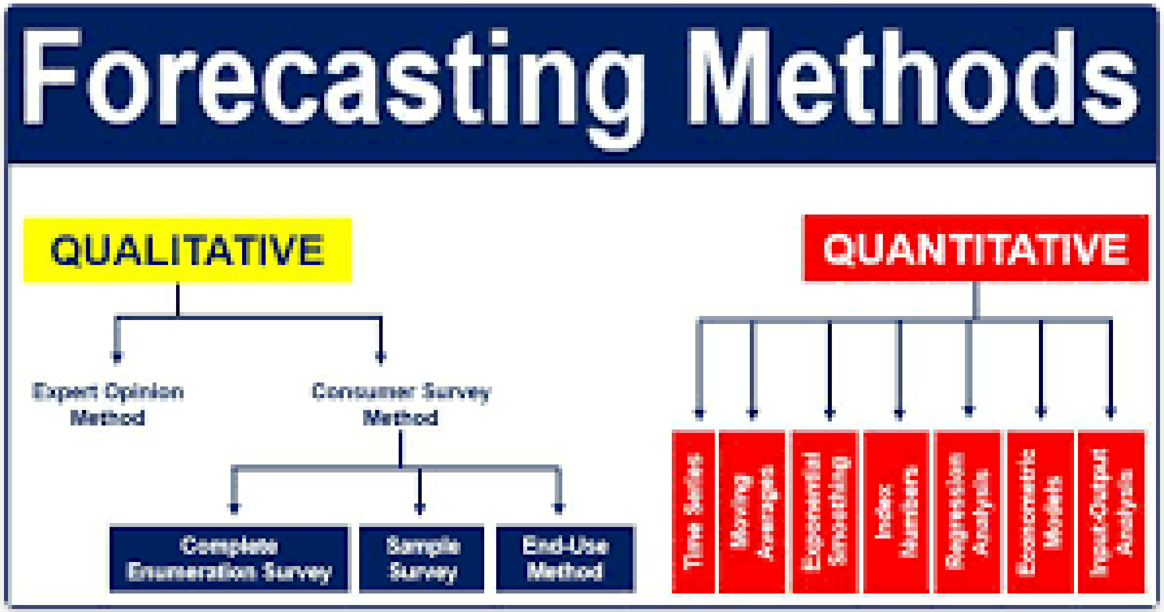
日期：2025年4月30日

第七講：人口推估

http://csyue.nccu.edu.tw

“Forecasting is very dangerous, especially about the future.”

--- Samuel Goldwyn



https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcSplX6Xaxd6mHFNO75OKaIrWxXjXutCq4azEA&usqp=CAU

https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcT8ukXai-mv4959ExESTcDq8A4Tt-_XD4jTLQ&usqp=CAU

https://64.media.tumblr.com/ce0b7b8077c4acba8207df1a469cc601/tumblr_inline_pjzrhwxE9Y1qzmz93_500.png

https://64.media.tumblr.com/9ff1e7b961225022693504f61c0a6879/tumblr_inline_pjzrhvqsWB1qzmz93_500.gifv

來源：呆伯特(Dilbert)漫畫

兩次普查間的人口估計

- 兩次戶口普查間的人數估計，通常有以下數種較常用的模型：

→線性內插(Linear Interpolation)

$$P(t) = (1-\frac{t}{n})\cdot P(0) + \frac{t}{n}\cdot P(n)$$

$$= P(0) + \frac{t}{n}\cdot\left[P(n)-P(0)\right], \quad 0 \le t \le n.$$

未來的人口推估(Projection)也可仿此，但建議時間 $t < 2n$。

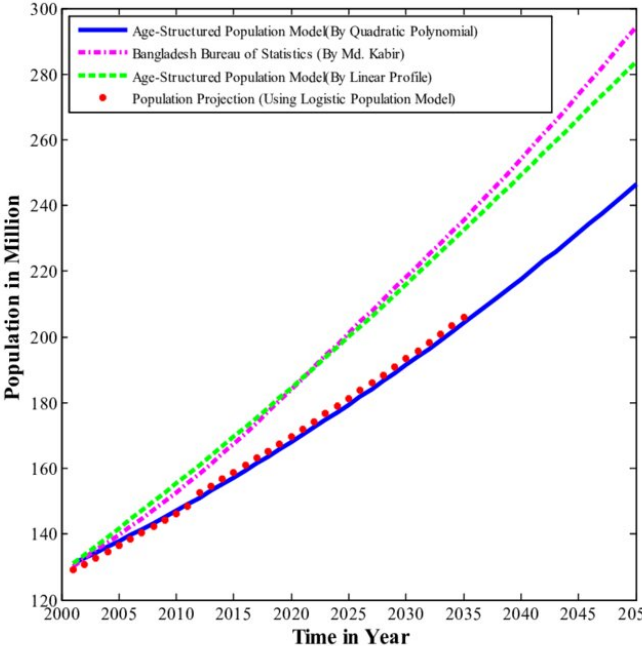
→多項式內插(Polynomial Interpolation)

若有 $n+1$ 次的歷史普查資料，則可配適 n 次多項式，即

$$P(0), P(1), P(2), \ldots, P(n) \to f(x)。$$

時間介於 0 與 n 間的任何數值都可由內插的方式獲得。

然而，因為二次以上的多項式變化較難預測，不建議由多項式函數預測未來的人口數(與線性內插類似)。



https://www.researchgate.net/profile/S-M-Mahmudul-Hasan/publication/283195249/figure/fig4/AS:667600921047042@1536179871190/Comparison-of-our-population-projection-with-different-methods-Rashed-Kabirhave-used_W640.jpg

→幾何(Geometric)模型

通常人口以幾何級數(複利)成長，即

$$P(t) = P(0) \cdot (1 + r_a)^t,$$

其中 r_a 是實質年成長率(annual effective rate)；或是

$$P(t) = P(0) \cdot e^{r_i t},$$

而 r_i 為瞬間成長率(continuous instantaneous growth rate)。

實質年成長率可由兩次普查的資料導出：

$$P(t) = P(0) \cdot (1 + r_a)^t \Rightarrow r_a = \left(\frac{P(n)}{P(0)} \right)^{\frac{1}{n}} - 1$$

同理，瞬間成長率也可由下式導出：

$$P(t) = P(0) \cdot e^{r_i t} \Rightarrow r_i = \frac{1}{n} \cdot \ln\left(\frac{P(n)}{P(0)} \right) = \ln(1 + r_a)$$

註：上述的內插法(包含幾何模型)，通常不宜用於預測距離現在較久遠以後的未來。

- 範例一：假設 P(80)=227.7及 P(90)=249.9 (單位：百萬)，試求P(88)及P(99)。

→線性內插：

$$P(88)=P(80)+\frac{8}{10}\big(P(90)-P(80)\big)=245.5$$

$$P(99)=P(80)+\frac{19}{10}\big(P(90)-P(80)\big)=269.9$$

→實質成長率：

$$r_a=\left(\frac{249.9}{227.7}\right)^{\frac{1}{10}}-1=0.93466\%$$

$$\Rightarrow P(88)=P(80)\cdot(1+r_a)^8=245.3；\mathrm{P}(99)=271.7$$

→瞬間成長率：

$$r_i=\frac{1}{10}\cdot\ln\left(\frac{249.9}{227.7}\right)=0.93032\%$$

$$\Rightarrow P(88)=P(80)\cdot e^{8\,r_i}=245.3；\mathrm{P}(99)=271.7$$

- 範例二：若某個國家的人口瞬間成長率為 r_i，計算該國人口加倍所需的時間。

→人口加倍意指：

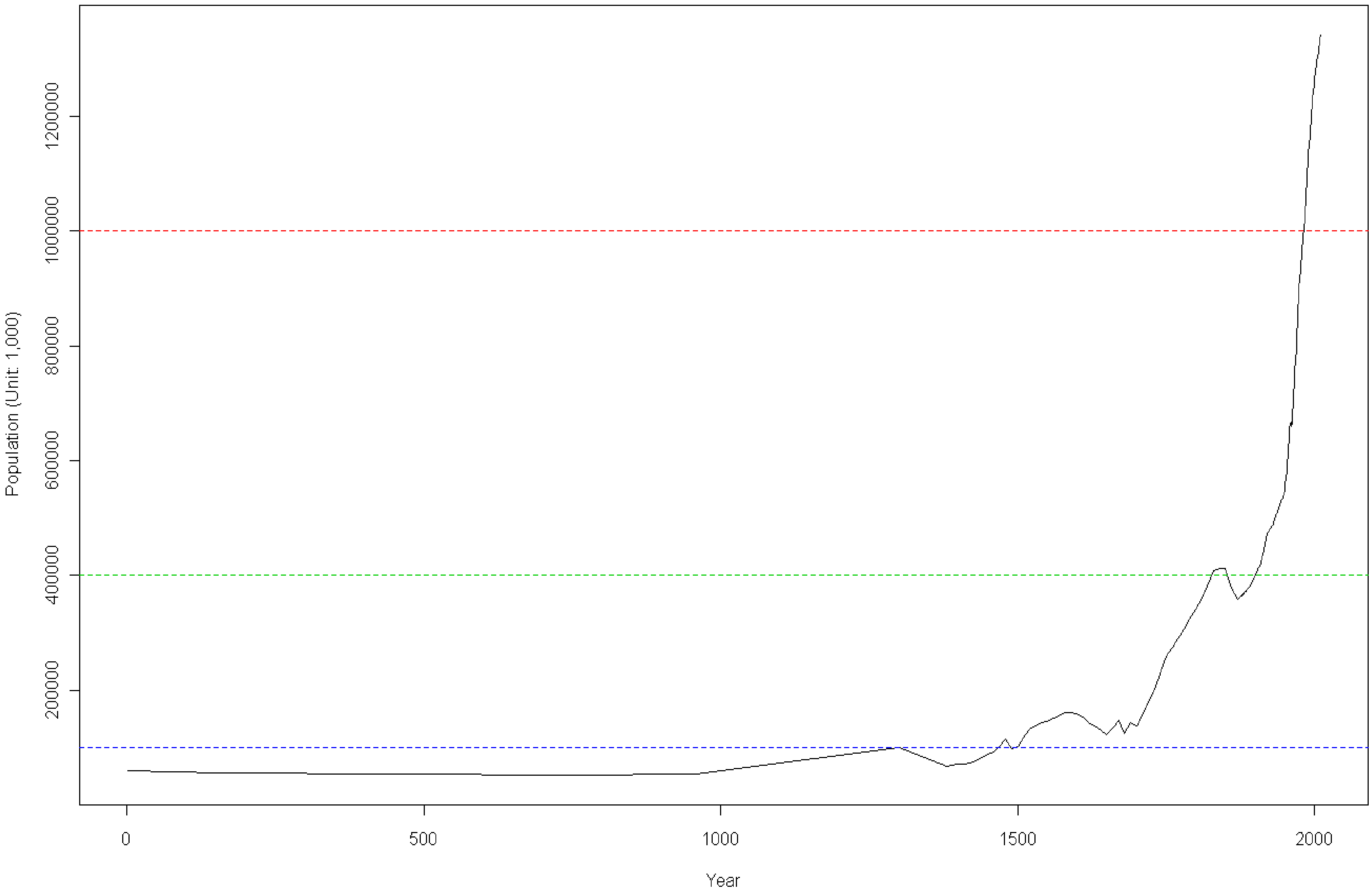
$$P \cdot e^{r_i t} = 2P \Rightarrow e^{r_i t} = 2$$

或可改寫為

$$r_i t = \ln 2 = 0.693 \Leftrightarrow t = \frac{0.693}{r_i} \approx \frac{70}{100 r_i}$$

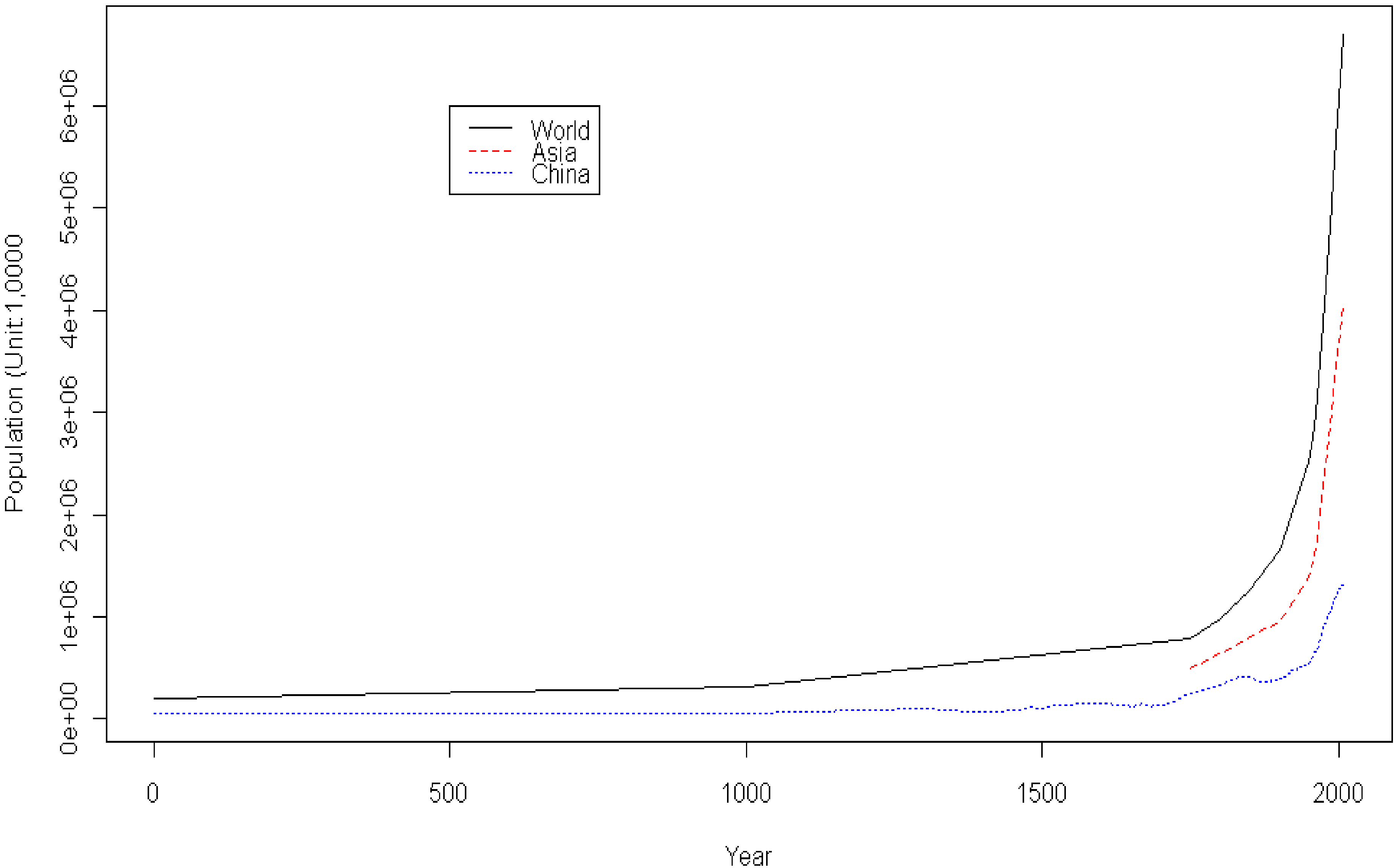
如果每年以 2%成長，則僅需35年人口即能加倍。(中國古代年成率有時不到0.1%，因此需要將近1000年方能加倍。)

中國兩千年來人口趨勢



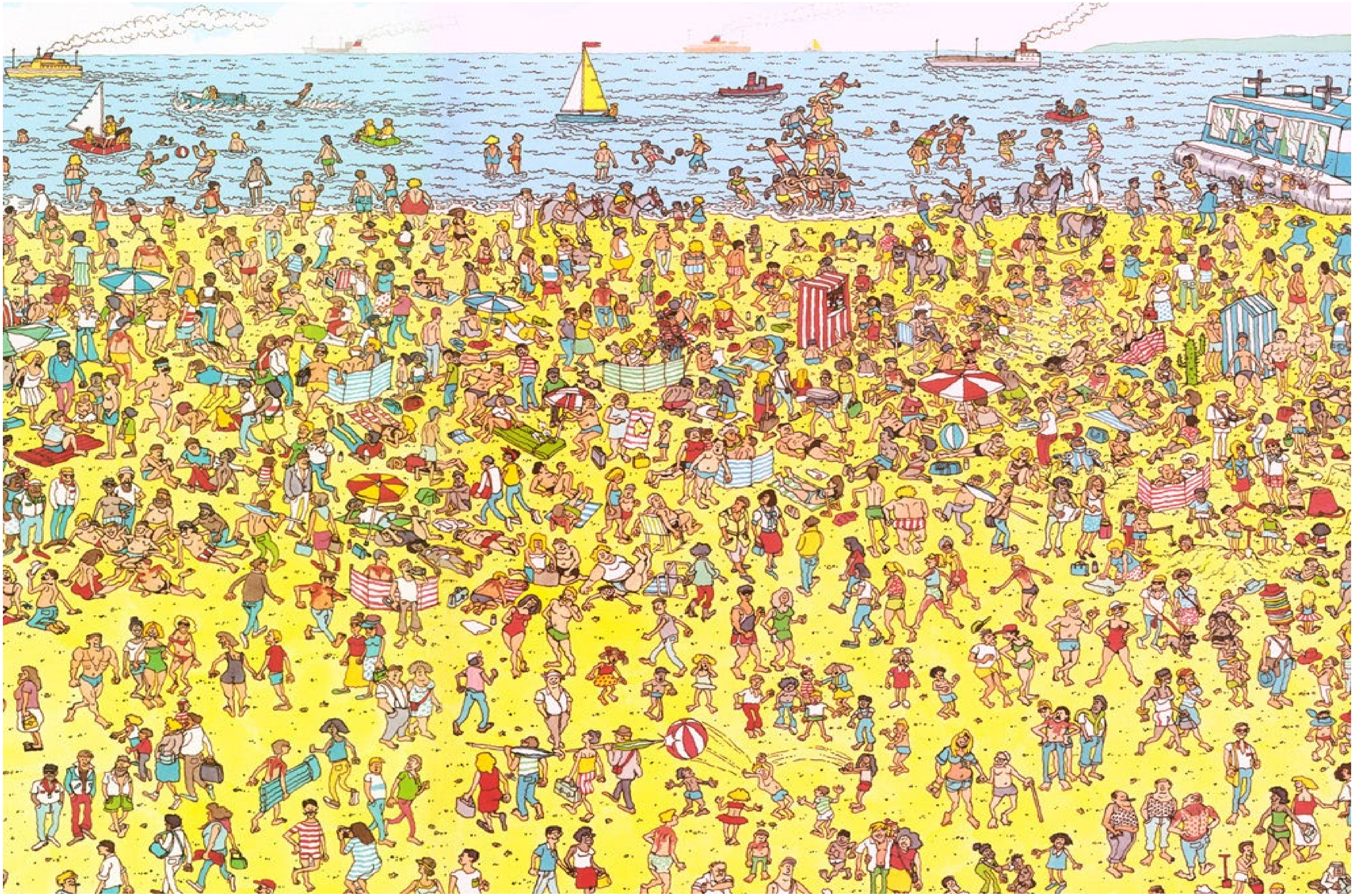
Population (Unit: 1,000)
200000
400000
600000
800000
1000000
1200000
0
500
1000
1500
2000
Year

世界及亞洲兩千年來人口趨勢



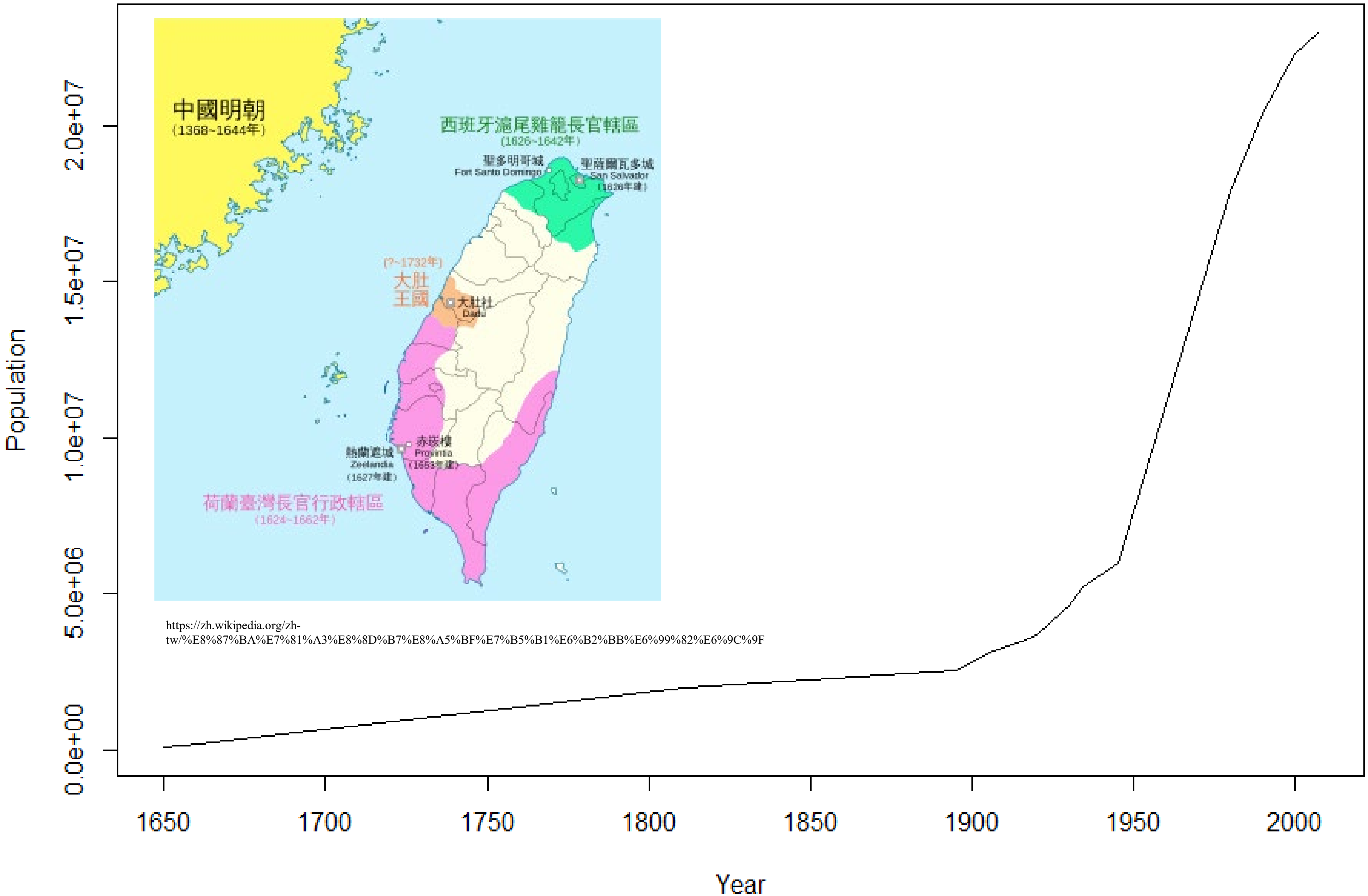
人口爆炸！！

Where's Waldo？？



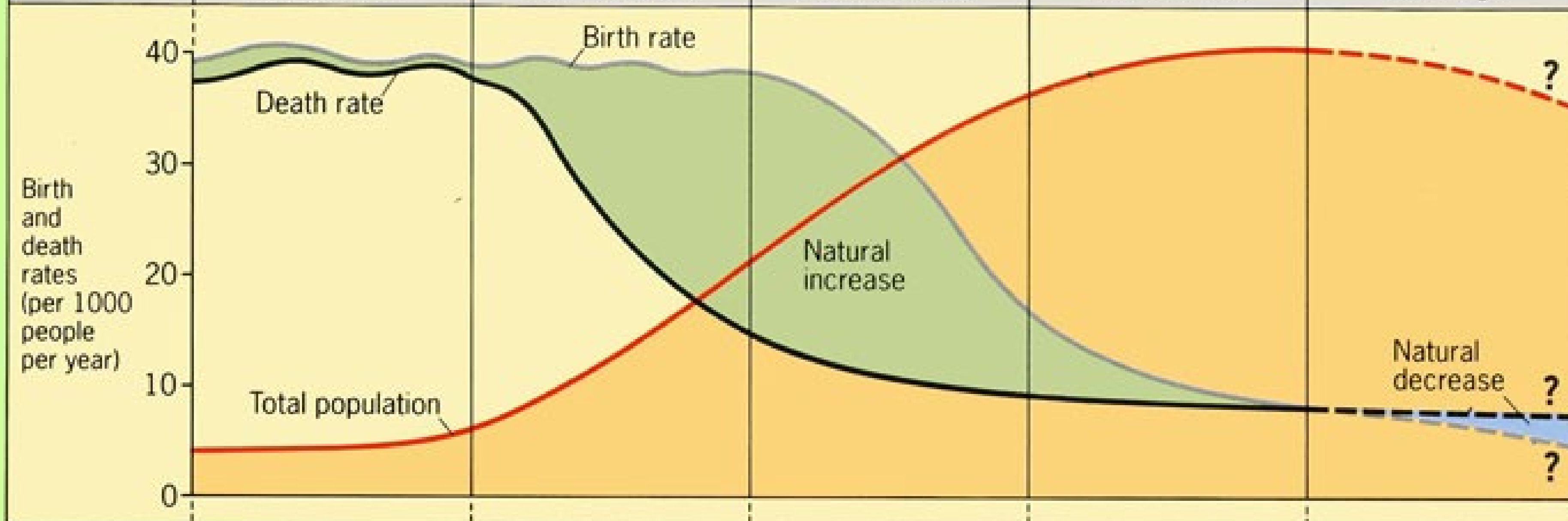
人口往都會地區集中。。。

台灣自荷蘭統治時期的人口數



https://zh.wikipedia.org/zh-tw/%E8%87%BA%E7%81%A3%E8%8D%B7%E8%A5%BF%E7%B5%B1%E6%B2%BB%E6%99%82%E6%9C%9F

人口轉型理論(續)



Stage	1 High stationary	2 Early expanding	3 Late expanding	4 Low stationary	5 ? Declining?
Examples	A few remote groups	Egypt, Kenya, India	Brazil	USA, Japan France, UK	Germany
Birth rate	High	High	Falling	Low	Very low
Death rate	High	Falls rapidly	Falls more slowly	Low	Low
Natural increase	Stable or slow increase	Very rapid increase	Increase slows down	Stable or slow increase	Slow decrease
Reasons for changes in birth rate	Many children needed for farming. Many children die at an early age. Religious/social encouragement. No family planning.		Improved medical care and diet. Fewer children needed.	Family planning. Good health. Improving status of women. Later marriages.	
Reasons for changes in death rate	Disease, famine. Poor medical knowledge so many children die.	Improvements in medical care, water supply and sanitation. Fewer children die.		Good health care. Reliable food supply.	

羅吉斯曲線(Logistic Curve)

- 人口成長可能因自然資源而有限制，不像之前模型中毫無限制。

→較常見的上限條件為 $\lim_{t\to\infty} P(t) = a$ ，又因為瞬間成長率滿足

$$\frac{1}{P(t)} \cdot \frac{dP(t)}{dt} = r_t = k\left[1 - \frac{P(t)}{a}\right] \to 0.$$

瞬間成長率已不再是定值，而將隨時間遞減。

由上式求解模型參數：

$$\frac{d\,P(t)}{P(t)(1-P(t)/a)} = k \cdot dt$$

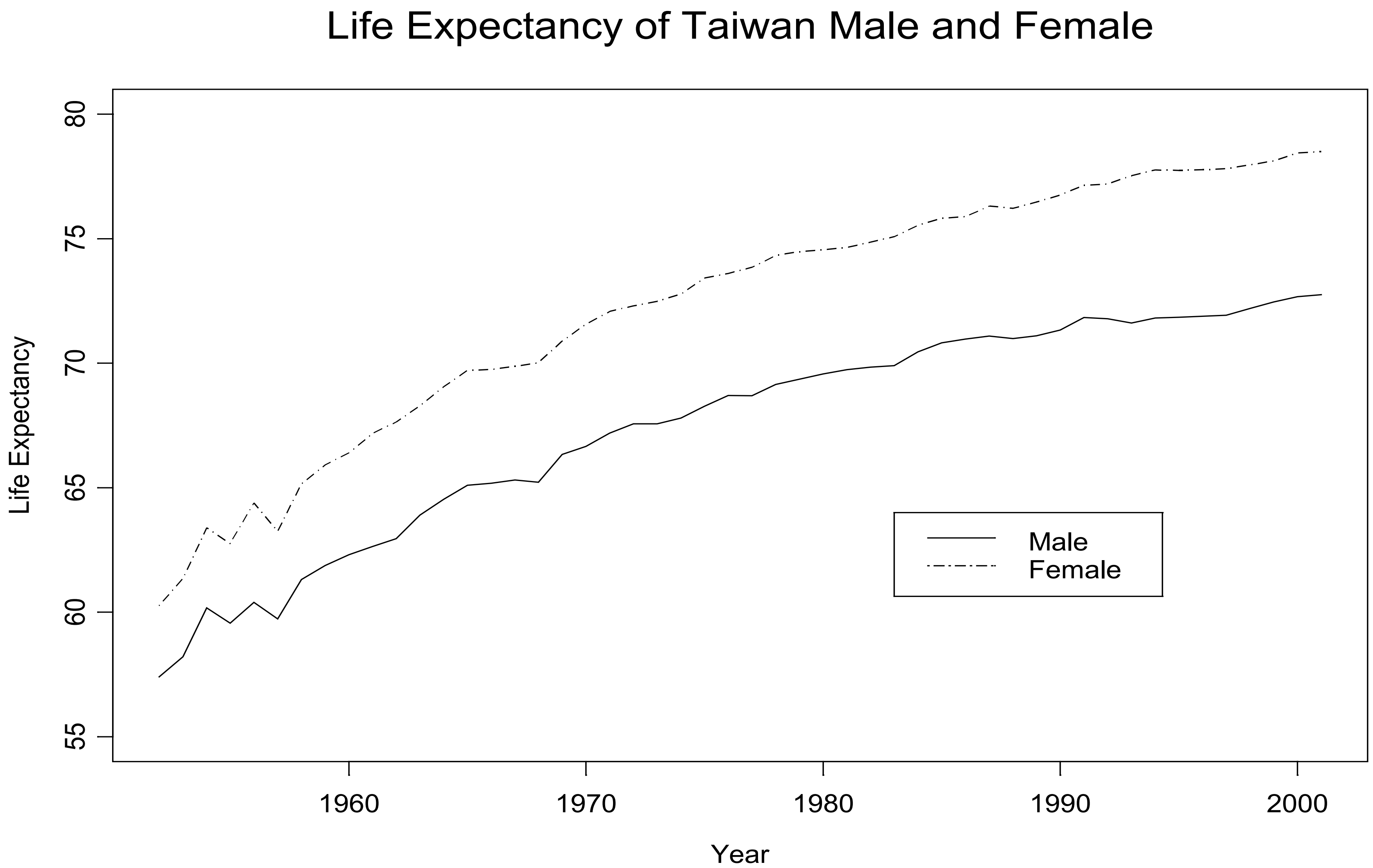
若改寫為

$$\left[\frac{1/a}{1-P(t)/a} + \frac{1}{P(t)}\right] dP(t) = k \cdot dt$$

$$\Leftrightarrow \quad -\ln\left[1-P(t)/a\right] + \ln P(t) = kt + c$$

$$\Leftrightarrow \quad \ln\left[\frac{P(t)}{1-P(t)/a}\right] = kt + c$$

$$\Leftrightarrow \quad P(t) = \frac{1}{A + Be^{-kt}}, \quad i.e., P(\infty) = \frac{1}{A}.$$



由歷年的平均壽命判斷，羅吉斯曲線似乎可用於預測台灣地區居民的壽命。

羅吉斯曲線估測曲線模式

(資料來源：內政部統計處, 2007/1/22修正)

→臺灣地區男性零歲平均餘命：

$$78 / [1 + EXP (0.414457 - 0.036731t)]$$

$$R\text{-square} = 0.988587$$

→臺灣地區女性零歲平均餘命：

$$82 / [1 + EXP (0.664580 - 0.036843t)]$$

$$R\text{-square} = 0.966273$$

■範例三：若P(0) = 10、k = 1.25%、a = 25 (單位：百萬)，試以羅吉斯曲線估計P(10)、P(50)。

→由已知條件可求得 $A = \frac{1}{a} = 4 \times 10^{-8}$另一參數 B 則可由下式求出：

$$B = \frac{1}{P(0)} - A = 6 \times 10^{-8}.$$

因此，
$$P(t) = \frac{1}{(4 + 6e^{-0.0125t}) \times 10^{-8}}.$$

P(10) = 10.76、P(50) = 13.87 (百萬)。

- 範例四：美國歷年戶口普查人口數字得知 P(70) = 205.1、P(80) = 227.7、P(90) = 249.9 (百萬)，估計羅吉斯模型假設下的人口上限。

→由羅吉斯模型的定義可知

$$a = \frac{\dfrac{1}{C_1} + \dfrac{1}{C_3} - \dfrac{2}{C_2}}{\dfrac{1}{C_1 C_3} - \dfrac{1}{C_2^2}}, \text{其中} \mathrm{P}(t_i) = C_i, t_1 < t_2 < t_3.$$

代入已知條件，

$$a = \frac{\dfrac{1}{C_1} + \dfrac{1}{C_3} - \dfrac{2}{C_2}}{\dfrac{1}{C_1 C_3} - \dfrac{1}{C_2^2}} = 420.4(\text{百萬})。$$

年輪組成法（Cohort Component Method）

- 選定一年為基年，以基年的單一年齡人口乘上該年齡的存活率，分別推計下一年度的男、女各年齡人口：

$$P_{x+1}(t+1) = P_x(t) - D_x(t)。$$

0歲部份使用生育率來推計未來出生人數。(Whelpton, 1936)

$$B(t) = \sum_{x=15}^{49} {}_5P_x^f(t) \times {}_5f_x(t)$$

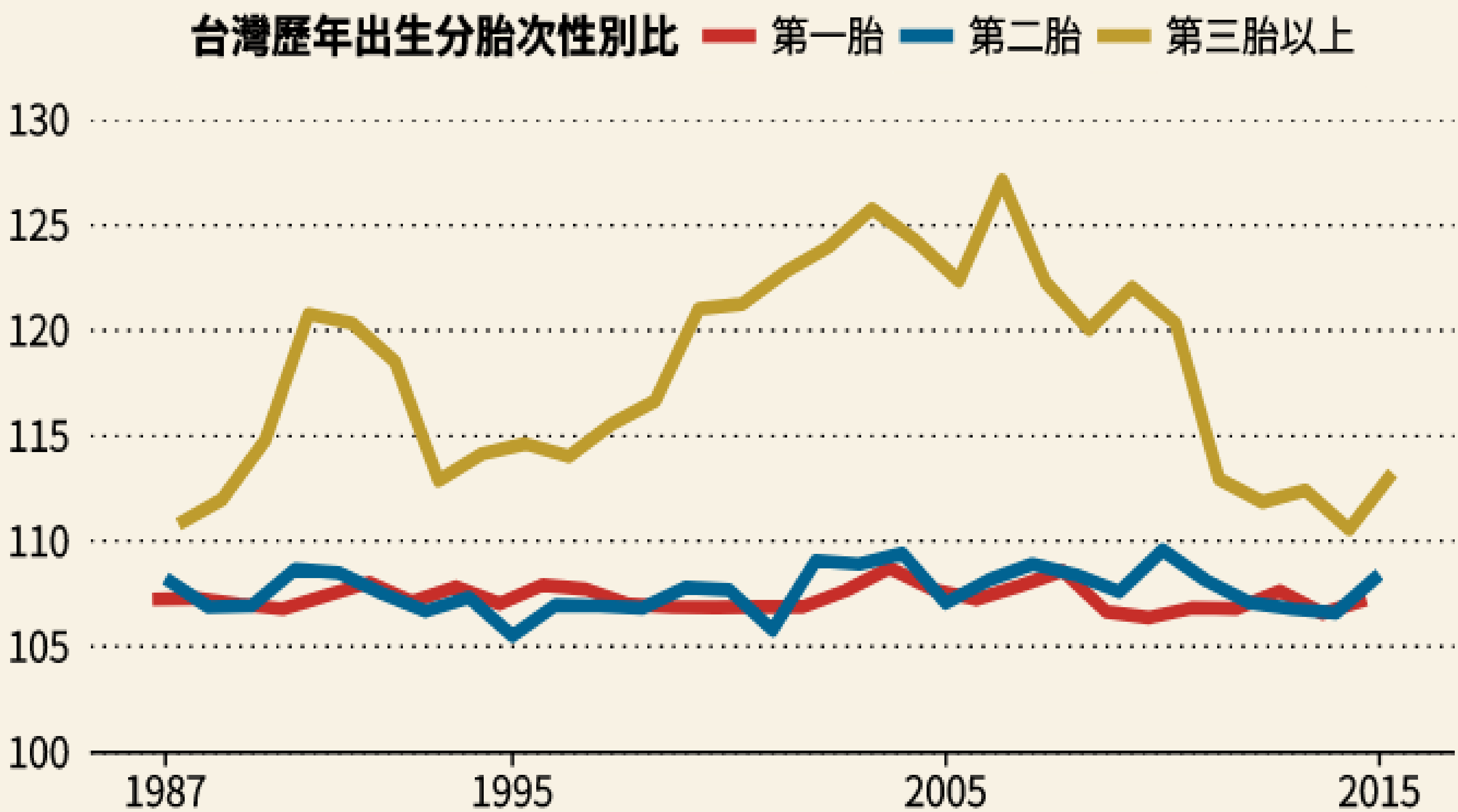
- 人口學家利用數學的方法，將人口推估所需的成份（Component）以一矩陣來表示，稱為Leslie矩陣（Leslie Matrix），記做 M。
- 第一列中的元素表示單一年齡育齡婦女的生育率，只在15至49歲有值，其餘為零。而對角線以下的元素則為男、女生單一年齡的存活率，表示今年t歲的人，乘上存活率可得明年t+1歲的人口數，其估計可寫成：

$$M X_t = X_{t+1}$$

$$\begin{bmatrix} 0 & 0 & \cdots & f_{15}^{t} & \cdots & f_{49}^{t} & 0 & \cdots \\ p_0^f & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots \\ 0 & p_1^f & \cdots & 0 & \cdots & 0 & 0 & \cdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \times \begin{bmatrix} P_0(t) \\ P_1(t) \\ P_2(t) \\ P_3(t) \\ \vdots \end{bmatrix} = \begin{bmatrix} P_0(t+1) \\ P_1(t+1) \\ P_2(t+1) \\ P_3(t+1) \\ \vdots \end{bmatrix}$$

- Leslie矩陣分別推估男性、女性人口數。其中男性出生數可由女性出生數，乘以出生性別比(Sex Ratio)，一般假設是1：1.05，臺灣過去曾超過1：1.1，最近幾年已下降至1：1.08。

臺灣歷年男女嬰出生性別比（胎次別）



http://talkecon.com/content/images/2016/06/sex_ratio-2.svg

註：消失的六萬女嬰！https://talkecon.com/missing_women_taiwan/

關於Leslie Matrix的故事

- Leslie (1900-1974)將年齡別死亡率及生育率以矩陣型態表示，建立以矩陣型態推估未來人口總數及其結構，延續之前人口推估研究的成果（例如：Whelpton）。

→Leslie任職於牛津(Oxford)的Bureau of Animal Population，在1945年提出Leslie Matrix，用於預測未來的動物總數。

- 某地區推估人數的增減可由Leslie矩陣，或是矩陣的特徵值(Eigenvalue)看出。因為

$$M^k X = M^{k-1}(MX) = M^{k-1}\sum_{i=1}^{n}\lambda_i v_i$$

$$\to \lambda_1^k v_i \ \ as\ k \to \infty,$$

其中 λ_i、v_i 為矩陣 M 的第 i 個特徵值、特徵向量。(令 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$)

→如果 $\lambda_1 > 1$ 則未來人口將增加；反之，未來人口將減少。但若特徵值為虛根時，則需透過矩陣乘積計算未來人口。

- 對應於Leslie矩陣最大特徵值的特徵向量v_1，將是該地區到達穩定時(參閱上一章Sharpe-Lotka定理)的各年齡層的人口結構。

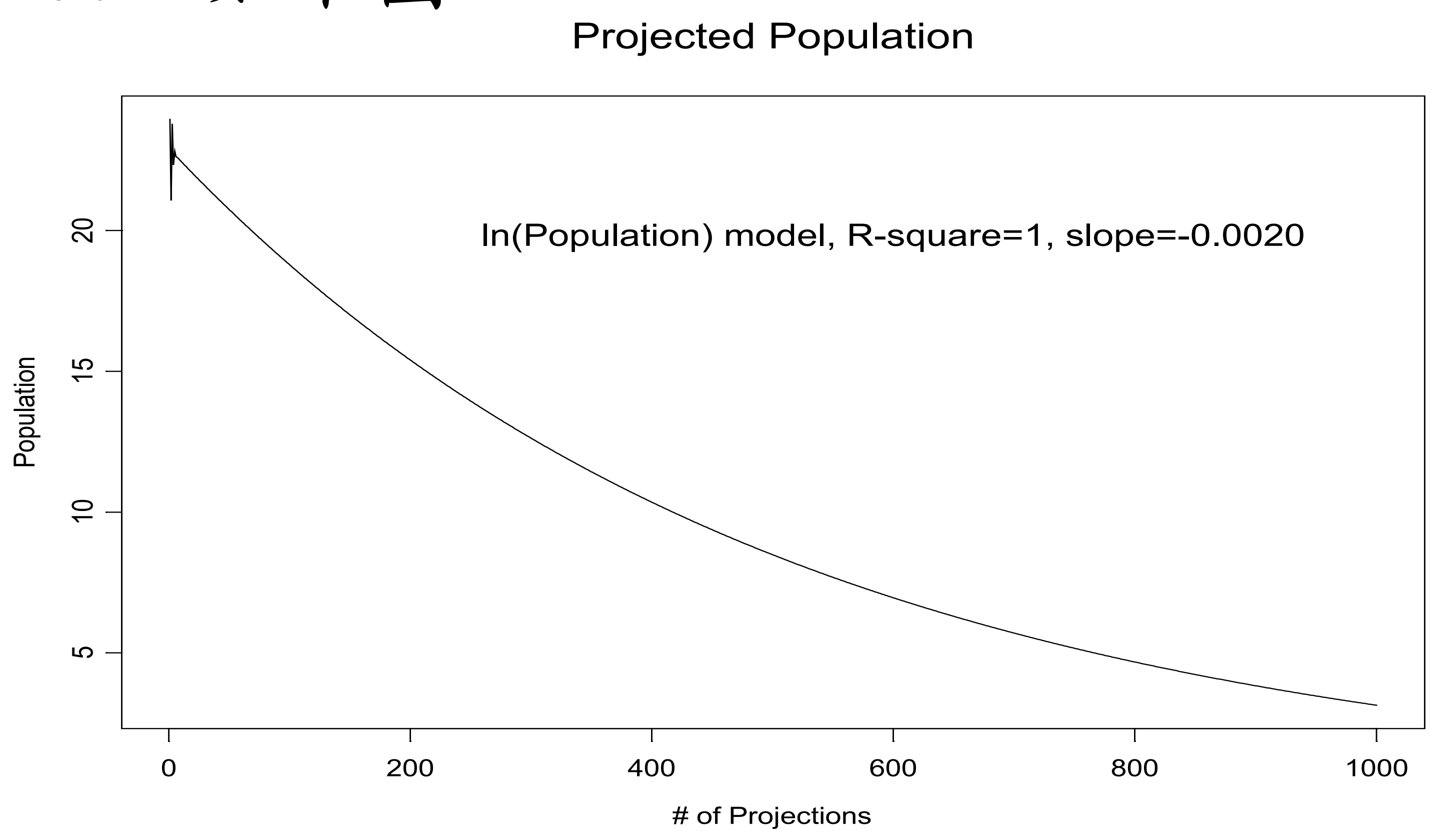
範例五：某地的兔子總數可由下列的Leslie矩陣推估，試求穩定時的年齡結構。

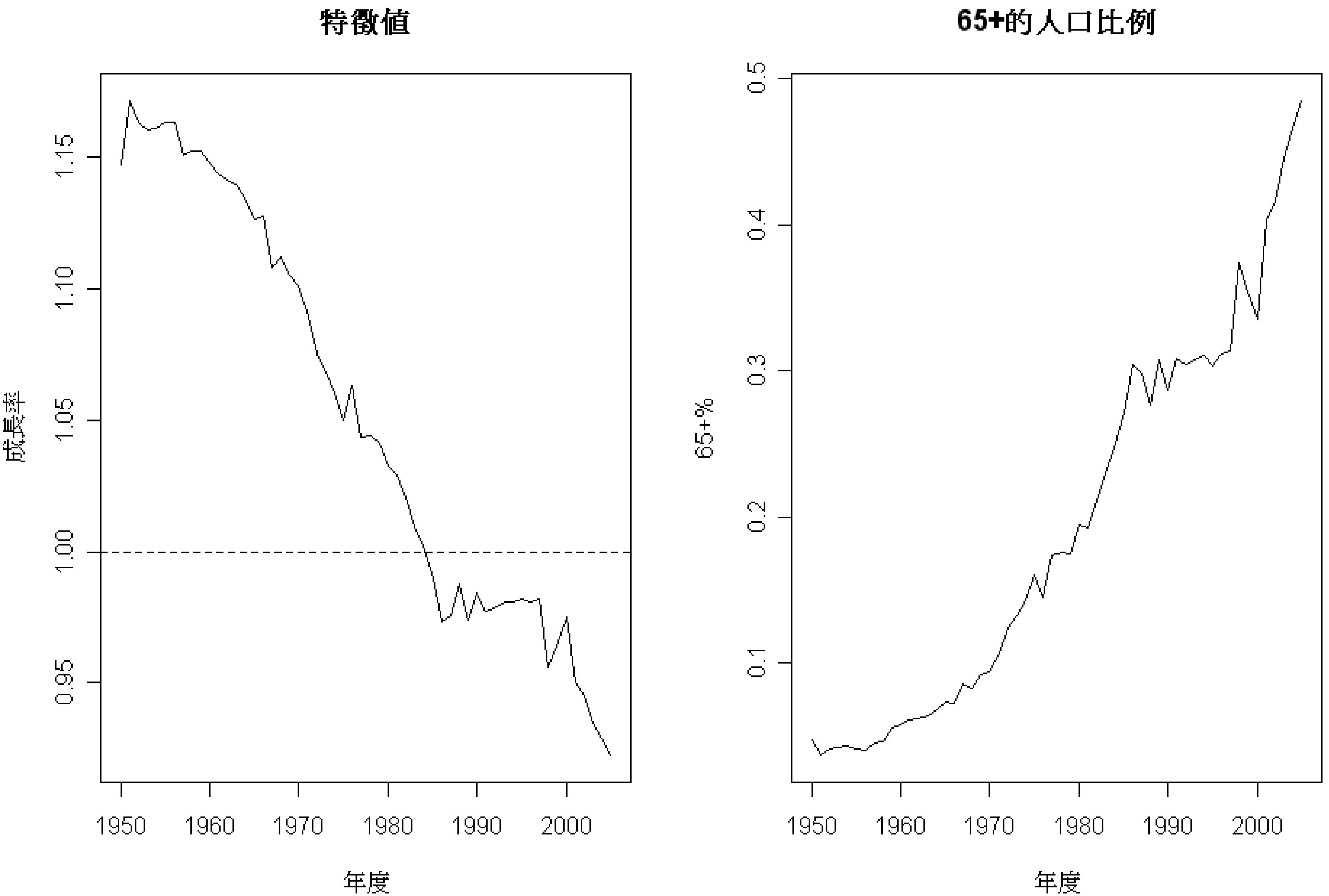
$$M=\begin{bmatrix} .3360 & .5152 & .1506 \\ .9946 & 0 & 0 \\ 0 & .9880 & 0 \end{bmatrix}$$

→代入$X=(5,10,15)$，即 0, 1, 2 歲的兔子比例各為1:2:3。

範例五(續)

→因為 M 的特徵值為虛根，只能藉由矩陣乘積求得穩定時的總數及結構。經過一千期計算得出 $X = (1.054, 1.050, 1.039)$，每年約下降0.2%，如下圖。





臺灣地區自1950年以來的人口成長（特徵值）

生育、死亡、遷移的假設

- 因為生育率與死亡率是Leslie矩陣的最重要參數，不同假設衍生出截然不同的推估結果。

→經歷經濟高速成長、進入已開發國家之林，臺灣生育率及死亡率逐年下降。然而，生育率自2010年後回升至1.2左右，未來趨勢仍是未定之數。

→然近年臺灣遷出遷入較為頻繁，2008年行政院經建會(現為國發會)已將遷移列入考量。

2024～2070年臺灣人口推估(生育率假設)

- 女性教育程度與就業機會提升
- 婚育價值觀的改變

遲婚、遲育

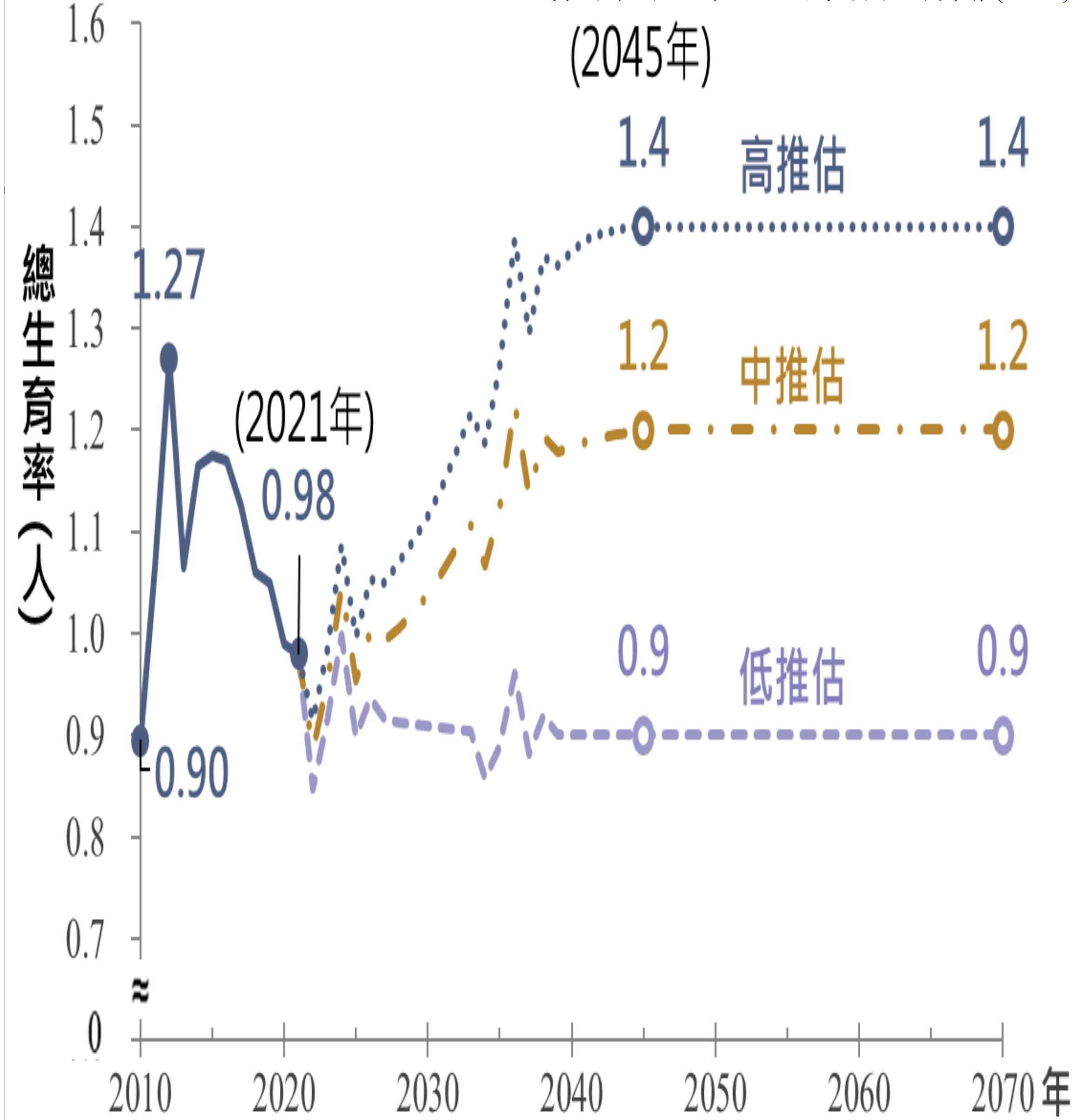
1991年至2021年
女性初婚由26.0歲延至30.4歲
生第一胎由25.5歲延至31.2歲

縮短育齡婦女生育期間

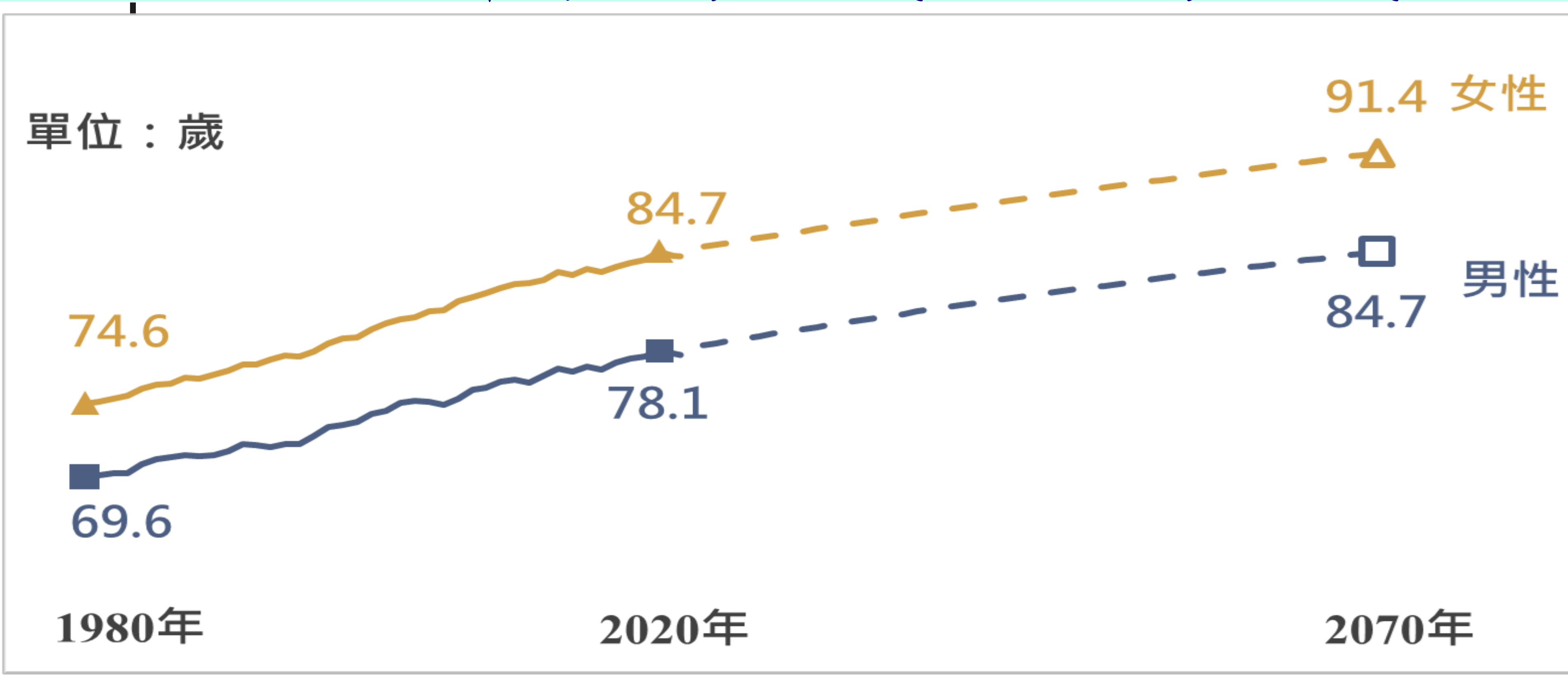
不婚、不育

生育數量
減少

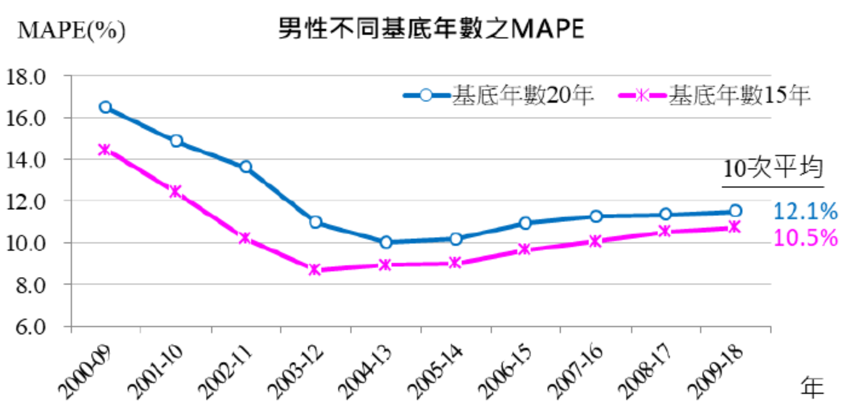
資料來源：行政院國家發展委員會(2024)



死亡假設：利用Gompertz模型及Lee-Carter模型，推估至2070年男性為84.7歲，女性為91.4歲。



■ 推估相關因素：「資料品質」、「地區人數」、「資料年數」與「推估年數」。



社會增加假設 (外國人戶籍遷入+本國人戶籍遷入/遷出)

1 疫情前穩定淨移入

2017-2019年平均每年淨移入女性8千人，男性3千人

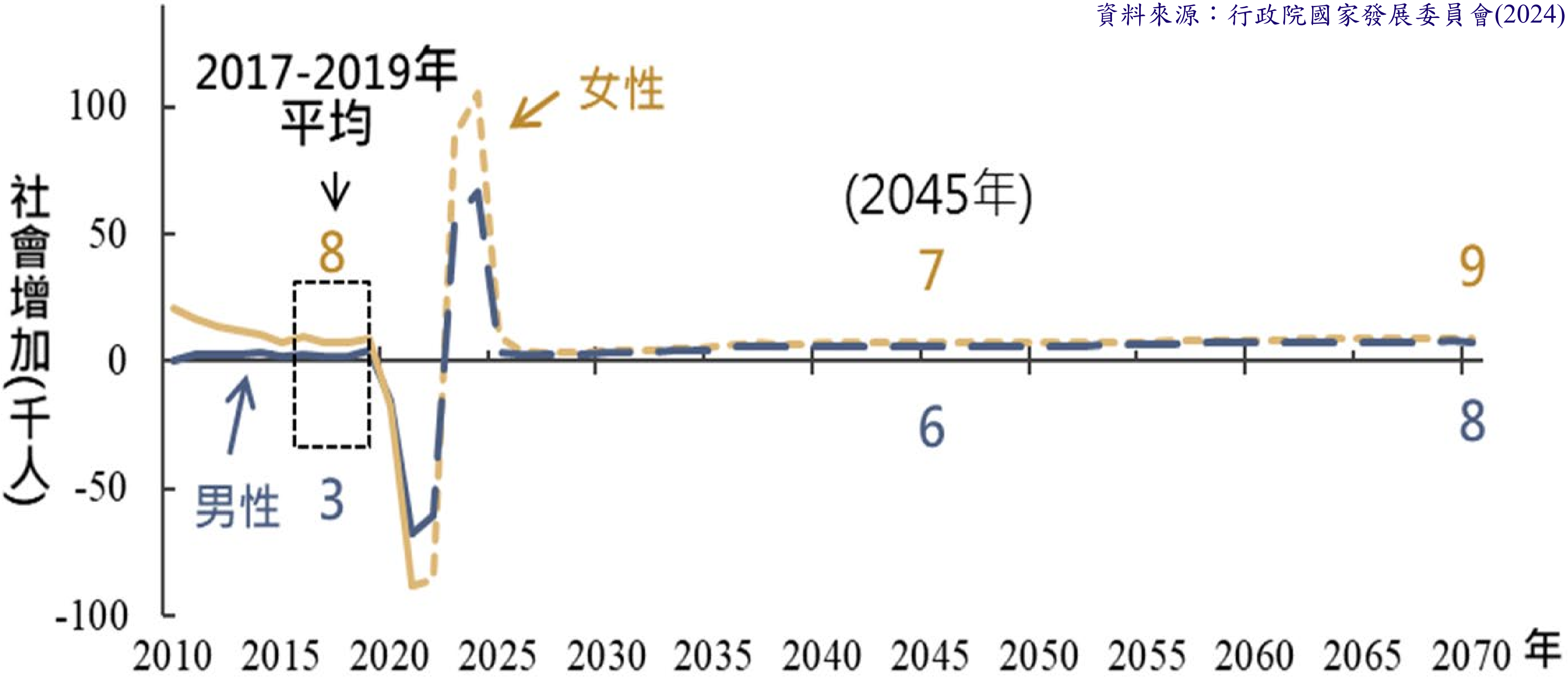
2 疫情後邊境開放，社會增加回流

- 外國人：假設疫情期間受影響者於2022-2024年回流9成
- 本國人：假設疫情期間戶籍被遷為遷出者於2023-2024年回流8成

3 長期趨勢回歸穩定淨移入

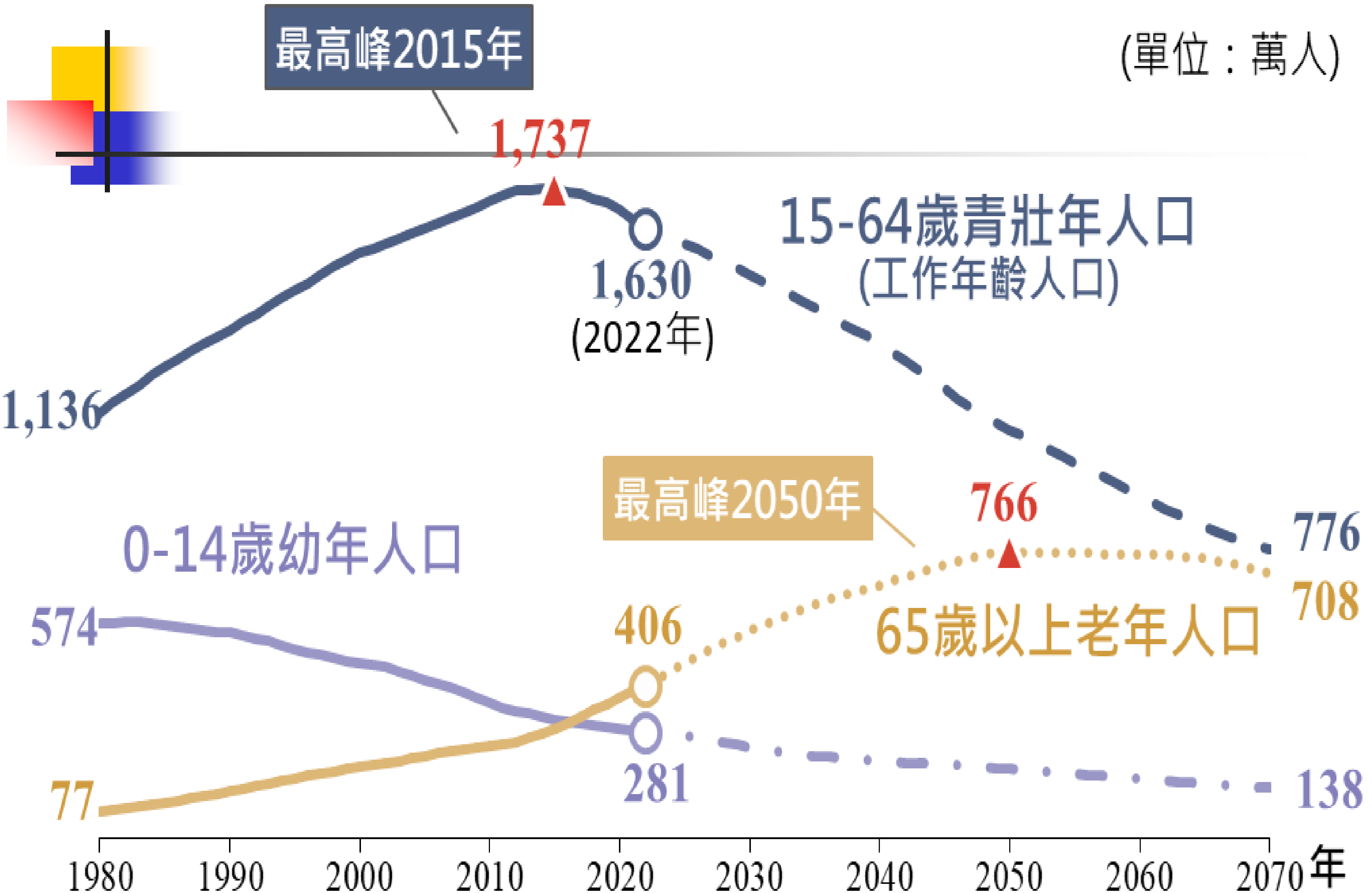
- 外國人：假設2045年後男性增至9千人，女性增至1.1萬人
- 本國人：依年齡別社會增加率推估

資料來源：行政院國家發展委員會(2024)



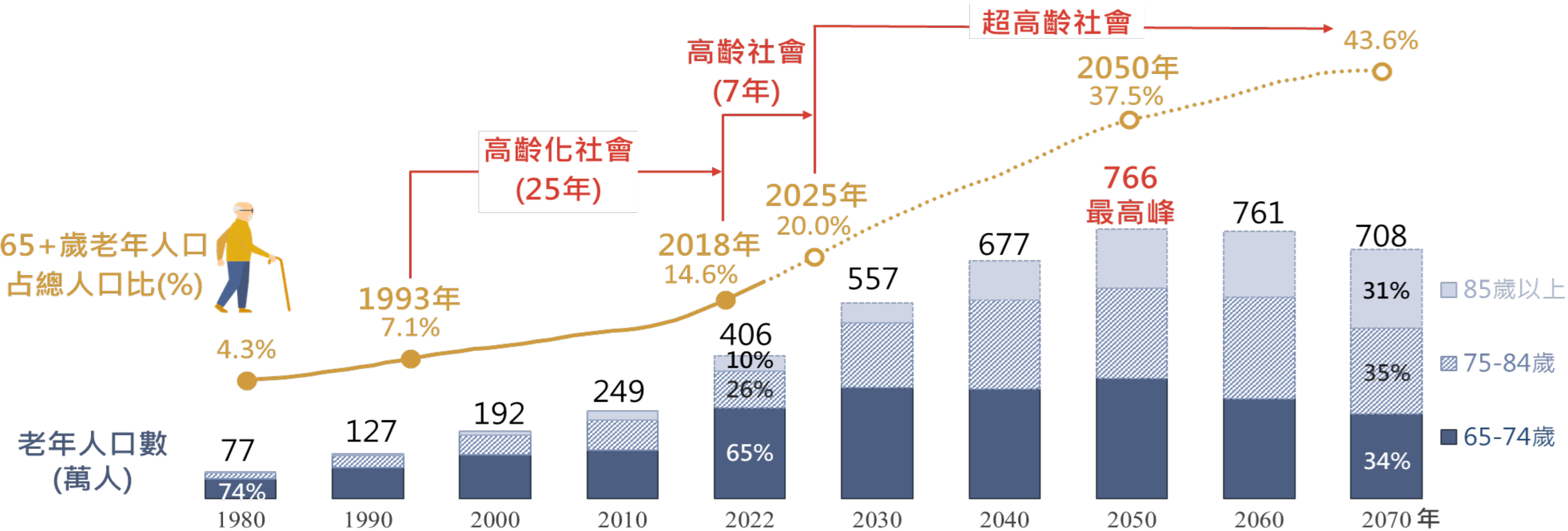
未來人口老化趨勢更加明顯(中推計)

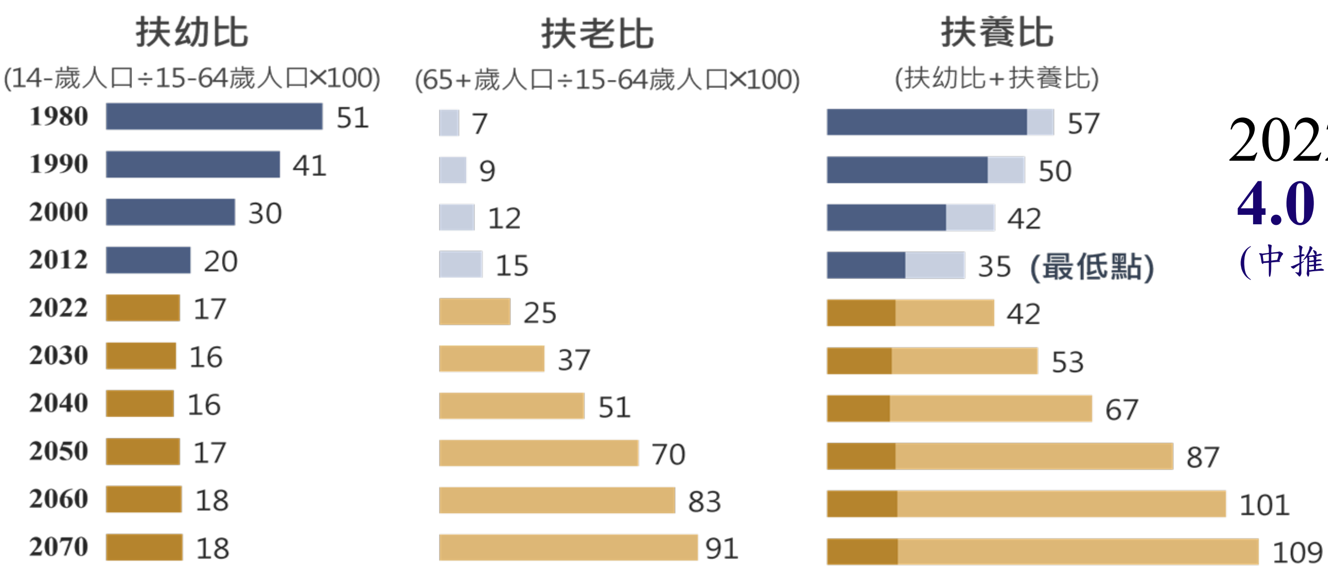
資料來源：行政院國家發展委員會(2024)

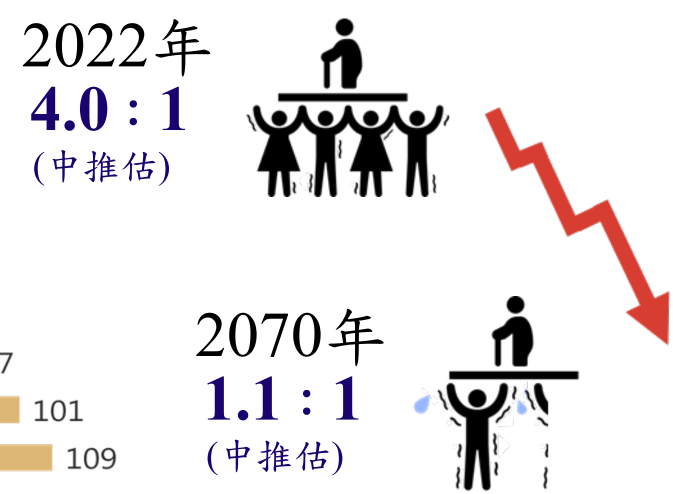


扶養比（扶老比）增加快速(中推計)

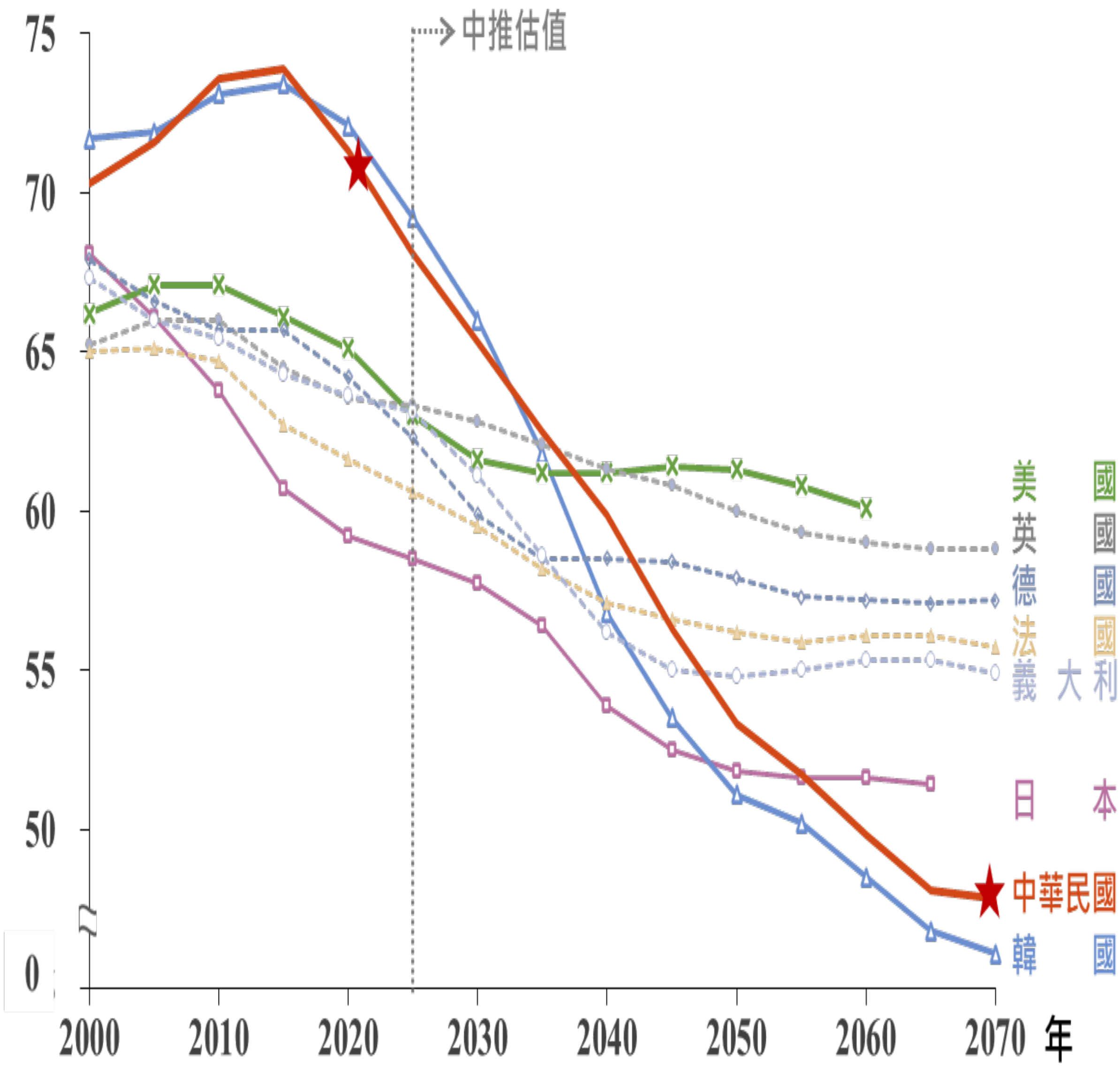
資料來源：行政院國家發展委員會(2024)







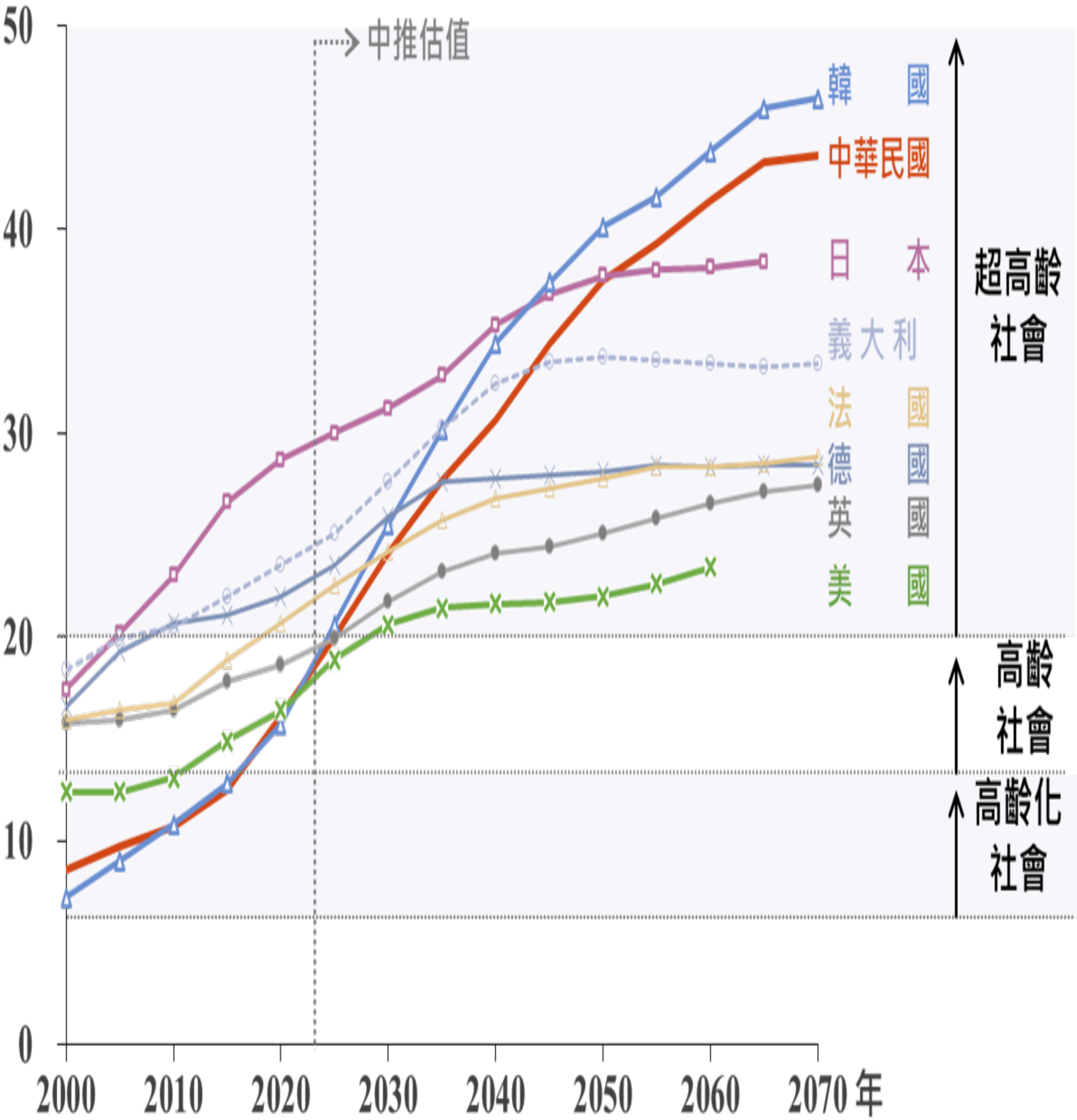
世界各國工作年齡人口占比



國別		2020年	2070年
韓　國	🇰🇷	72.1	46.1
中華民國	🇹🇼	71.3	47.8
美　國	🇺🇸	65.1	60.1 (2060年)
德　國	🇩🇪	64.2	57.2
義大利	🇮🇹	63.6	54.9
英　國	🇬🇧	63.5	58.8
法　國	🇫🇷	61.6	55.7
日　本	🇯🇵	59.2	51.4 (2065年)

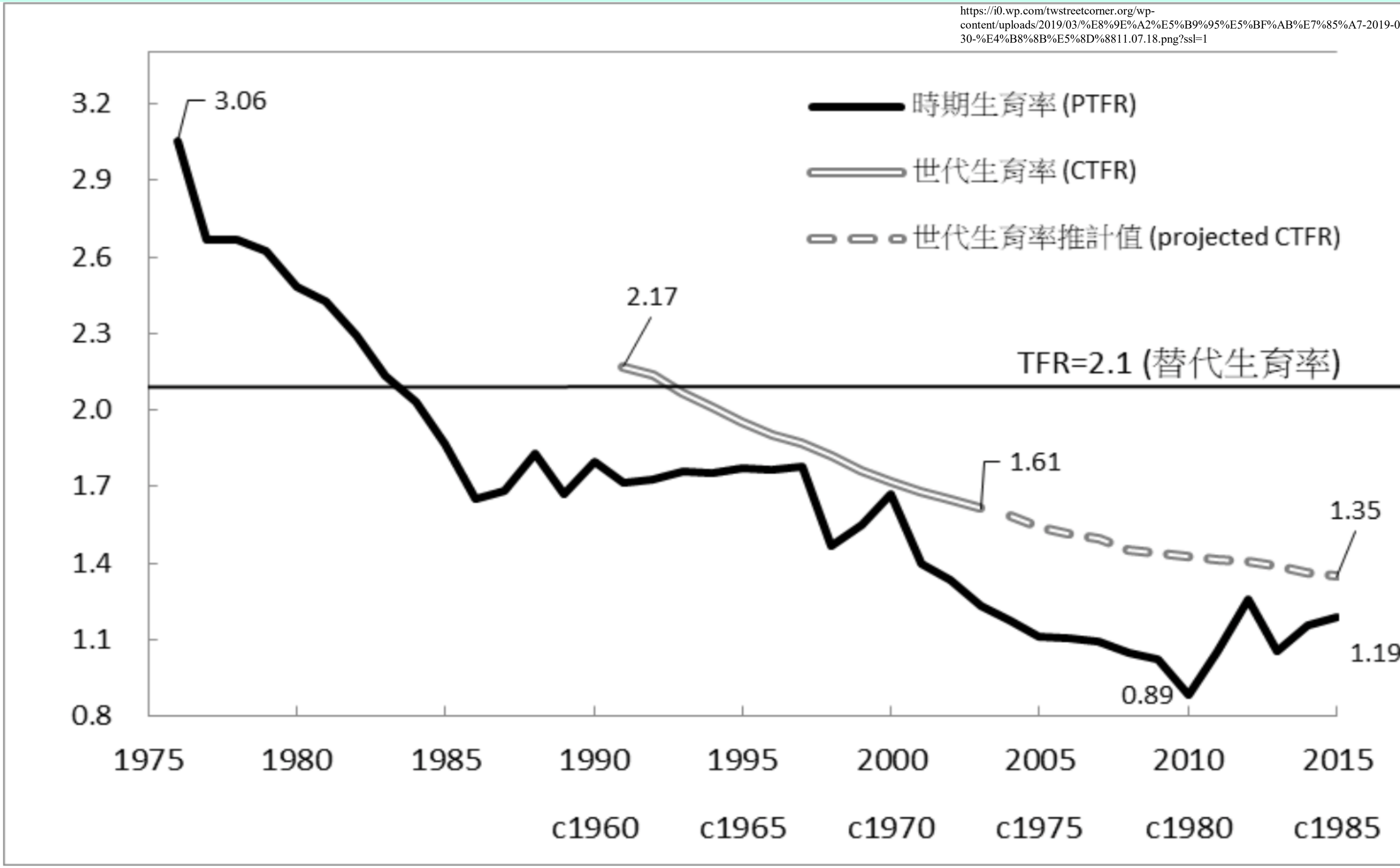
資料來源：行政院國家發展委員會(2025)

世界各國65歲以上人口占比



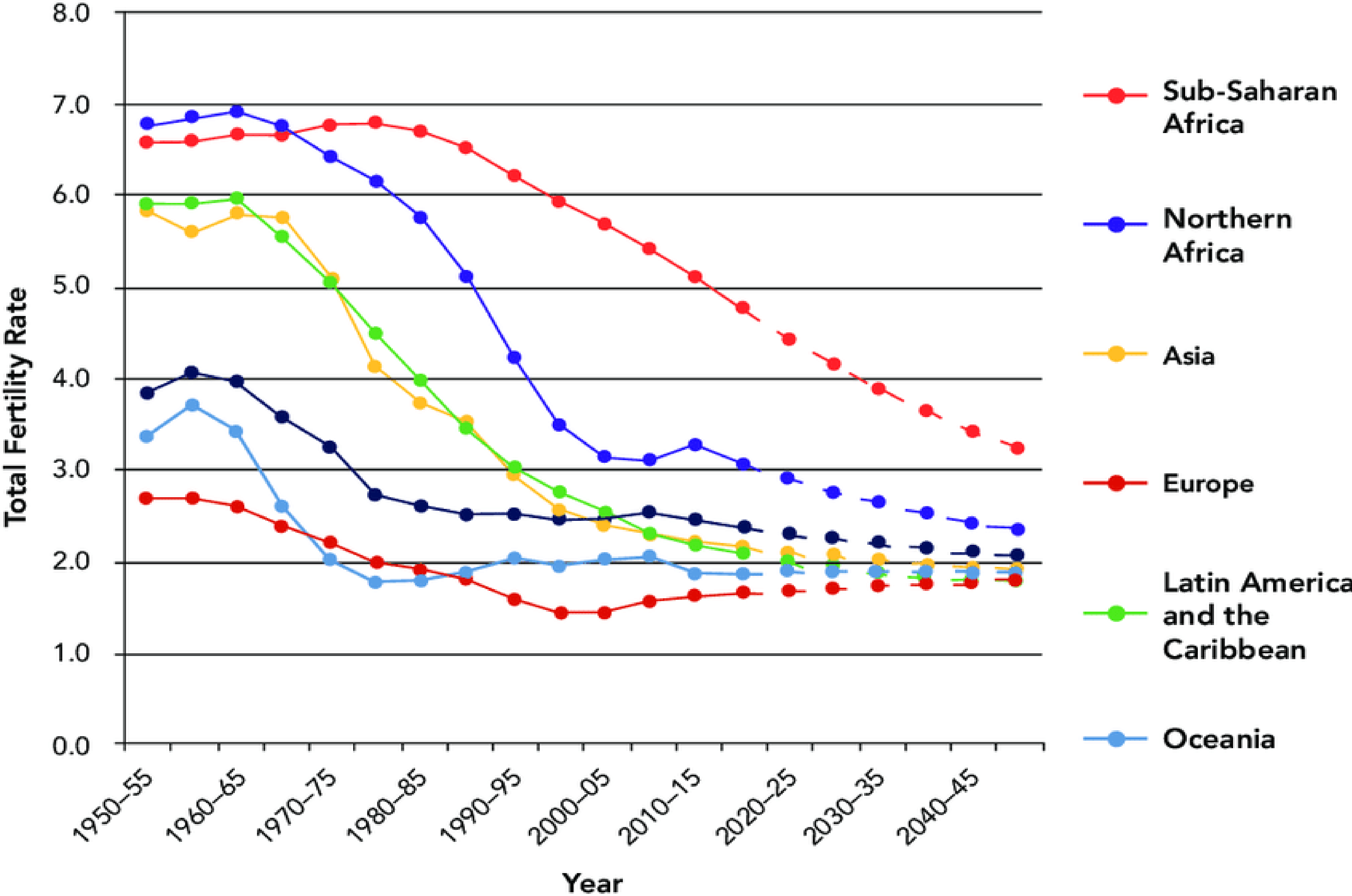
國別	2020年	2070年
韓　國	15.7	46.4
中華民國	16.1	43.6
美　國	16.4	23.4(2060年)
英　國	18.6	27.4
法　國	20.7	28.8
德　國	22.0	28.4
義大利	23.5	33.4
日　本	28.7	38.4(2065年)

臺灣歷年總生育率

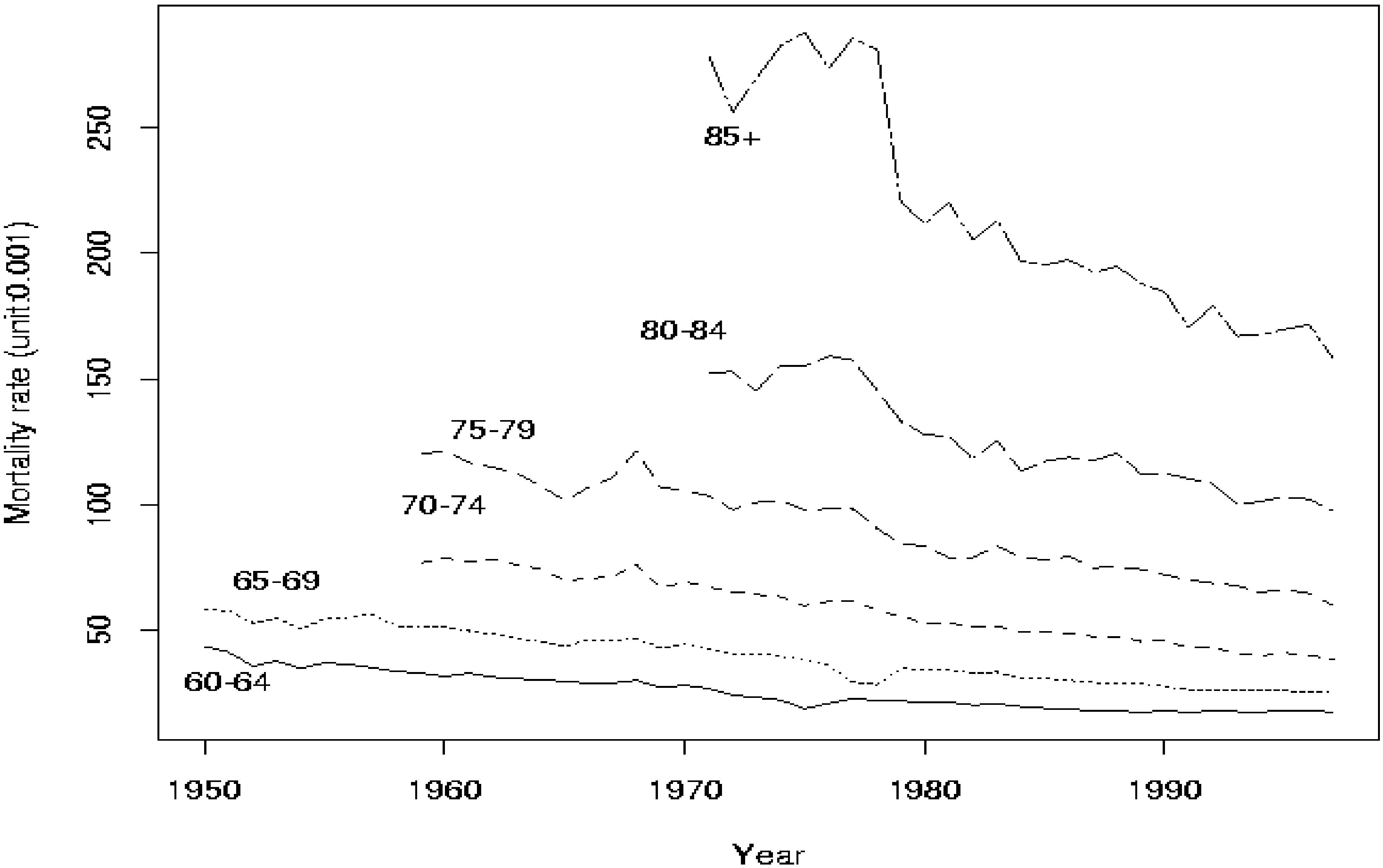


資料來源：Human Fertility Database。C 代表世代（cohort），c1960 指 1960 年出生之女性。CTFR 推計方法請見 Myrskylä, Goldstein, and Cheng (2013) "New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals", Population and Development Review, 39(1): 31-56.

世界各國歷年及推估之總生育率



台灣地區男性死亡率(高齡部份)



- 因為台灣地區生育率與死亡率的數值變動幅度很大，一般的參數(Parametric)模型的預測效果普遍較差，建議可採用半參數(Semi-parametric)或無母數(Non-parametric)的方法。

→較常見的方法有多變量分析中的主成份法(Principal Component Analysis)，以及獨立成分法(Independent Component Analysis)。近年廣受使用的Lee-Carter模型可視為(一個)主成份法的特例。

推估(Projection)與預測(Forecast)

- 「推估」通常指定對未來的假設，可視為在某些假設條件成立下的條件計算值(Conditional Calculation)。
- 「預測」指的是最有可能情形下的結果。
- 國內的人口推估大致可定義為情境推估，以民國95年至140年的推估為例，除了一套死亡與遷移假設，生育率的假設有高、中、低三種推計。

情境推估的隱憂

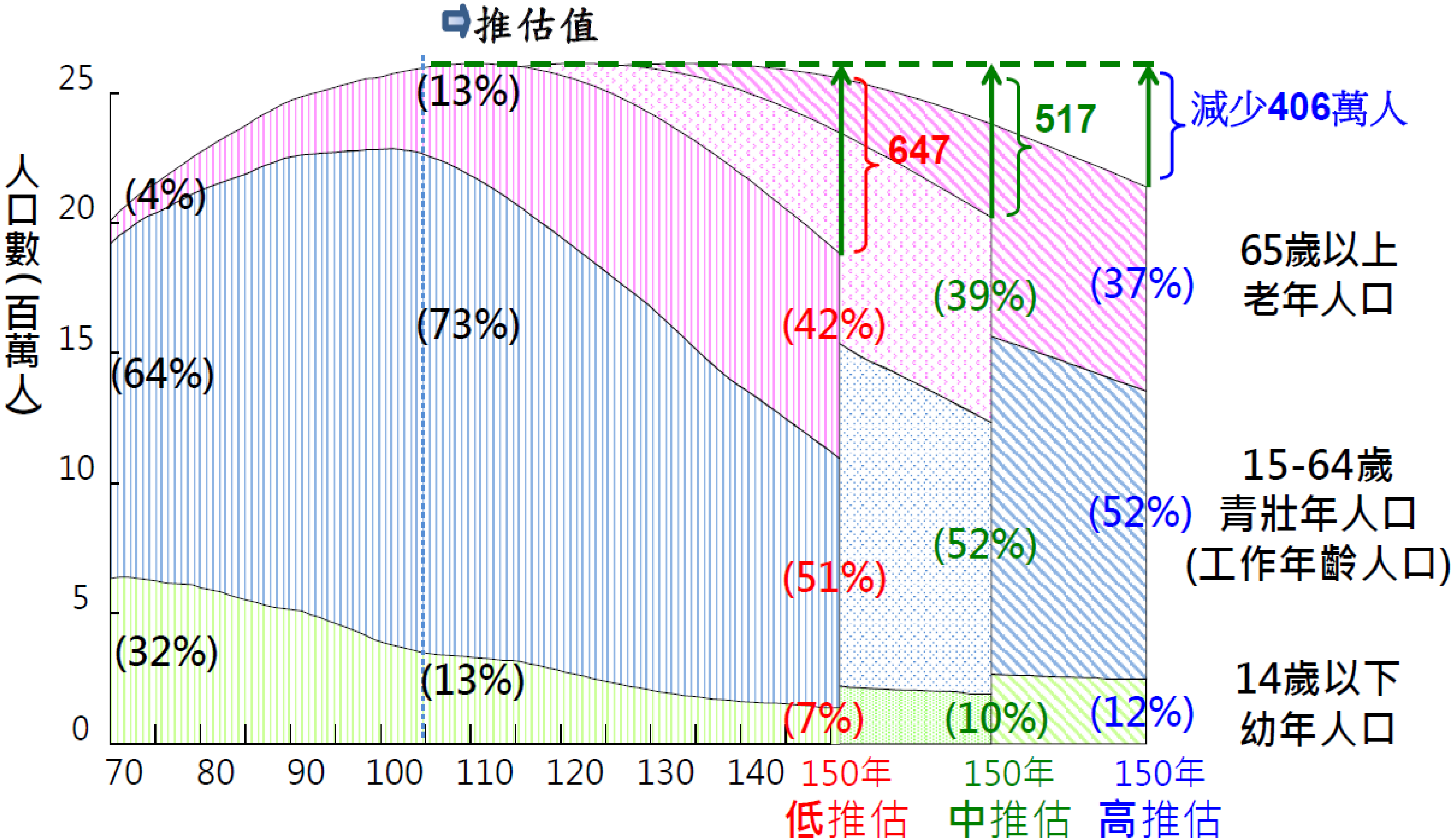
- 近年生育率和死亡率的下降迅速，增加推估的困難，以專家意見或政府施政期許來訂定未來可能，或有改進的空間。
- 情境推估無法提供推估值的發生可能，限制了人口推估的機率詮釋。
- 不同情境給予未來生育、死亡、遷移等數值的可能數值，未考慮變數之間的相關性。

專家意見

- 聯合國及歐美各國的人口推估，仍以年輪組成法為主，但未來的生育率與死亡率多依賴專家意見，缺乏機率上的涵義。
- 近年有三種機率人口推估，可用來彌補專家意見的不足。
  1. 隨機推估(Stochastic Forecast)
  2. 模擬情境(Random Scenario Method)
  3. 推估誤差法(ex post Method)

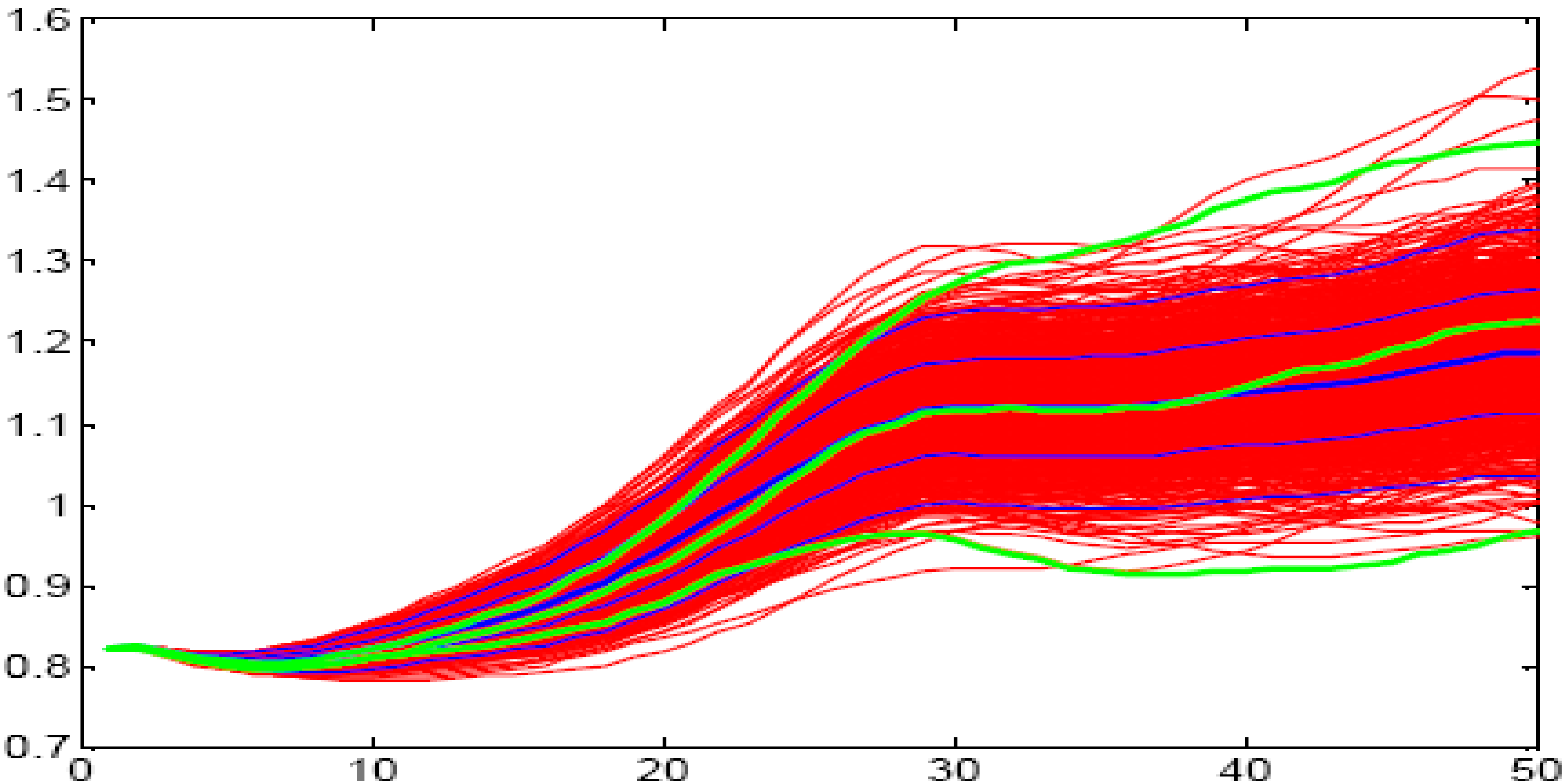
傳統方法推估之呈現：(專家意見)

國發會（前身為經建會）民國105年公佈的人口推估



註：推估分成高、中、低三種推估情境。

隨機方法推估之呈現：(電腦模擬)



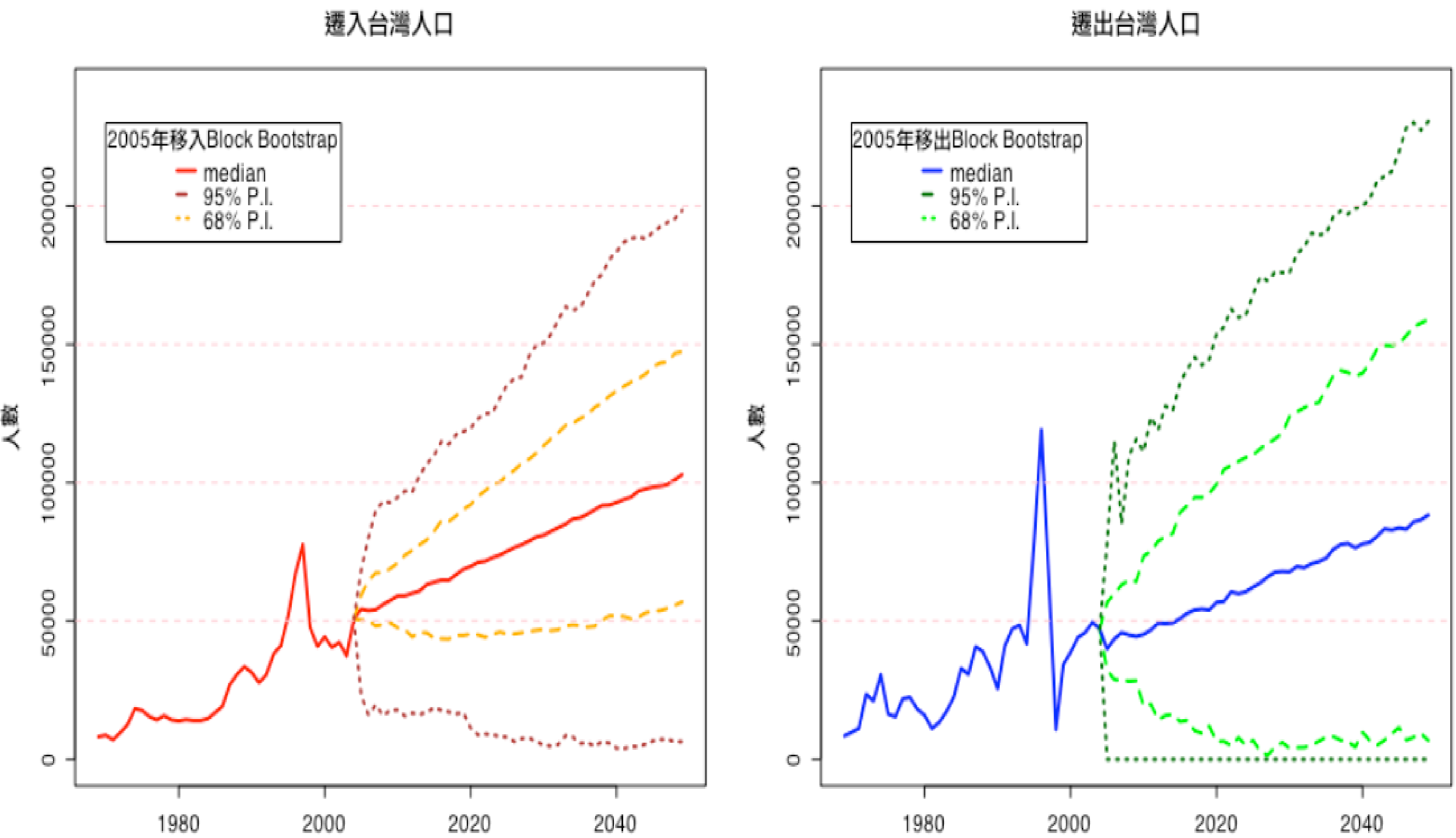
Red:　德國 2002-2050年依賴人口比之模擬路徑

Green: 加總後中位數、最小值及最大值對應之路徑

Blue:　平均值、 68%預測區間及95%預測區間

資料來源：Bertz & Lipps (2004)

隨機方法推估之呈現：(區塊拔靴法)



- 引進隨機推估用於台灣地區的人口推估

資料來源：李芯柔、余清祥 (2008)

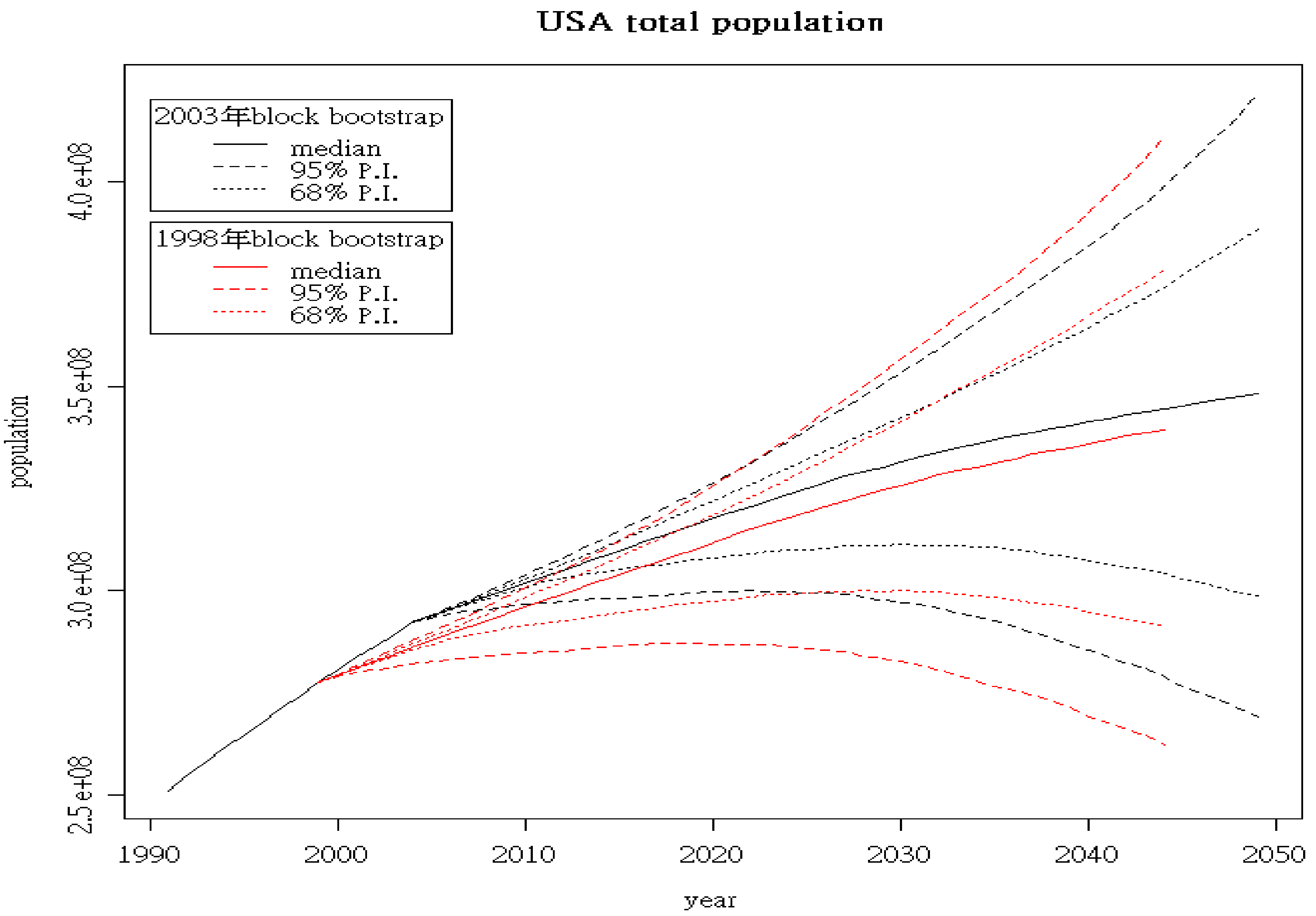
區塊(Block)拔靴法

- 區塊拔靴法的重複抽樣方式類似一般的拔靴法，只是每次抽取一個「區塊」的資料，當資料服從均衡(Stationary)假設時，區塊拔靴法大多都適用。

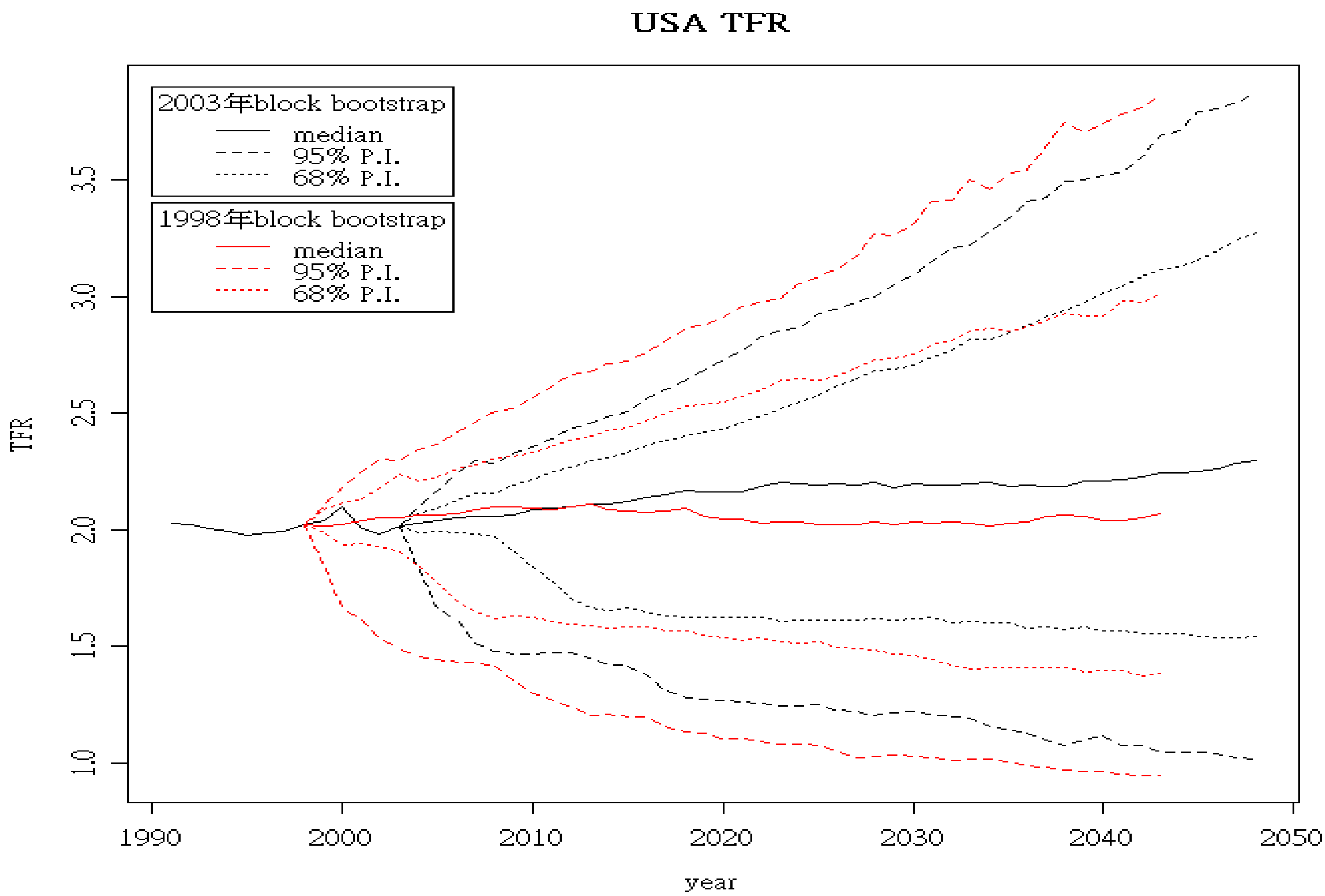
→區塊拔靴法用於相關資料有不錯的效果，雖然不如獨立樣本時一般拔靴法的準確，但比Subsampling效果好，而且不需要資料滿足很強的條件，加上操作時不需對資料給予任何假設，實證上是很好的選擇。

- 區塊拔靴法的抽樣方式類似拔靴法，但不是對觀察值直接抽樣，而是對相鄰觀察值的差異抽樣，而且抽取時將連續一串的差異值抽出。例如：若區塊長度為b、且抽到第 k 個觀察值，則第 k 個至第$k+b-1$個差異值被抽出，最後一個觀察值加上這些差異值即為預測值。
- R的模組「boot」也有處理時間數列資料(相關資料的一種)的功能，細節可查閱「tsboot」指令。(這個指令可指定區塊長度為定值、或是服從幾何分配。)

美國總人口預測

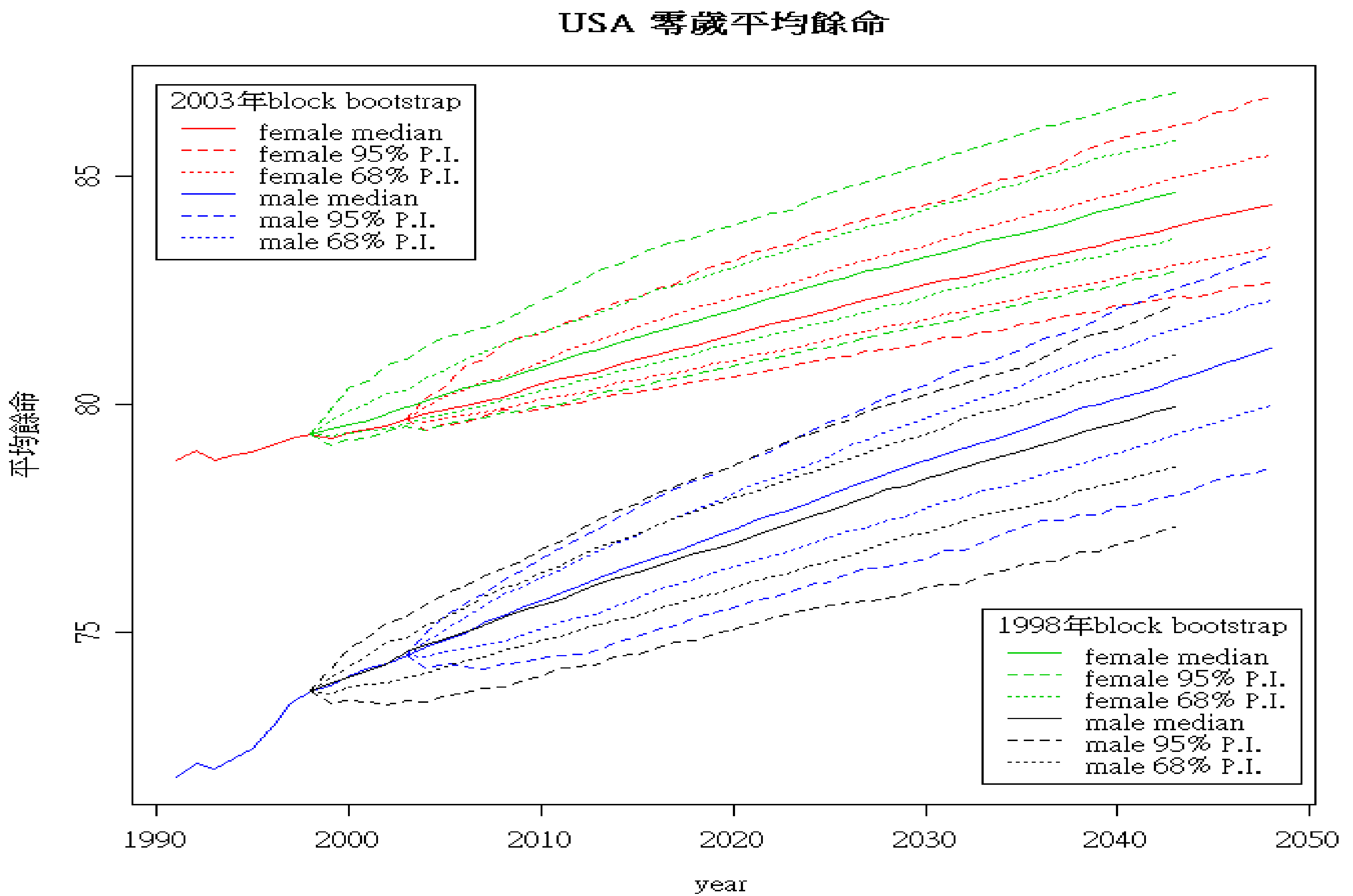

USA total population
2003年block bootstrap
median
95% P.I.
68% P.I.
1998年block bootstrap
median
95% P.I.
68% P.I.
population
2.5 e+08
3.0 e+08
3.5 e+08
4.0 e+08
1990
2000
2010
2020
2030
2040
2050
year

美國總生育率預測

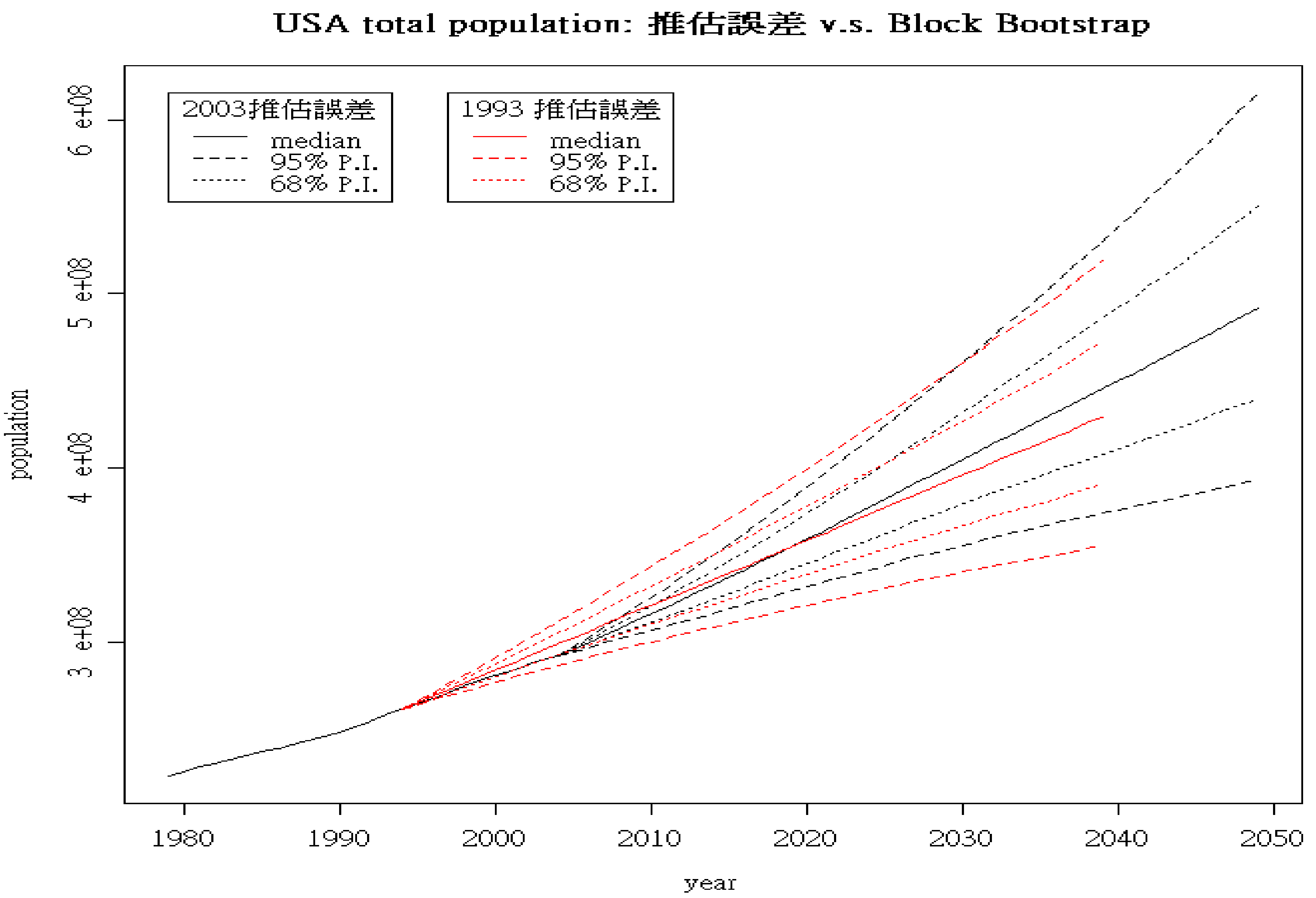


USA TFR
2003年block bootstrap
median
95% P.I.
68% P.I.
1998年block bootstrap
median
95% P.I.
68% P.I.
TFR
3.5
3.0
2.5
2.0
1.5
1.0
1990
2000
2010
2020
2030
2040
2050
year

美國零歲平均餘命預測


USA 零歲平均餘命
2003年block bootstrap
female median
female 95% P.I.
female 68% P.I.
male median
male 95% P.I.
male 68% P.I.
1998年block bootstrap
female median
female 95% P.I.
female 68% P.I.
male median
male 95% P.I.
male 68% P.I.
85
80
75
平均餘命
1990
2000
2010
2020
2030
2040
2050
year

美國人口推估（Stoto修正）

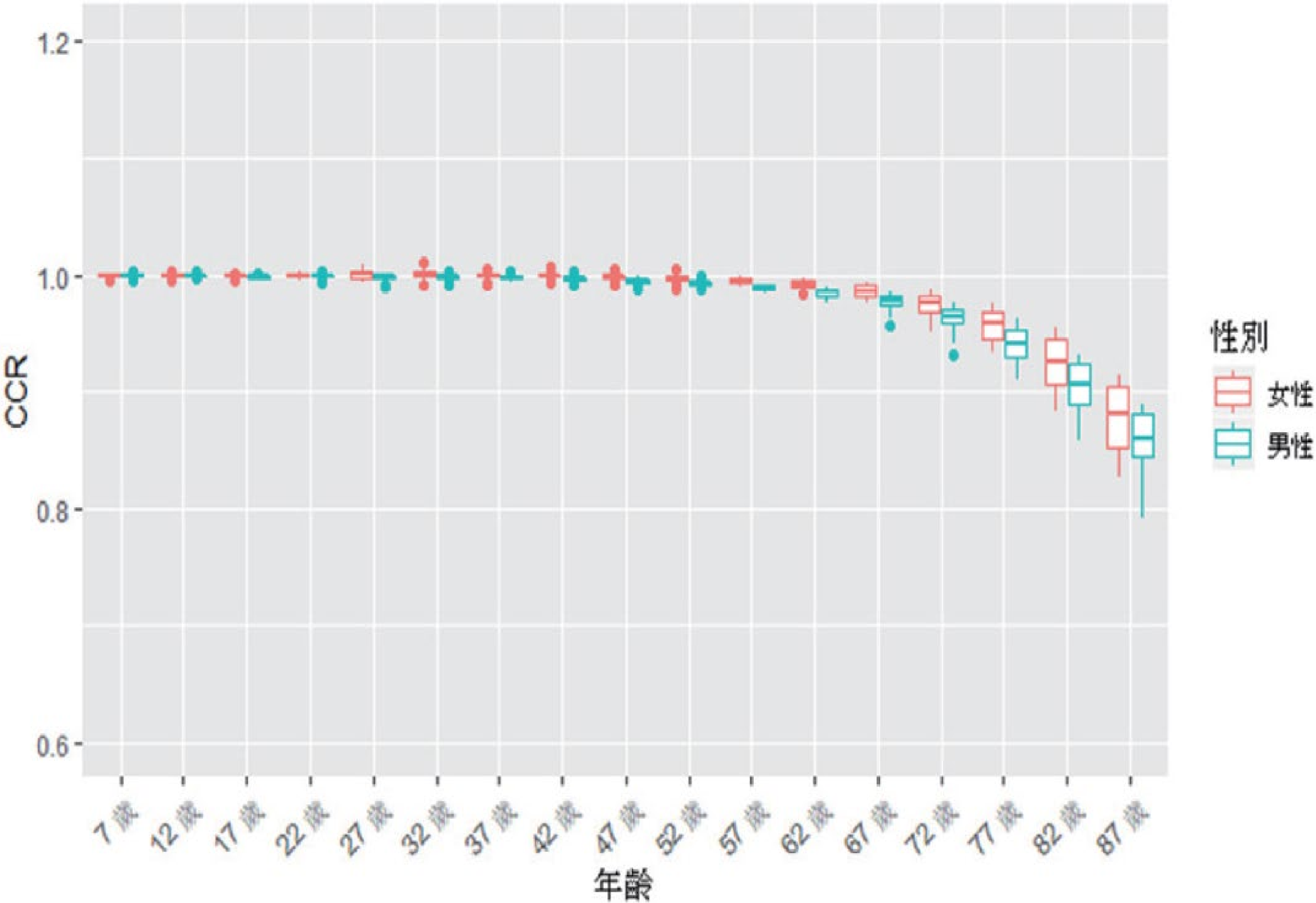


小區域(Small Area)人口推估

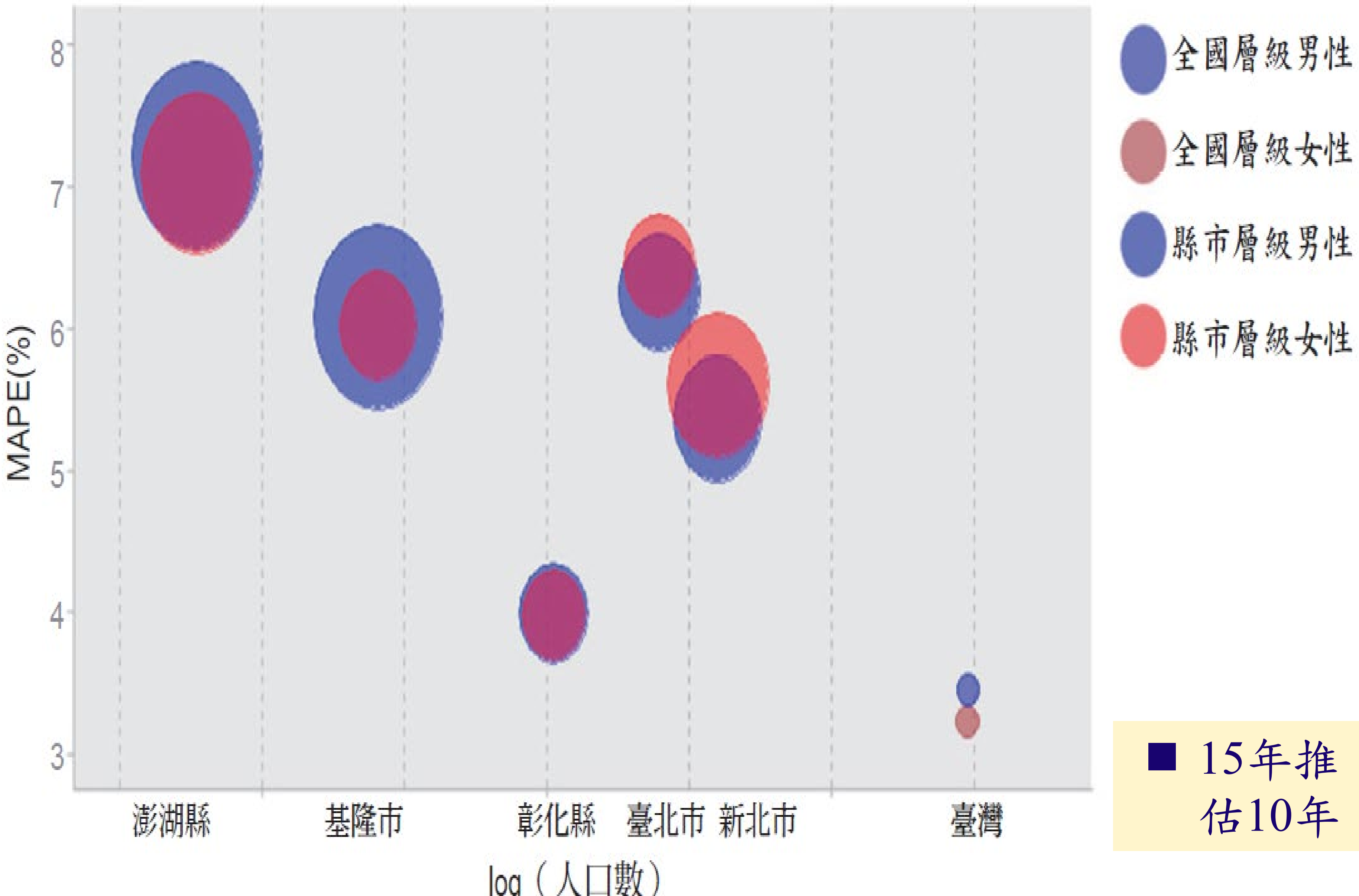
- 年輪組成法為我國官方的人口推估方法，需要詳細生育、死亡、遷移等資料，很難直接套用至縣市層級及以下之人口推估。
- 小區域推估可採Hamilton and Perry（簡稱HP法）的年輪變動比（Cohort Change Ratio；簡稱CCR），在此以臺灣各級行政區域為研究區域，驗證HP法是否能用於推估臺灣縣市、鄉鎮市區層級的人口及結構。

→單齡vs.五齡組、1975～2019年。

臺灣歷年5～84歲年輪變動比(CCR)

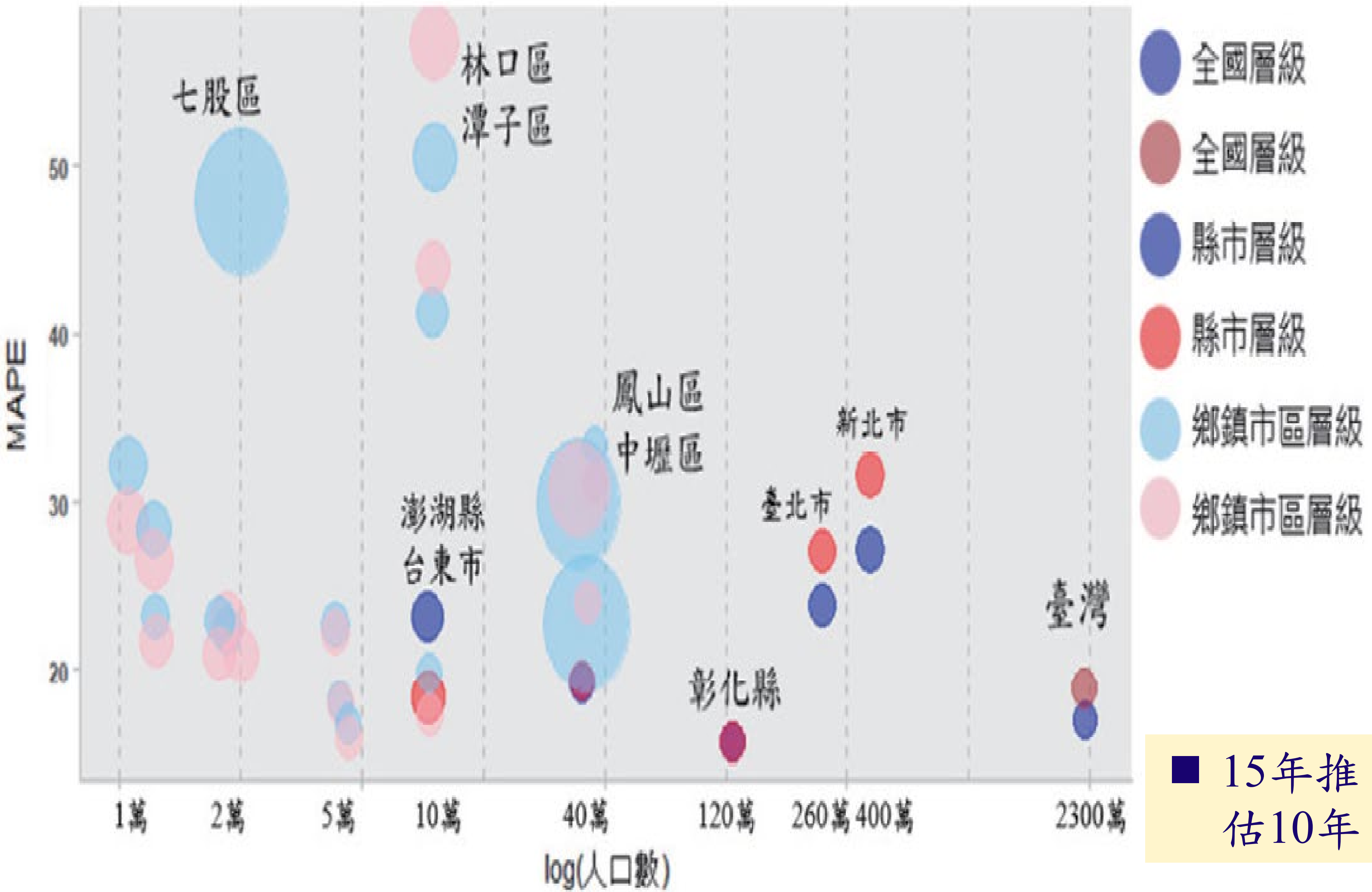


臺灣縣市HP法單齡推估誤差（1975～2019年）



■ 15年推估10年

臺灣HP法五齡推估誤差（1975～2019年）



■ 15年推估10年

參考文獻：

- 陳政勳與余清祥(2010),「小區域人口推估研究：臺北市、雲嘉兩縣、澎湖縣的實證分析」，《人口學刊》
- 王信忠、金碩、余清祥(2012),「小區域死亡率推估之研究」，《人口學刊》
- Wang, H., Yue, C. J., and Chong, C. (2018), "Mortality Models and Longevity Risk for Small Populations," *Insurance: Mathematics and Economics*
- Yue, C.J., Wang, H., & Wang, T. (2021), "Using Graduation to Modify the Estimation of Lee-Carter Model for Small Populations", *North American Actuarial Journal*
- 余清祥、王信忠、陳譽騰(2021)，「年輪變動比用於小區域人口推估的探討」，《人口學刊》