

人口統計(Demography)

授課教師：余清祥教授

日期：2023年5月10日

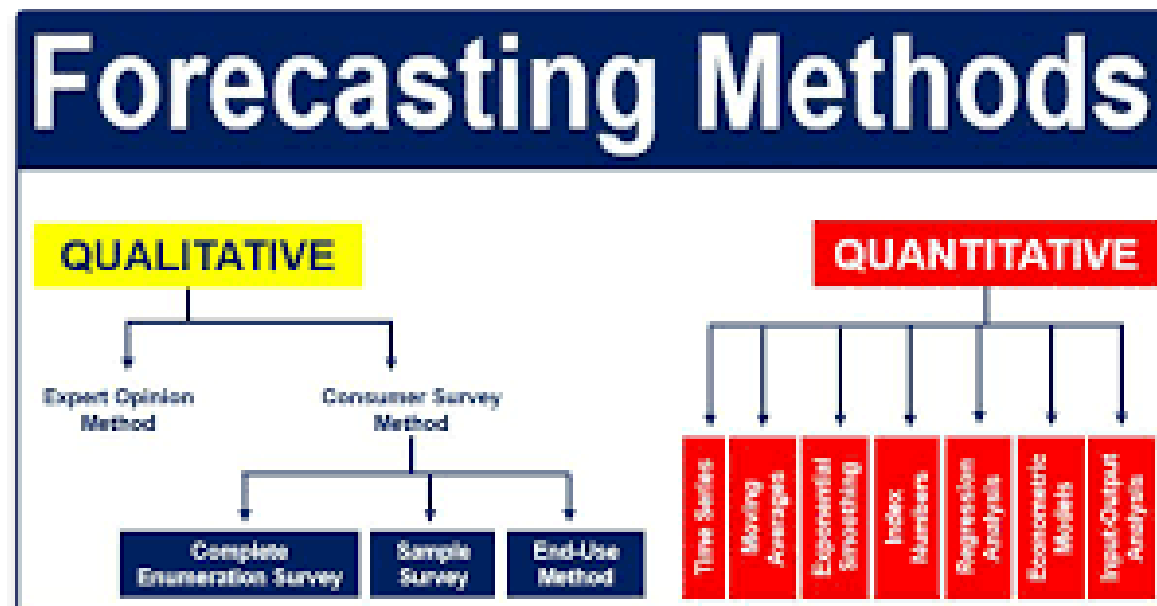
第七講：人口推估

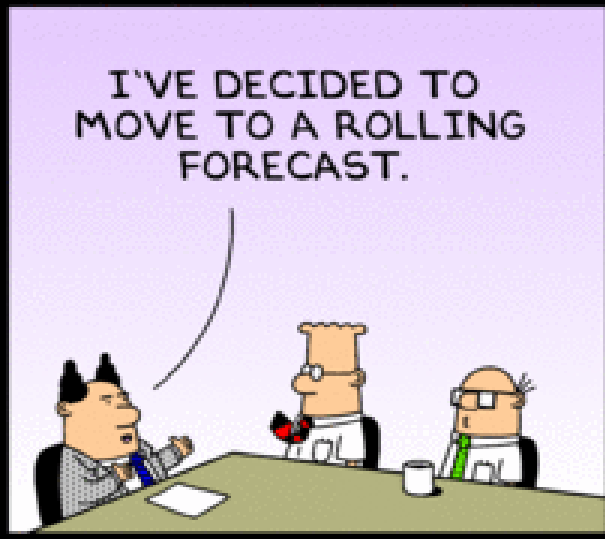
<http://csyue.nccu.edu.tw>



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

--- Samuel Goldwyn





Dilbert.com DilbertCartoonist@gmail.com



7-2-10 © 2010 Scott Adams, Inc./Dist. by UFS, Inc.



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



© 1994 Scott Adams, Inc. All Rights Reserved



© 1994 Scott Adams, Inc. All Rights Reserved



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)

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

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

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

$$P(0), P(1), P(2), \dots, P(n) \rightarrow f(x)。$$

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)。



→ 幾何(Geometric)模型

通常人口以幾何級數(複利)成長，即

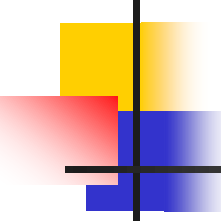
$$P(t) = P(0) \cdot (1 + r_a)^t,$$

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：

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

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

→ 實質成長率：

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

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

→ 瞬間成長率：

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

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

- 
- 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：

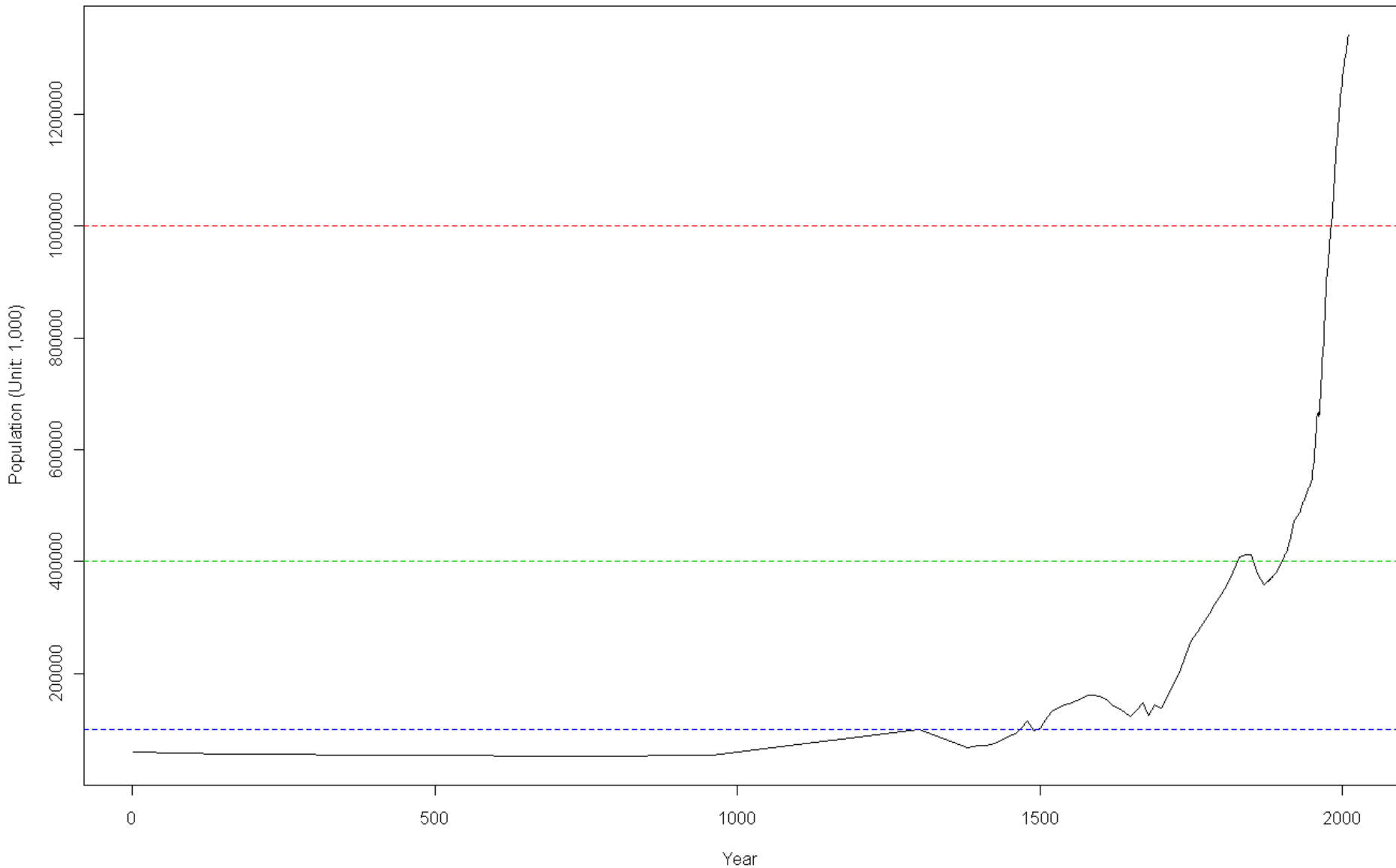
$$P \cdot e^{r_i t} = 2P \Rightarrow e^{r_i t} = 2$$

或可改寫為

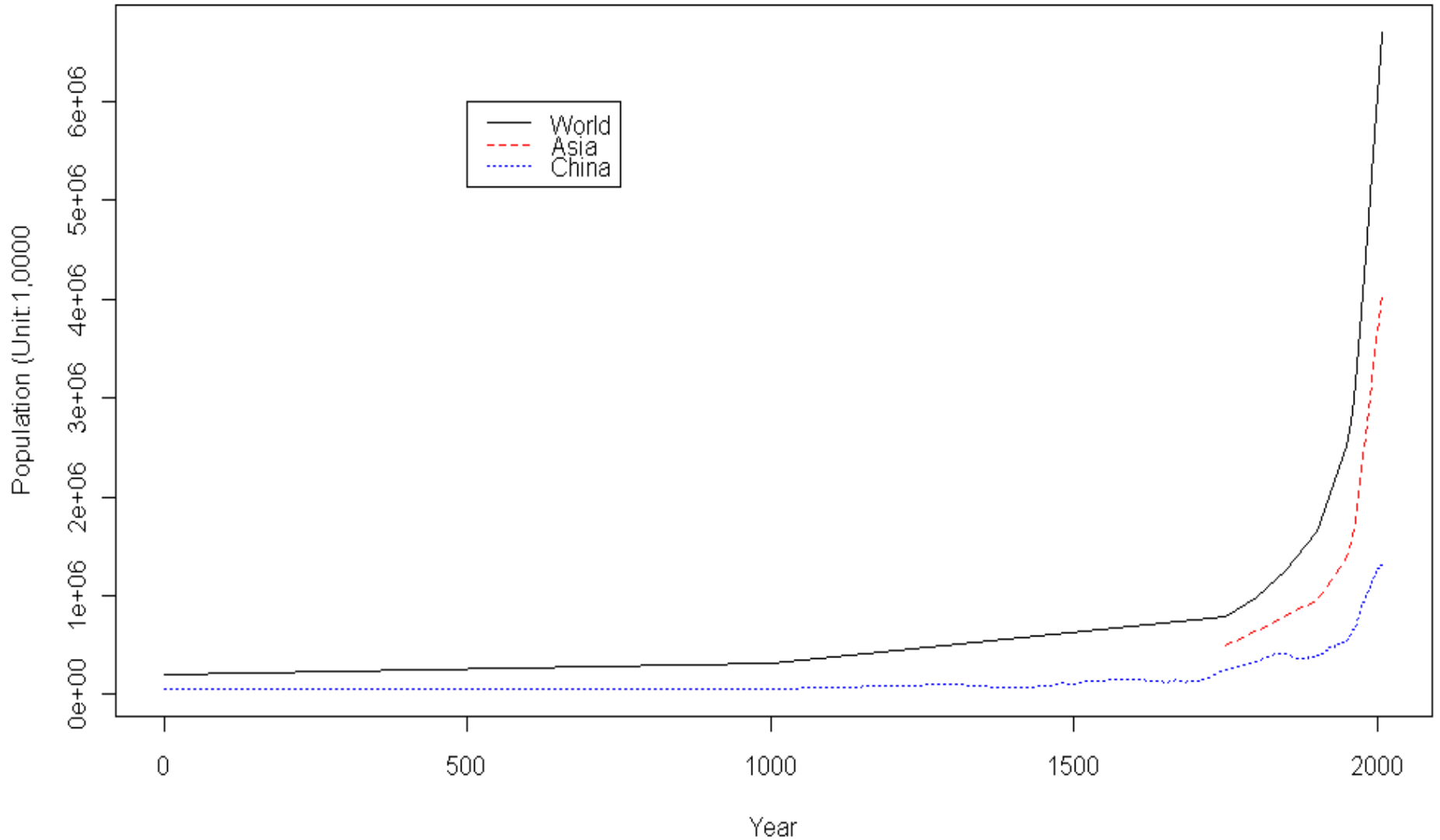
$$r_i t = \ln 2 = 0.693 \Leftrightarrow t = \frac{0.693}{r_i} \approx \frac{70}{100r_i}$$

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



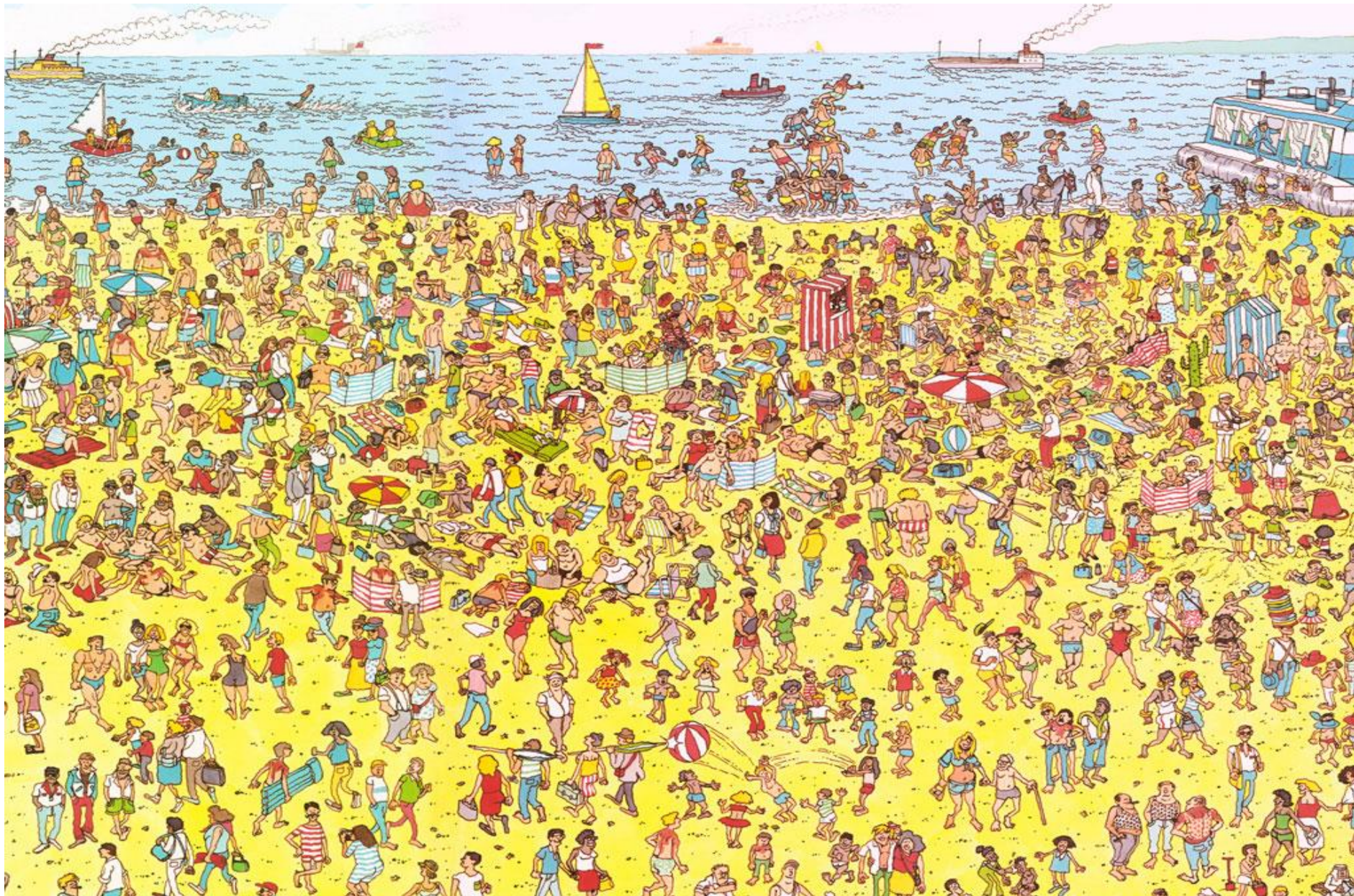
世界及亞洲兩千年來人口趨勢



人口爆炸！！



Where's Waldo ? ?

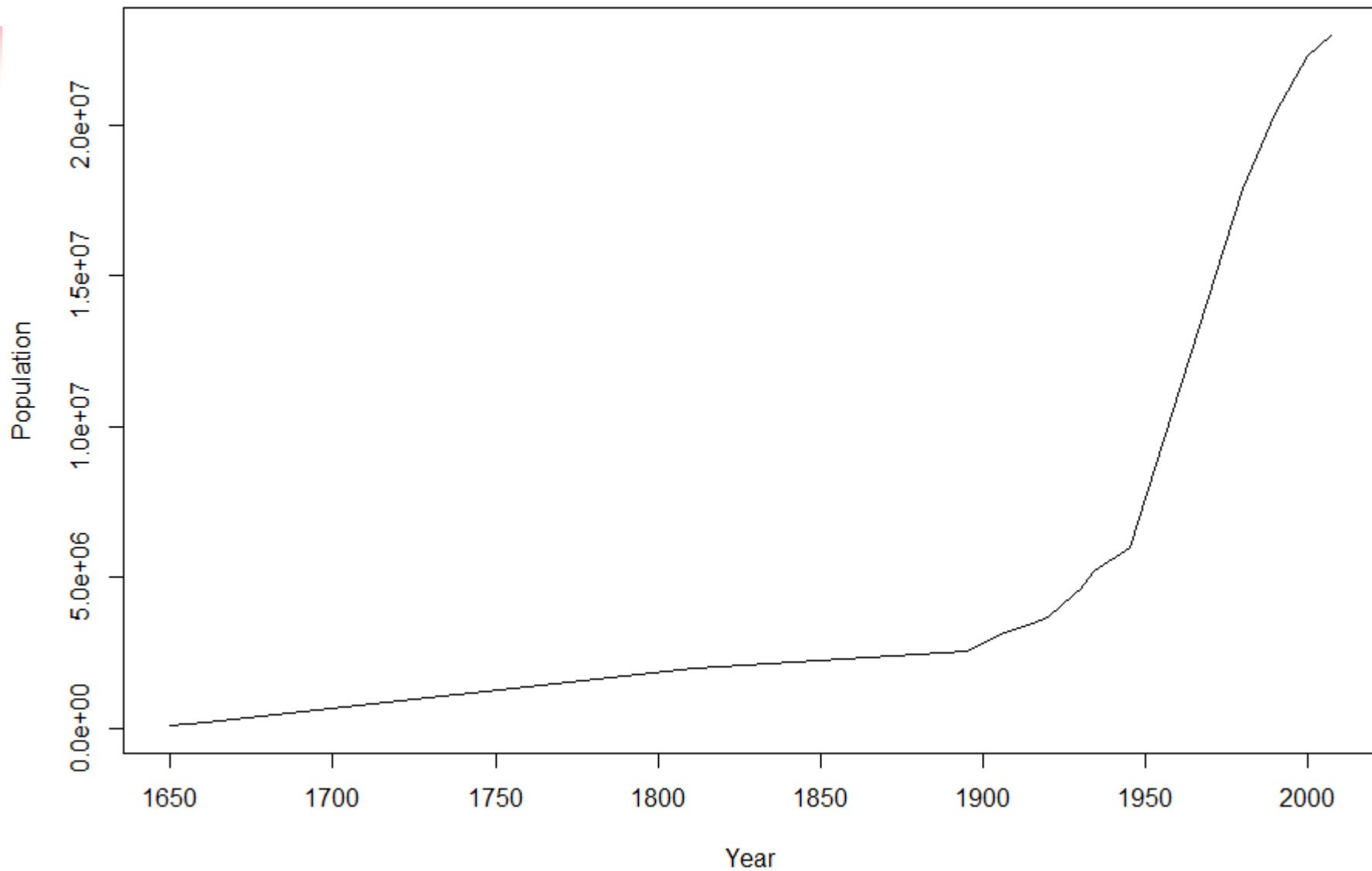


人口往都會地區集中。。

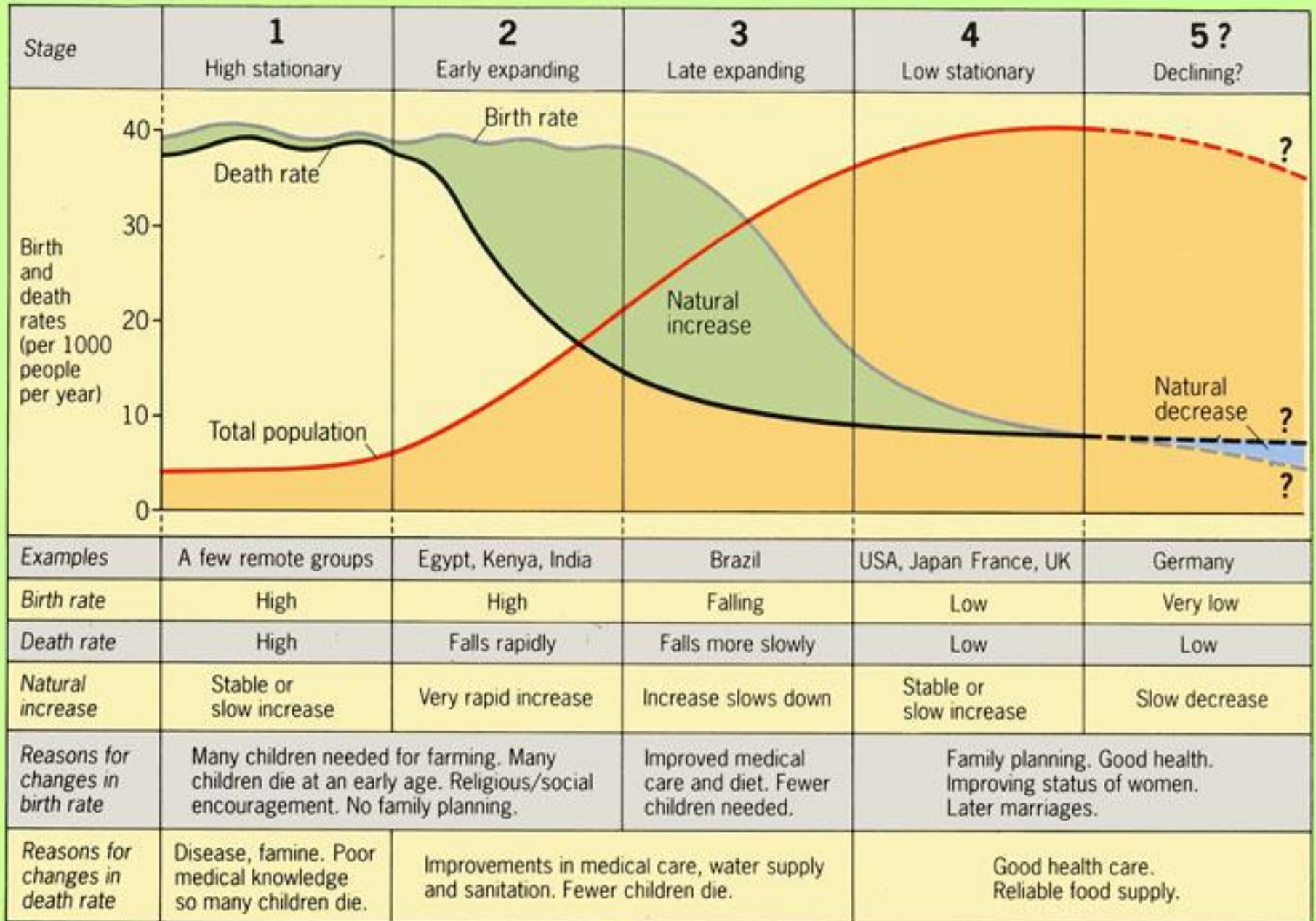




台灣自荷蘭統治時期的人口數



人口轉型理論(續)





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 $\lim_{t \rightarrow \infty} P(t) = a$ ，又因為瞬間成長率滿足

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

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：

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

若改寫為

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

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

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

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

Life Expectancy of Taiwan Male and Female



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：


$$78 / [1 + \text{EXP} (0.414457 - 0.036731t)]$$

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

→ 臺灣地區女性零歲平均餘命：

$$82 / [1 + \text{EXP} (0.664580 - 0.036843t)]$$

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


- 
- 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

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

代入已知條件，

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

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)

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

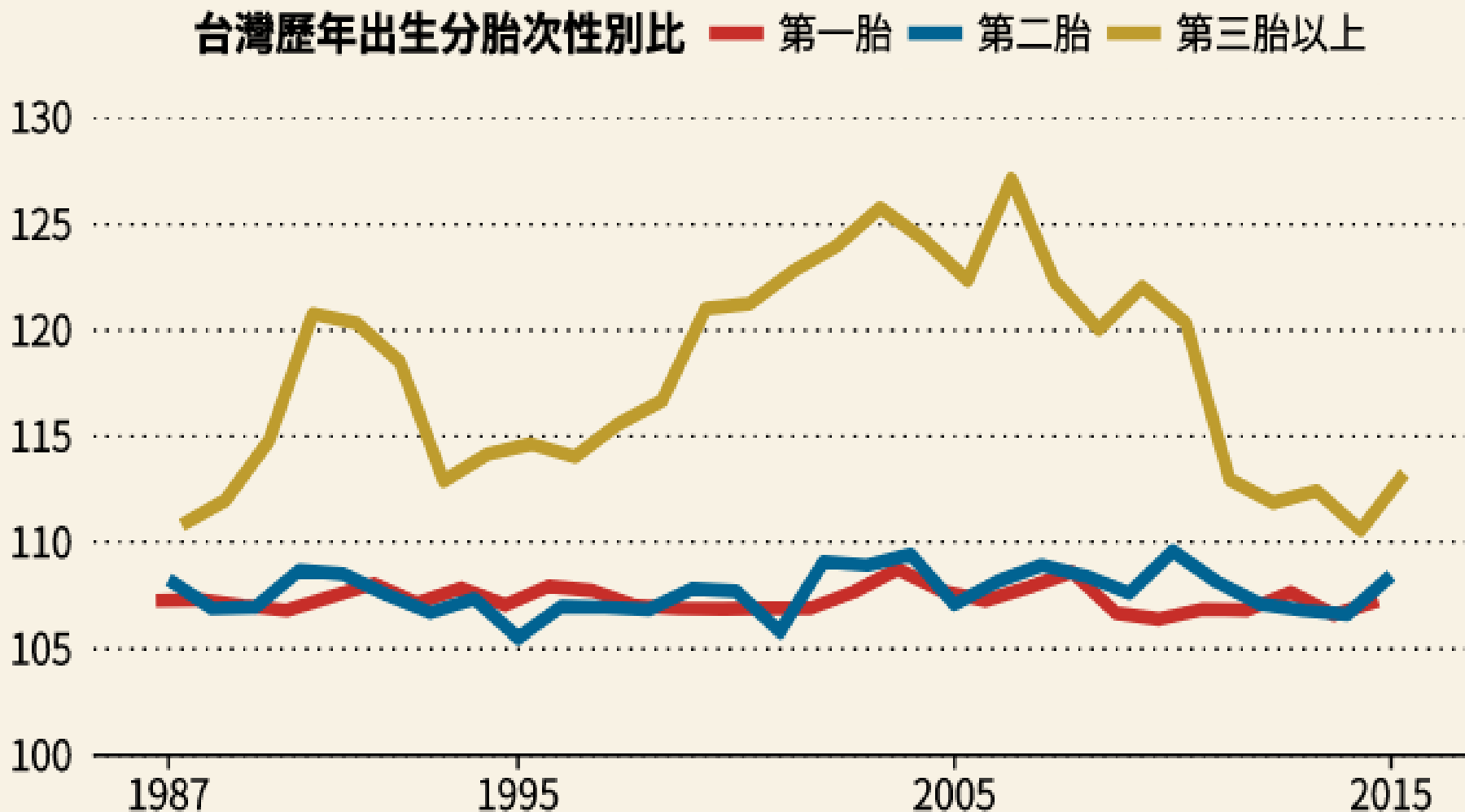
- 
- 人口學家利用數學的方法，將人口推估所需的成份（Component）以一矩陣來表示，稱為Leslie矩陣（Leslie Matrix），記做 M 。
 - 第一列中的元素表示單一年齡育齡婦女的生育率，只在15至49歲有值，其餘為零。而對角線以下的元素則為男、女生單一年齡的存活率，表示今年 t 歲的人，乘上存活率可得明年 $t+1$ 歲的人口數，其估計可寫成：

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

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

- 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）



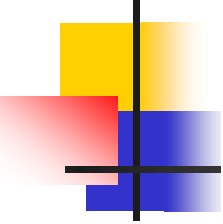
http://talkecon.com/content/images/2016/06/sex_ratio-2.svg

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

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

$$\rightarrow \lambda_1^k v_i \text{ as } k \rightarrow \infty,$$

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

→ 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。

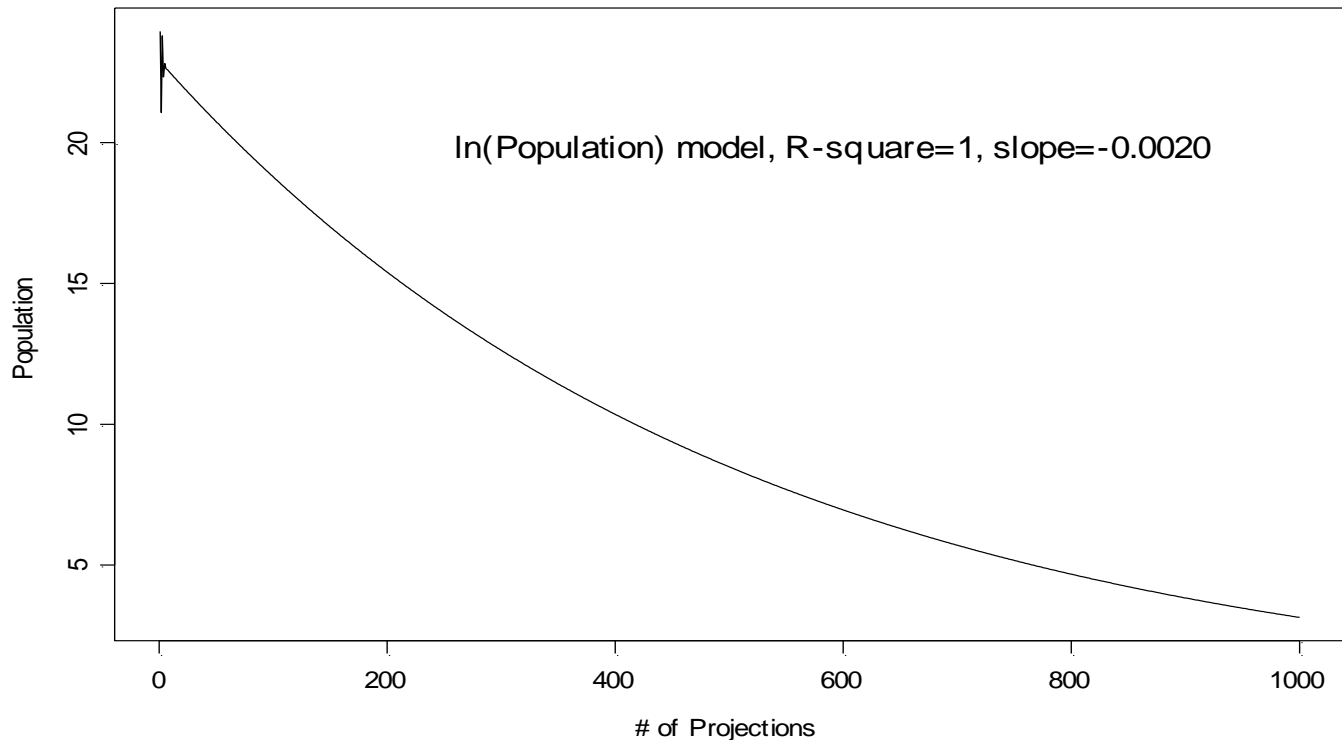
$$M = \begin{bmatrix} .3360 & .5152 & .1506 \\ .9946 & 0 & 0 \\ 0 & .9880 & 0 \end{bmatrix}$$

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

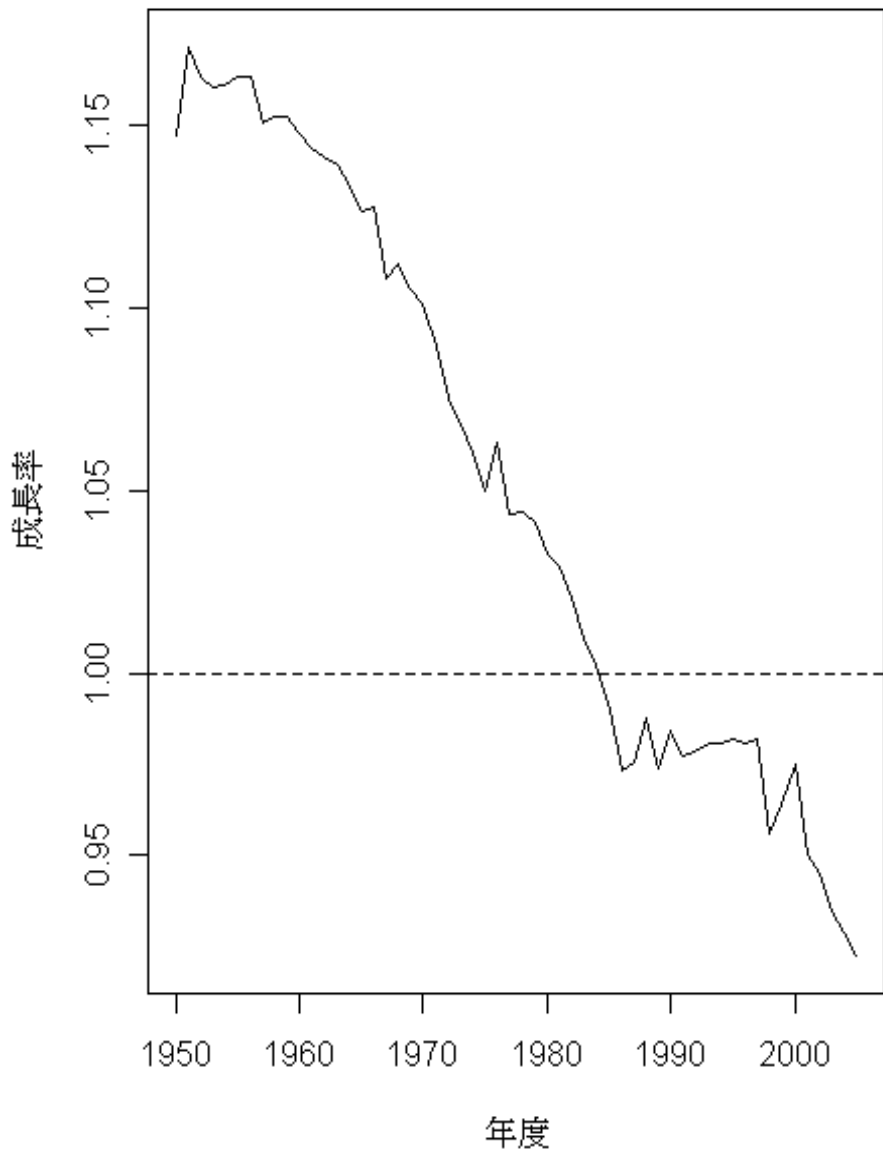
■ 範例五(續)

→ 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。

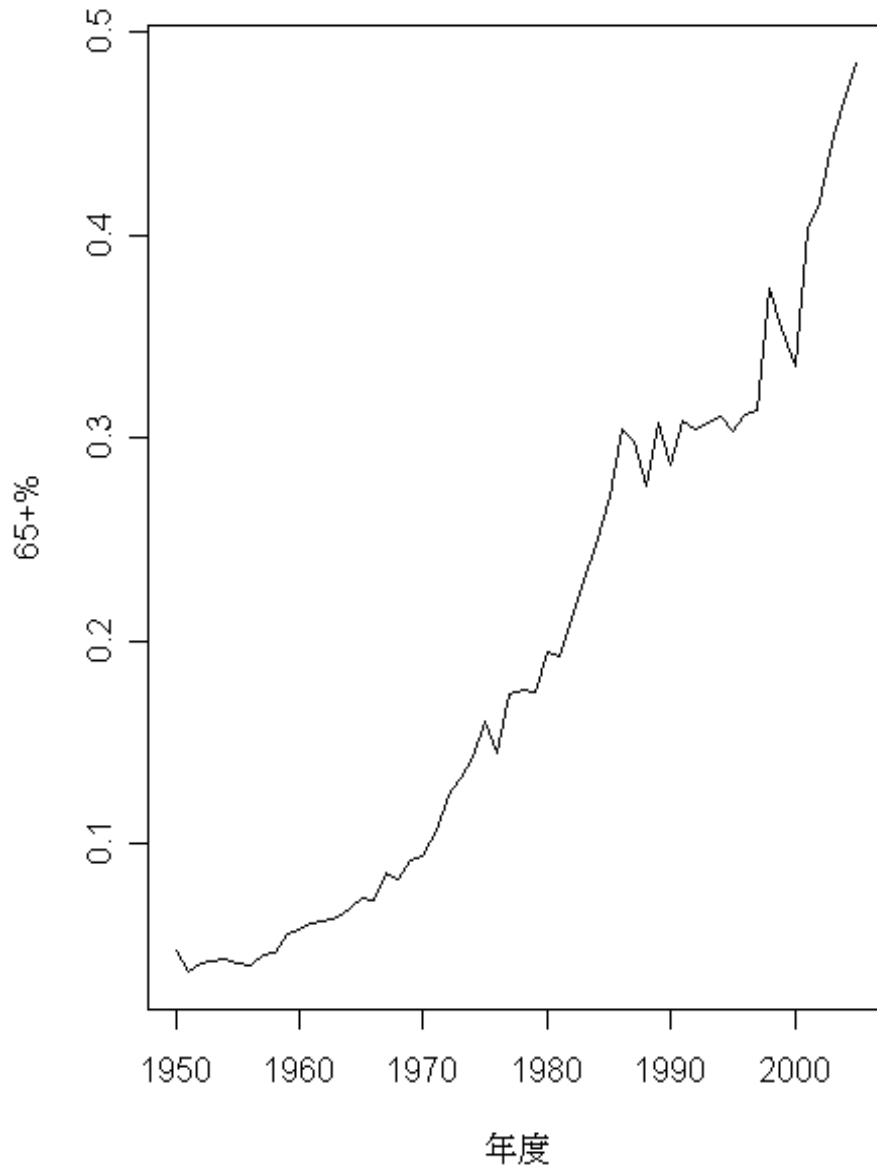
Projected Population



特徵值



65+的人口比例



臺灣地區自1950年以來的人口成長（特徵值）

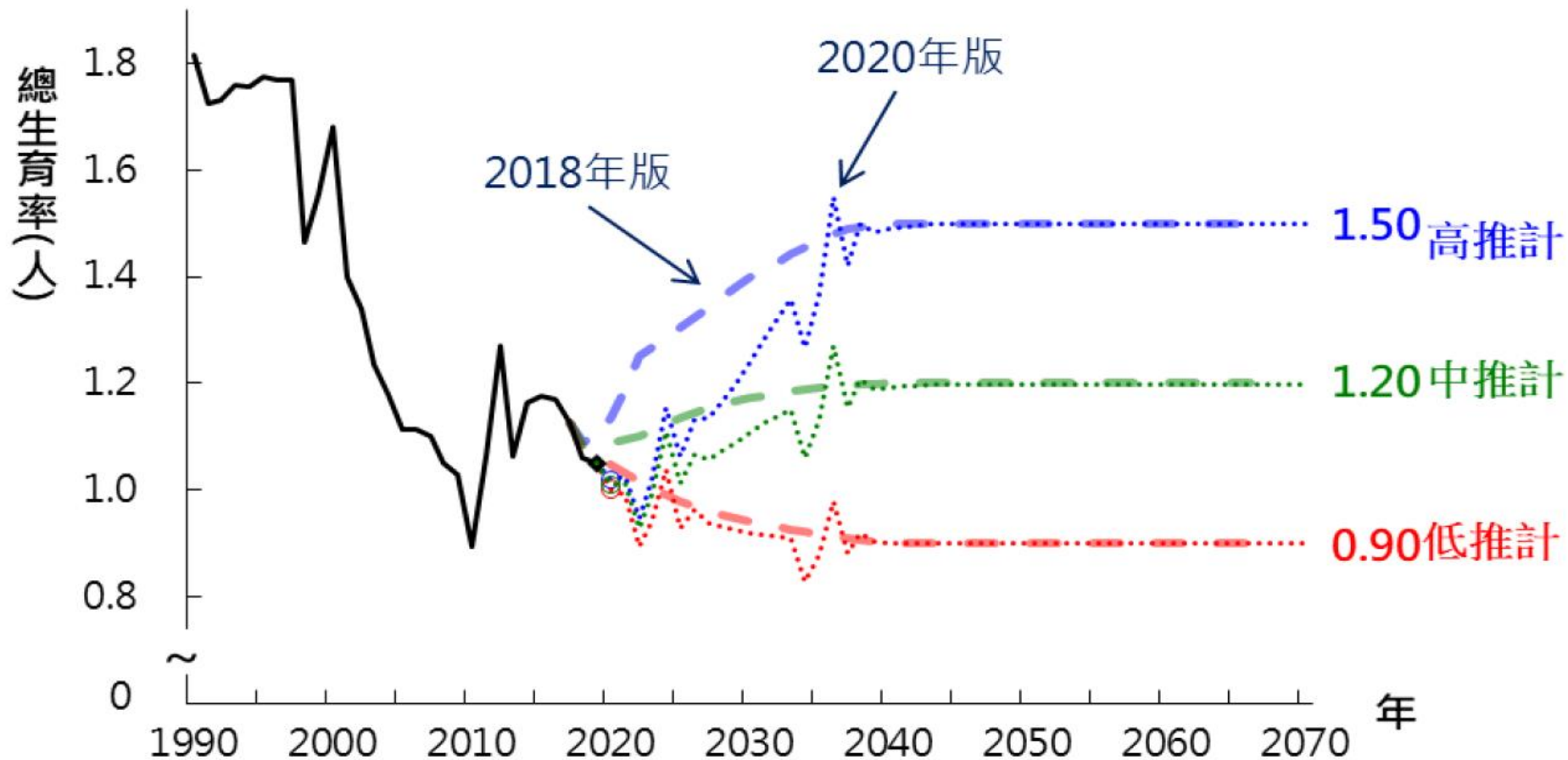


生育、死亡、遷移的假設

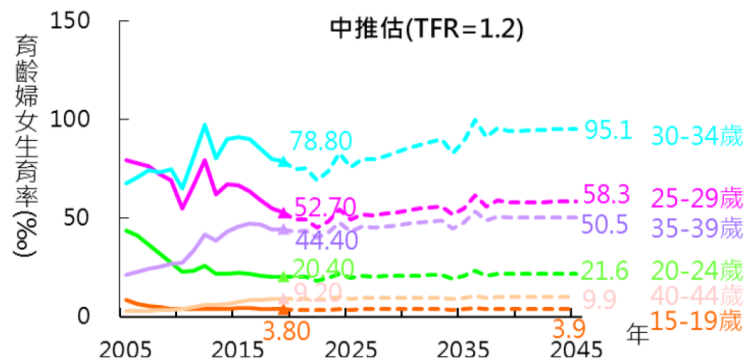
- 因為生育率與死亡率是Leslie矩陣的最重要參數，不同的假設數值將衍生出截然不同的推估結果。
 - 經歷經濟高速成長、進入已開發國家之林後，臺灣生育率及死亡率逐年下降。然而，生育率自2010年後回升至1.2左右，未來趨勢仍是未定之數。
 - 然而，近年臺灣遷出遷入較為頻繁，2008年行政院經建會已將遷移列入考量。

2020~2070年臺灣人口推估(生育率假設)

資料來源：行政院國家發展委員會(2020)



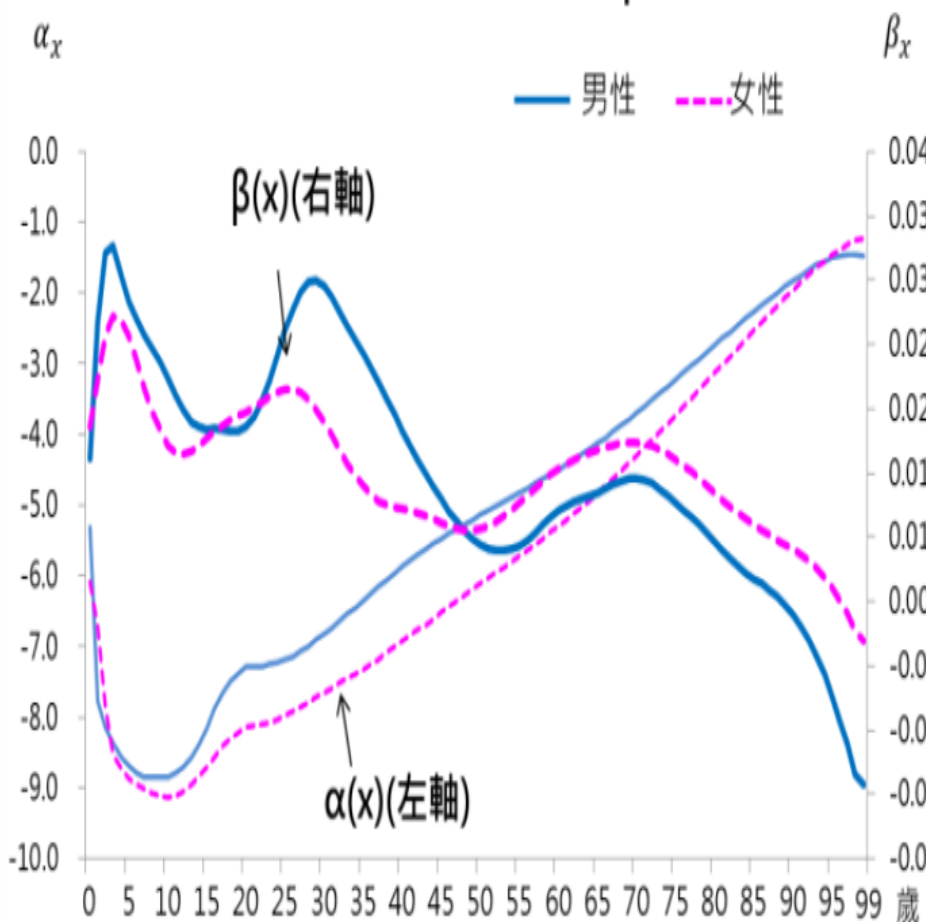
| 推估版本 | 基期年實績 | 高推估 | 中推估 | 低推估 | 說明 |
|-----------------|-------|------|------|-------|---|
| 2014 年版 | 1.065 | 1.40 | 1.10 | 0.80 | 至 2036 年達成假設水準後維持 |
| 2016 年版 | 1.175 | 1.50 | 1.20 | 0.90 | 至 2041 年達成假設水準後維持 (另高推估先於 2030 年達成 1.4 之目標) |
| 2018 年版 | 1.125 | 1.50 | 1.20 | 0.90 | 至 2040 年達成假設水準後維持 (另高推估分別先於 2022 年達 1.25、2030 年達成 1.4 之目標) |
| 2020 年版 (暫訂) | 1.050 | 1.50 | 1.20 | 0.90- | 至 2045 年達成假設水準後維持 |



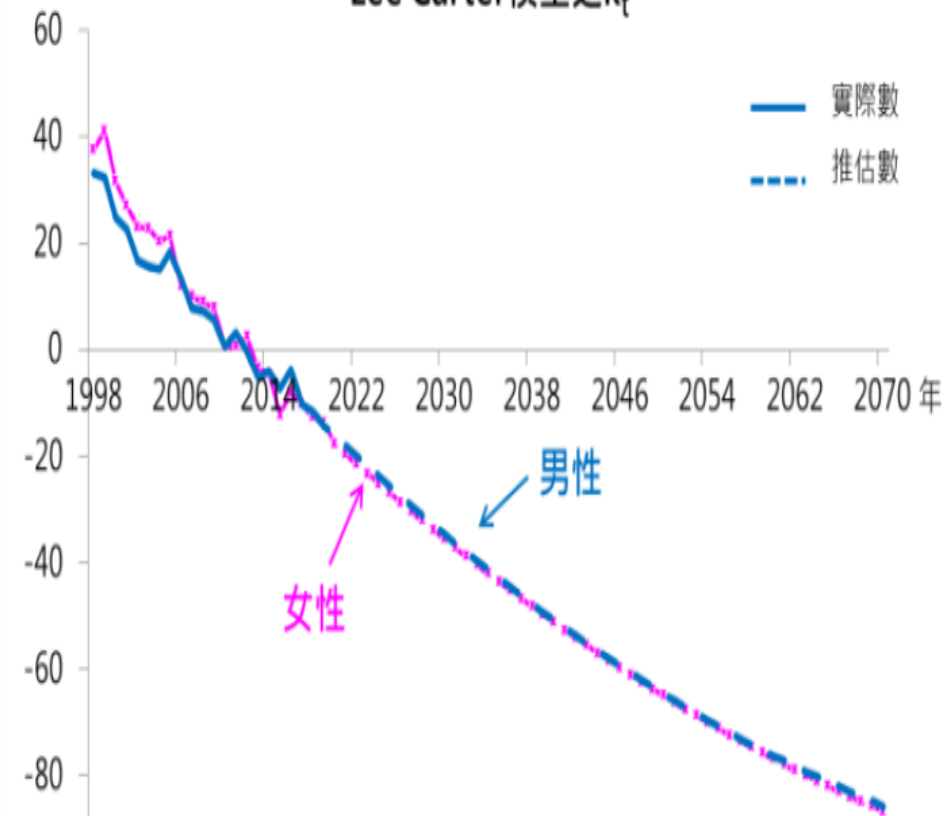
死亡假設：15資料代入Lee-Carter模型

資料來源：行政院國家發展委員會(2020)

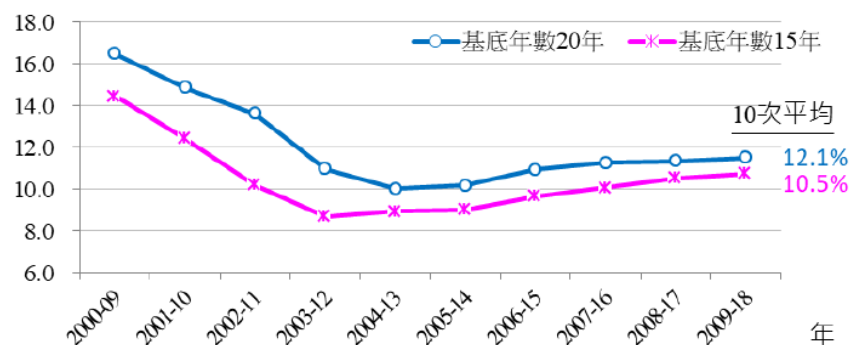
Lee Carter模型之 α 、 β



Lee Carter模型之 k_t



MAPE(%) 男性不同基底年數之MAPE

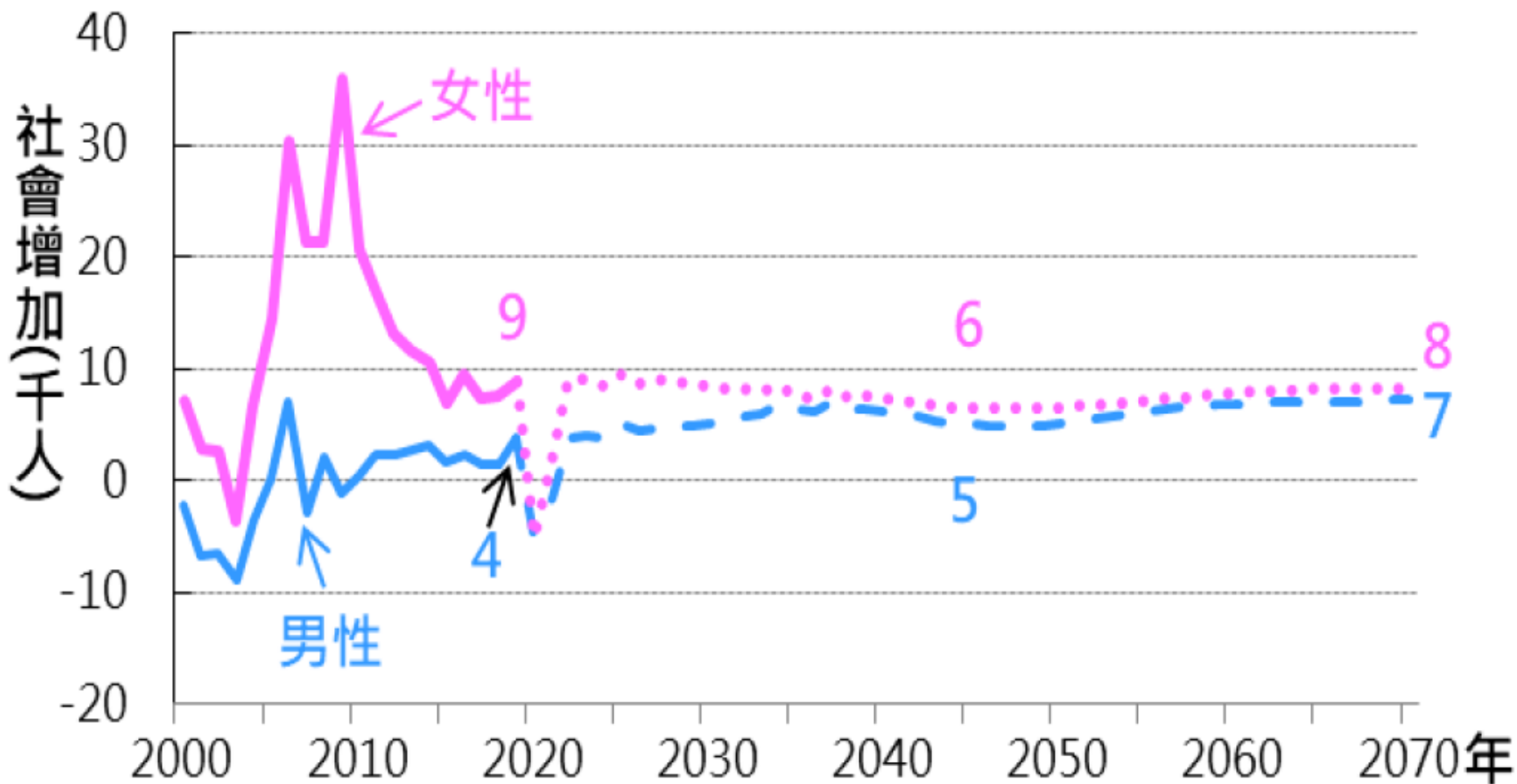


■ 推估相關因素：「資料品質」、「地區人數」、「資料年數」與「推估年數」。

淨遷徙假設（本國人、外國人）

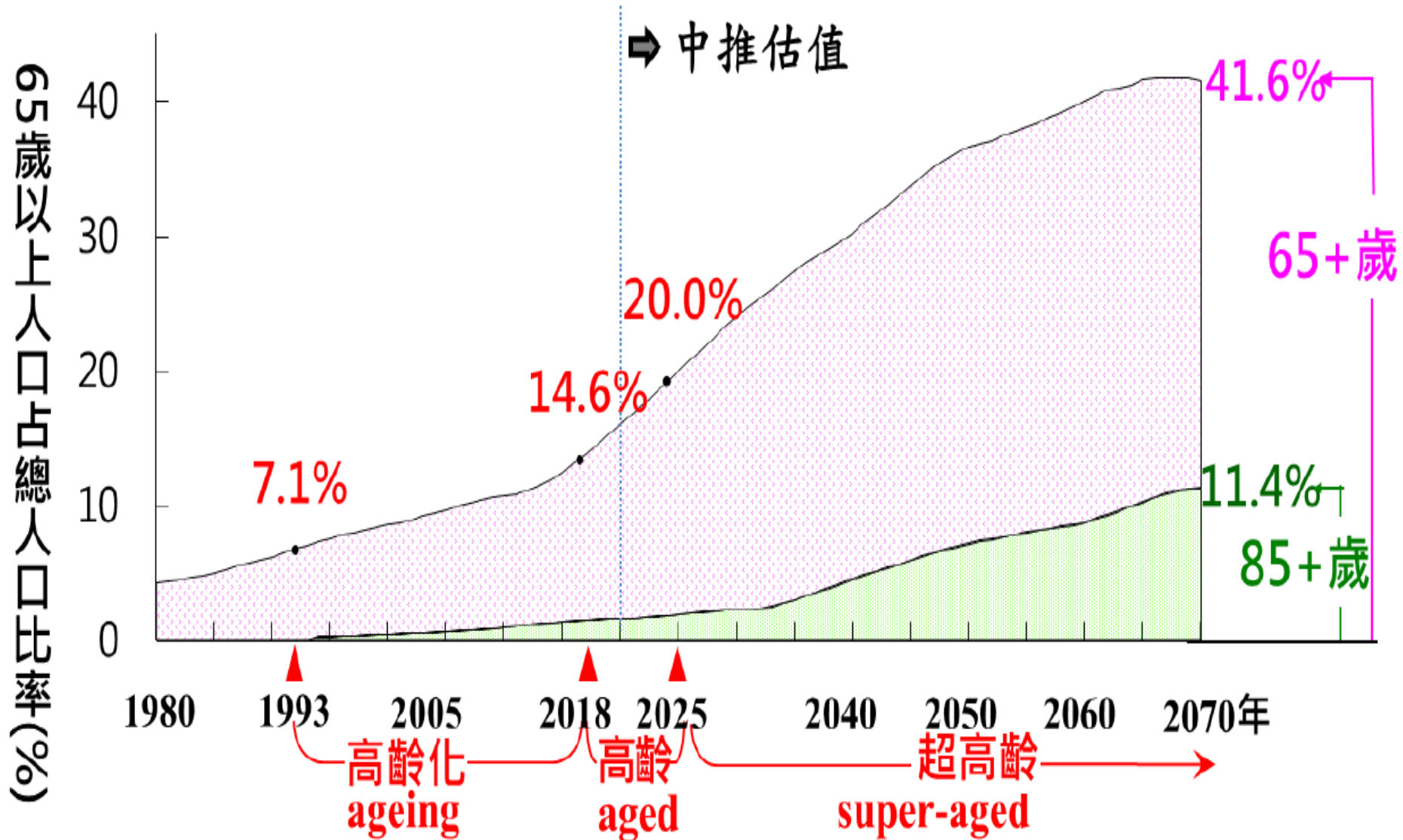
- 本國人：年齡、性別的國際淨遷徙（2021年加入武漢肺炎）
- 外國人：各年齡2010~2019年中位數，以3歲移動平均修勻，推估未來6年、之後維持固定。

資料來源：行政院國家發展委員會(2020)



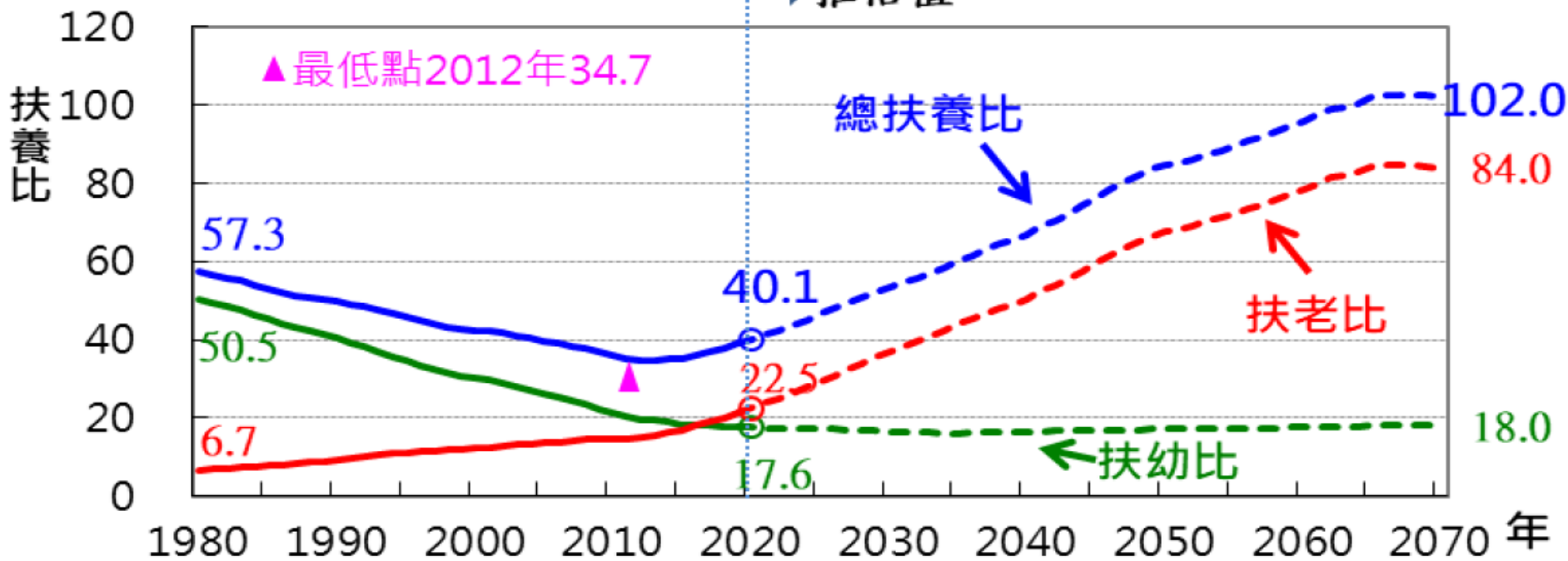
未來人口老化趨勢更加明顯 (中推計)

資料來源：行政院國家發展委員會(2020)

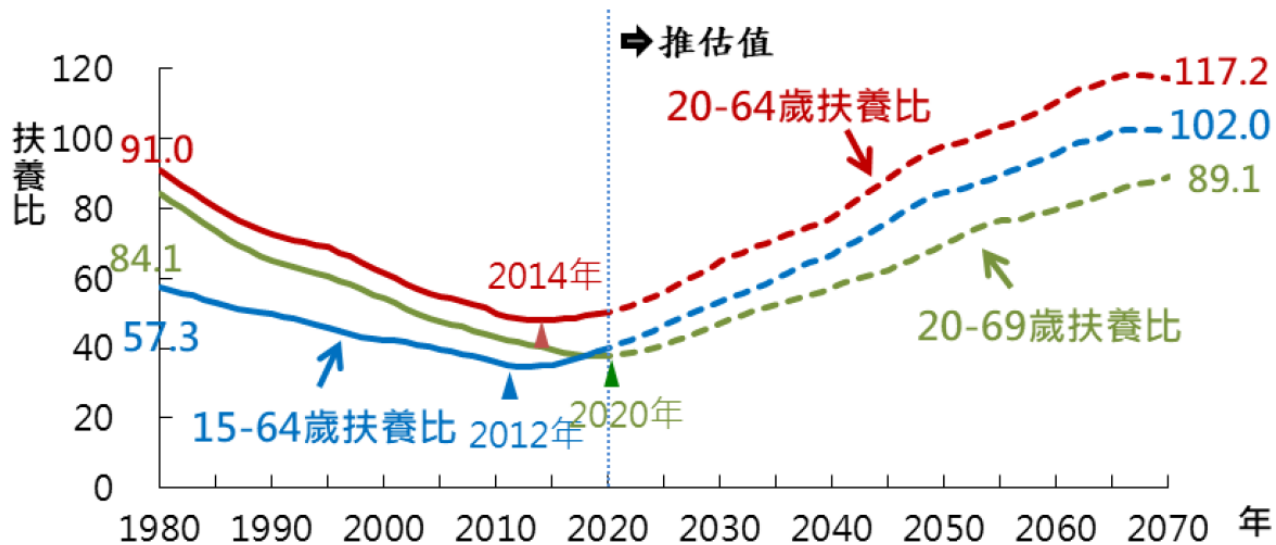


扶養比（扶老比）增加快速（中推計）

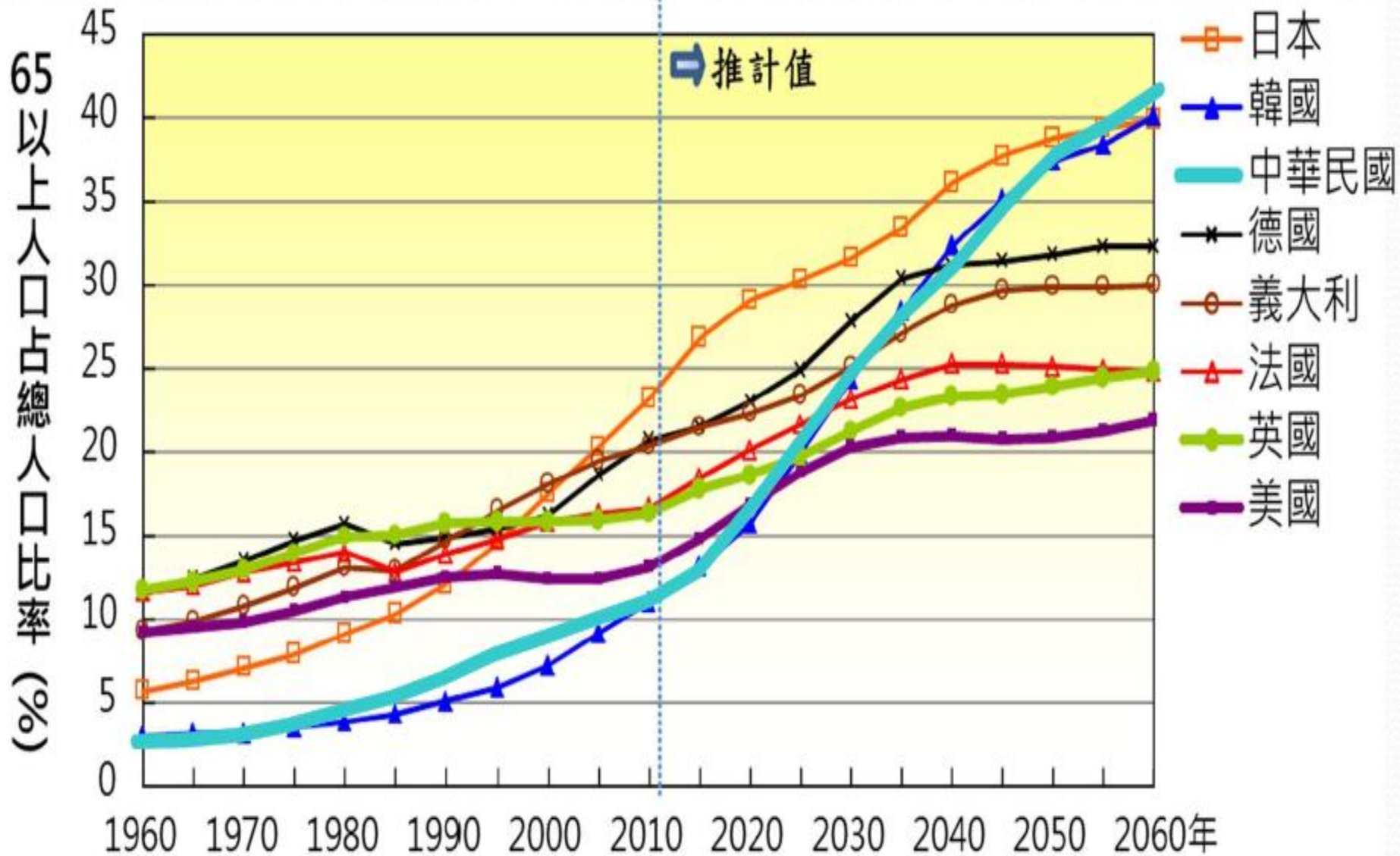
資料來源：行政院國家發展委員會(2020)



註：延後退休年齡
可稍微舒緩扶
養比的壓力。

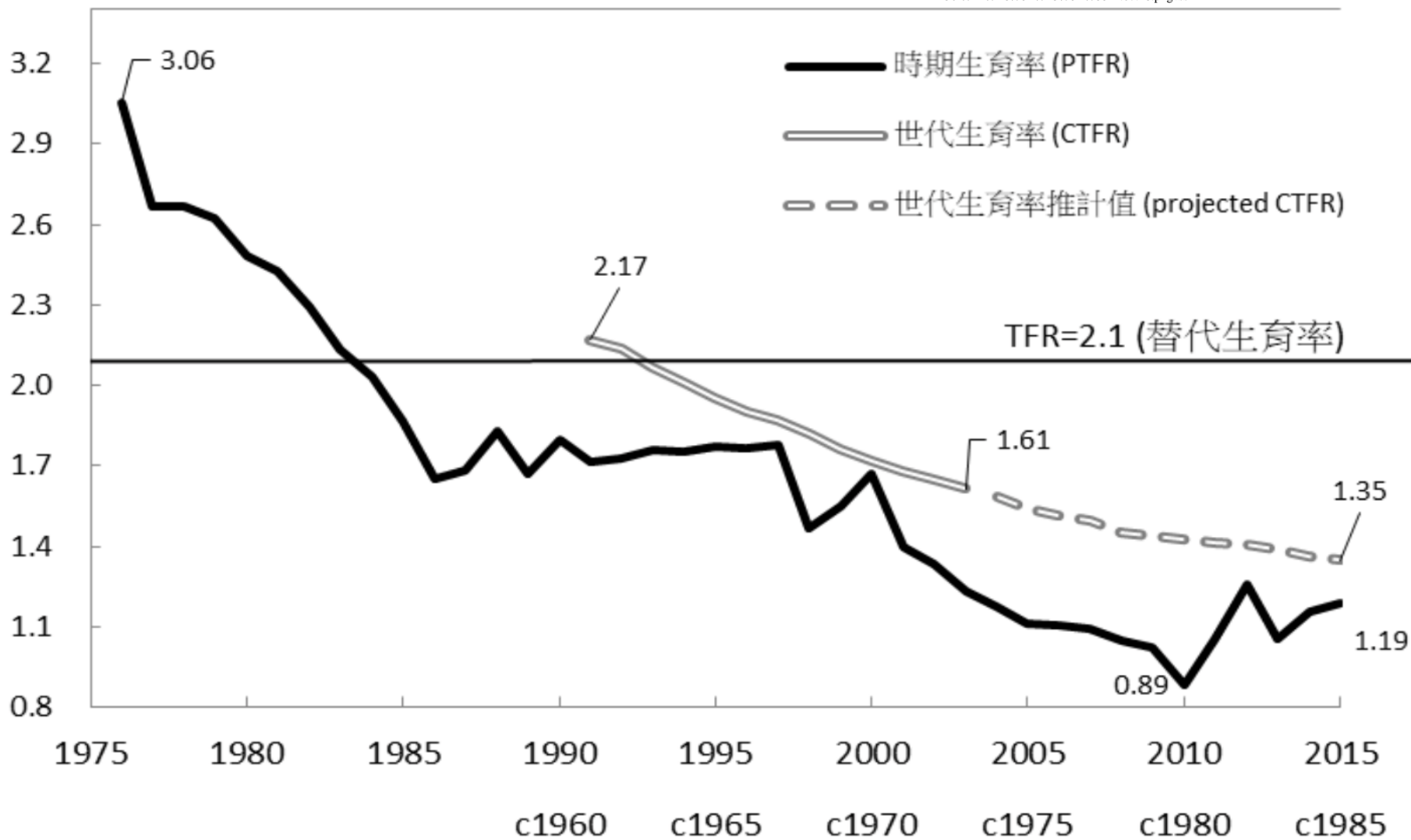


世界各國65歲以上人口占總人口比率



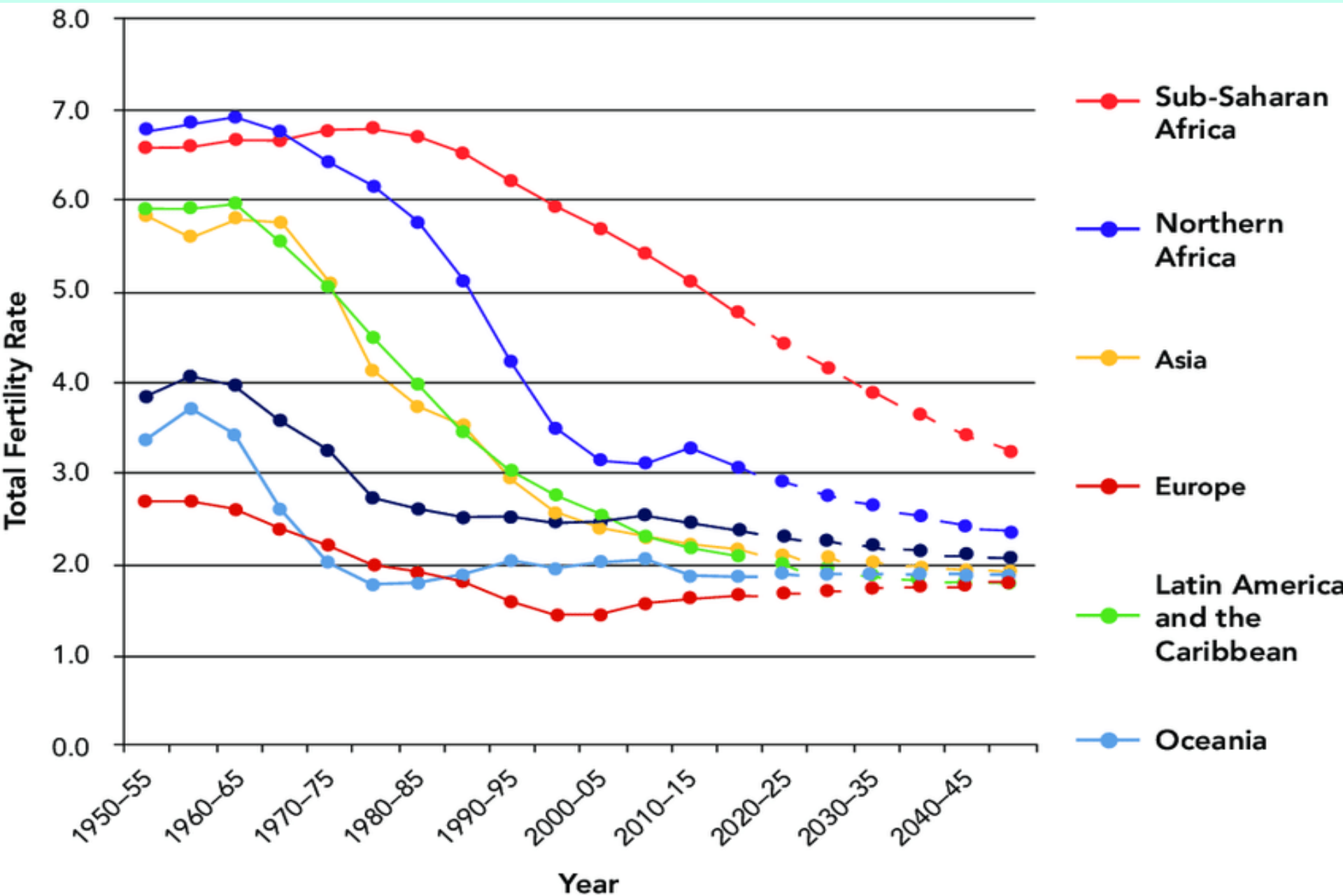
臺灣歷年總生育率

<https://i0.wp.com/twstreetcorner.org/wp-content/uploads/2019/03/%E8%9E%A2%E5%B9%95%E5%BF%AB%E7%85%A7-2019-03-30-%E4%B8%8B%E5%8D%8811.07.18.png?ssl=1>



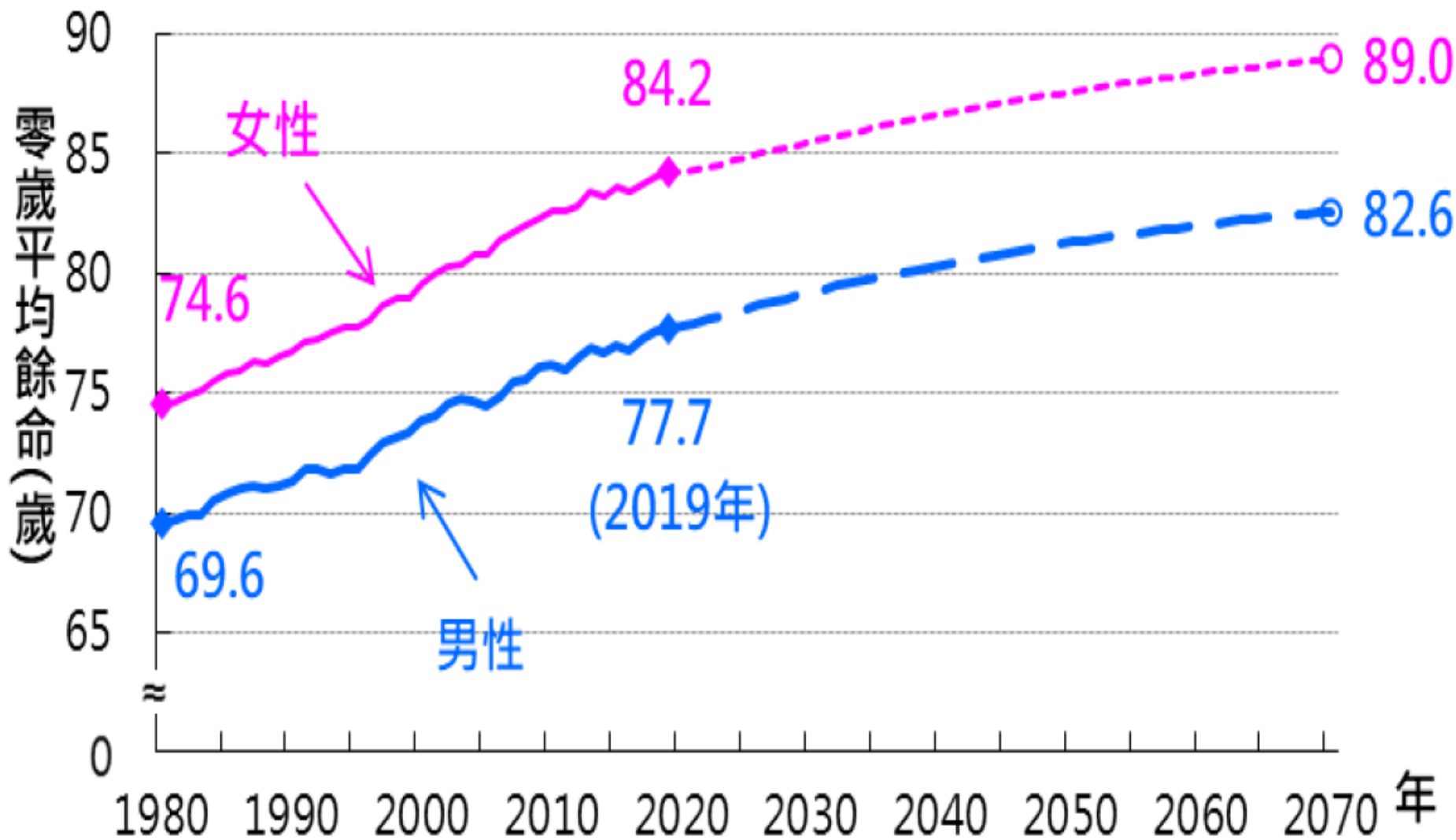
資料來源：Human Fertility Database。C 代表世代 (cohort)，c1960 指 1960 年出生之女性。CTFR 推計方法請見 Myrskylä, Goldstein, and Cheng (2013) "New Cohort Fertility Forecasts for the Developed World: Rises, Falls, and Reversals", Population and Development Review, 39(1): 31-56.

世界各國歷年及推估之總生育率



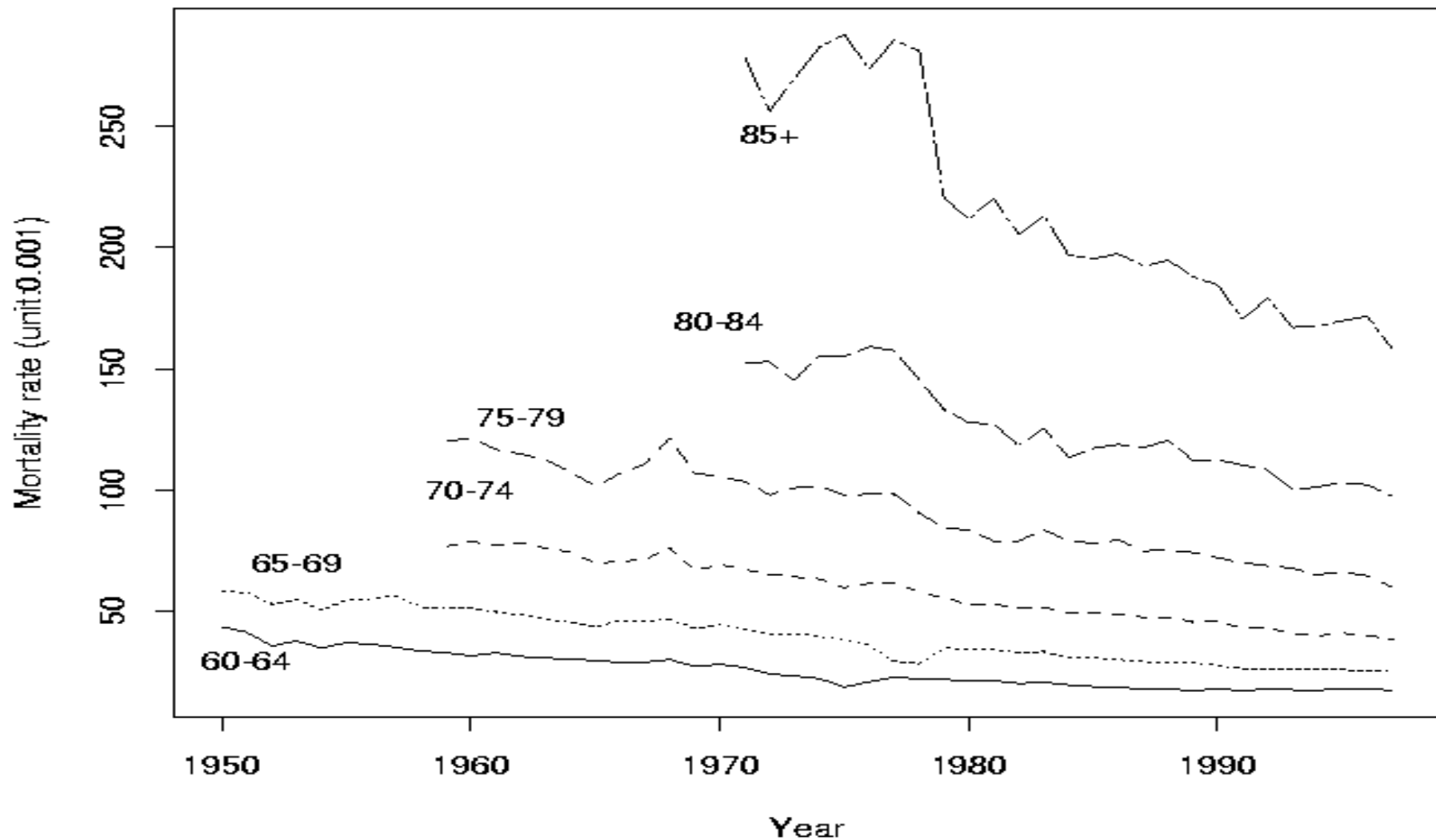
臺灣居民平均壽命的趨勢

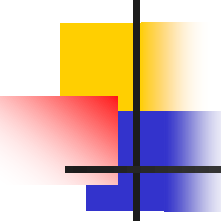
資料來源：行政院國家發展委員會(2020)



註：除了Lee-Cater模型，高齡死亡率也採Gompertz模型。

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。



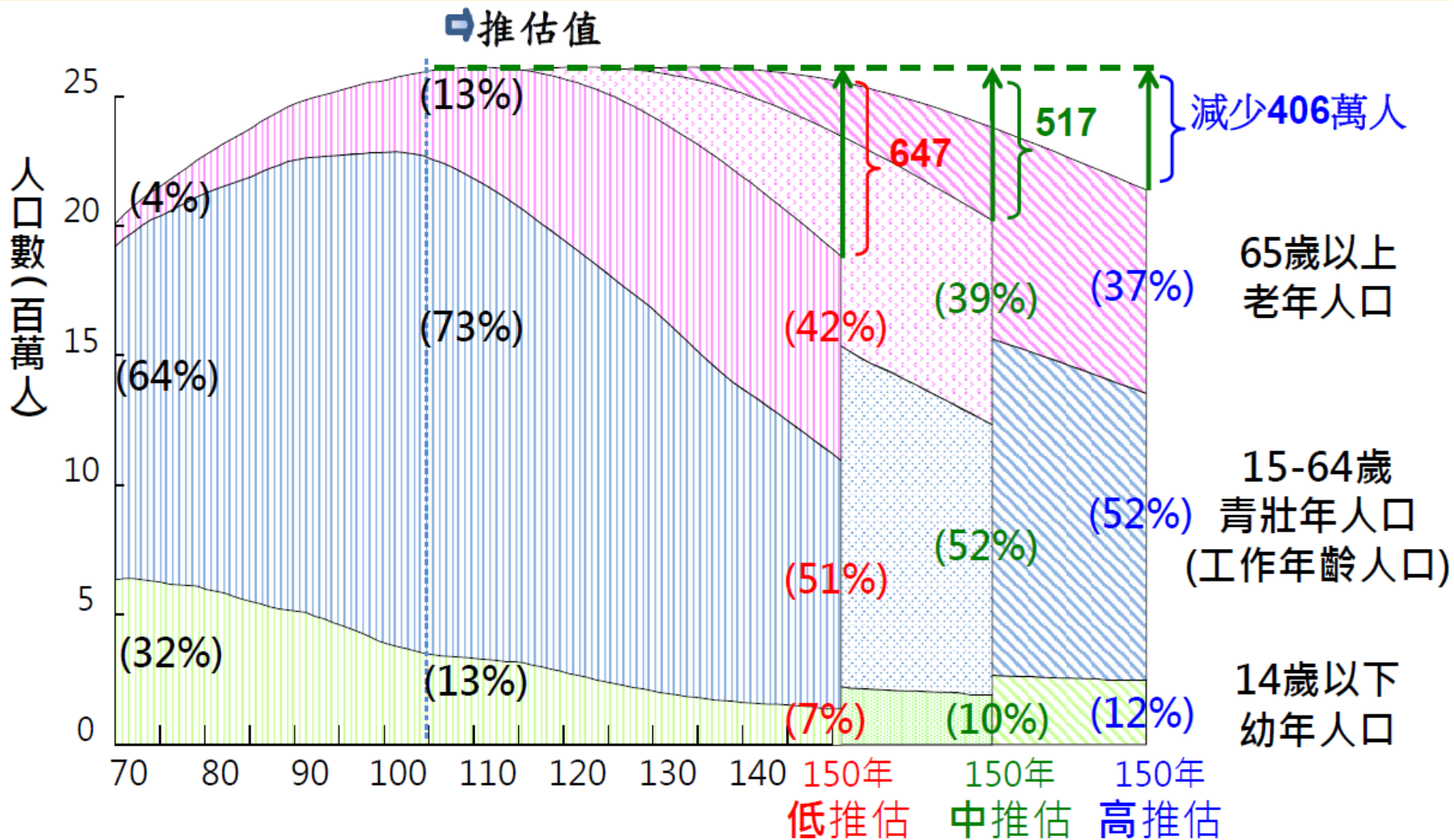
專家意見

- 聯合國及歐美各國的人口推估，仍以人口變動合成要素法為主，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)

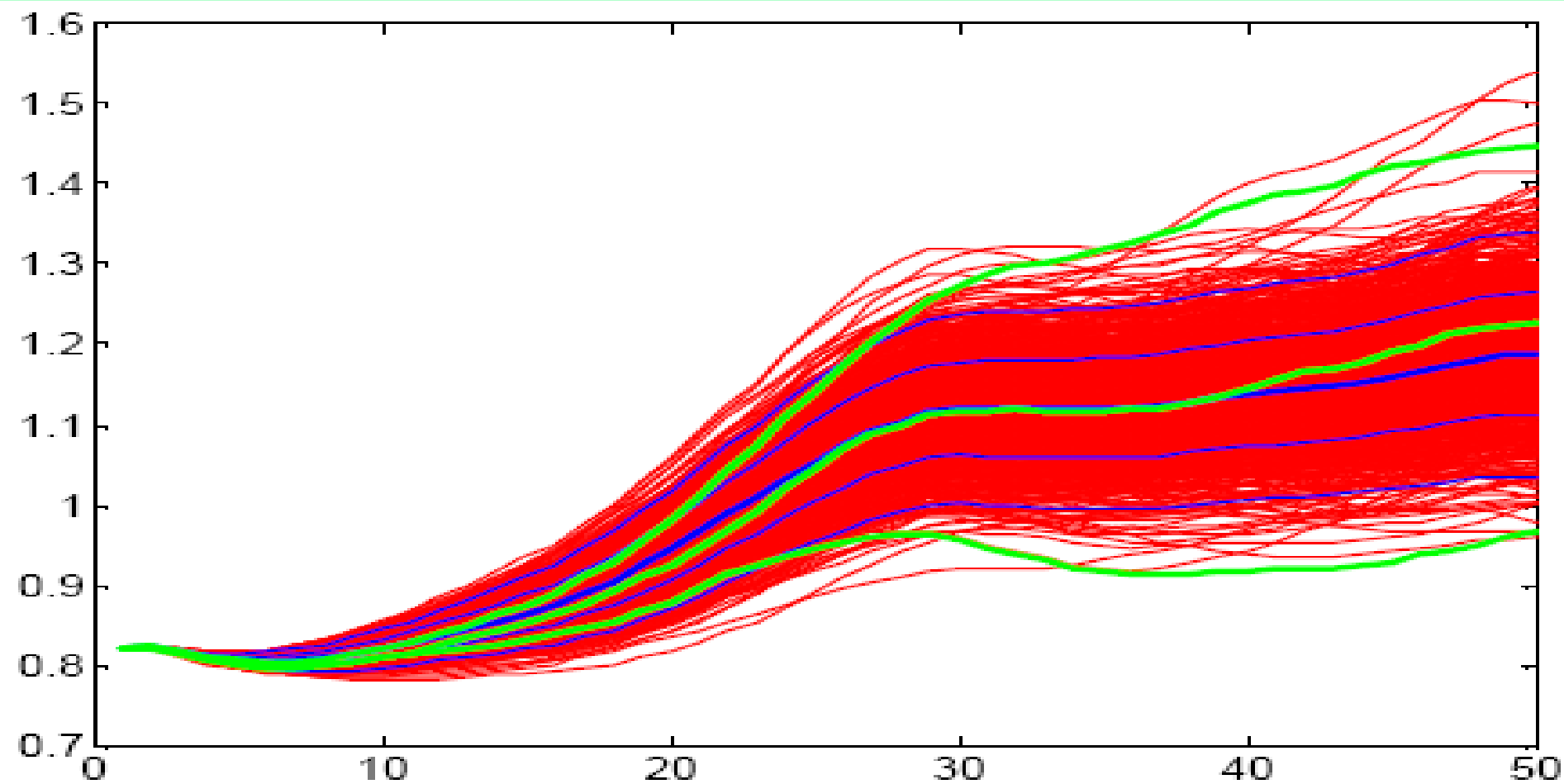
國發會 (前身為經建會) 民國105年公佈的人口推估



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

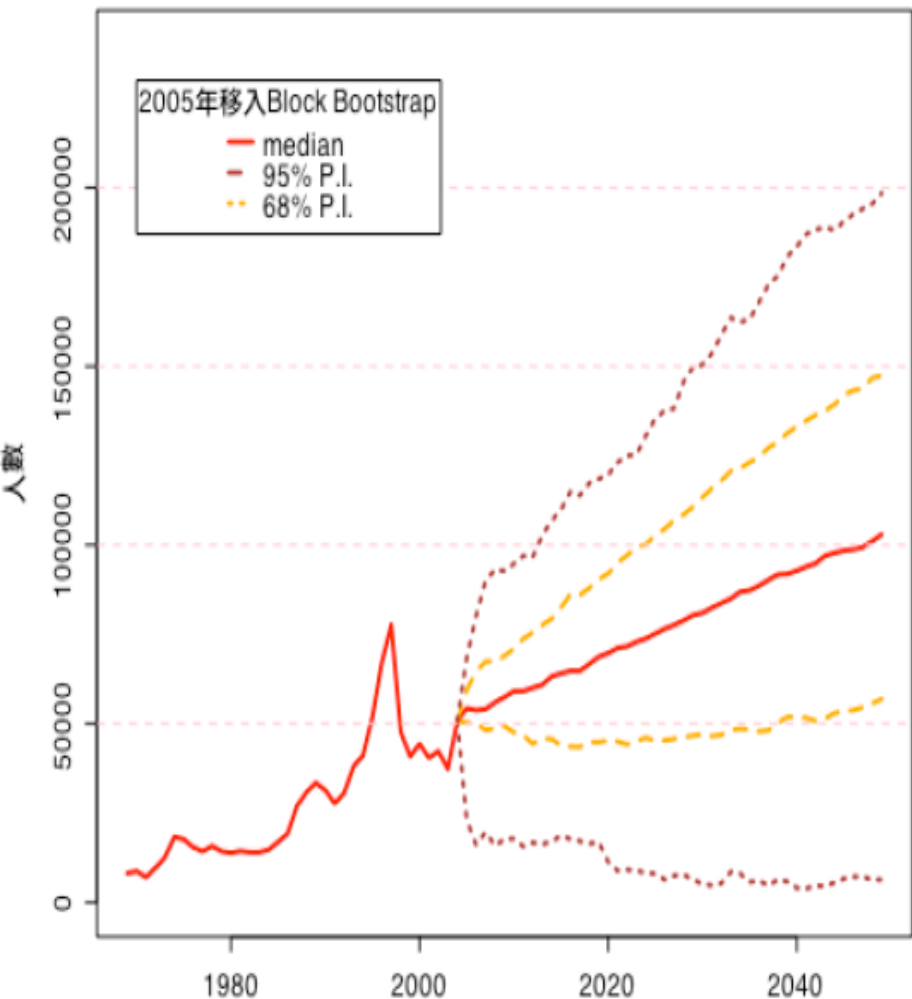
Blue: 平均值、68%預測區間及95%預測區間

資料來源：Bertz & Lipps (2004)

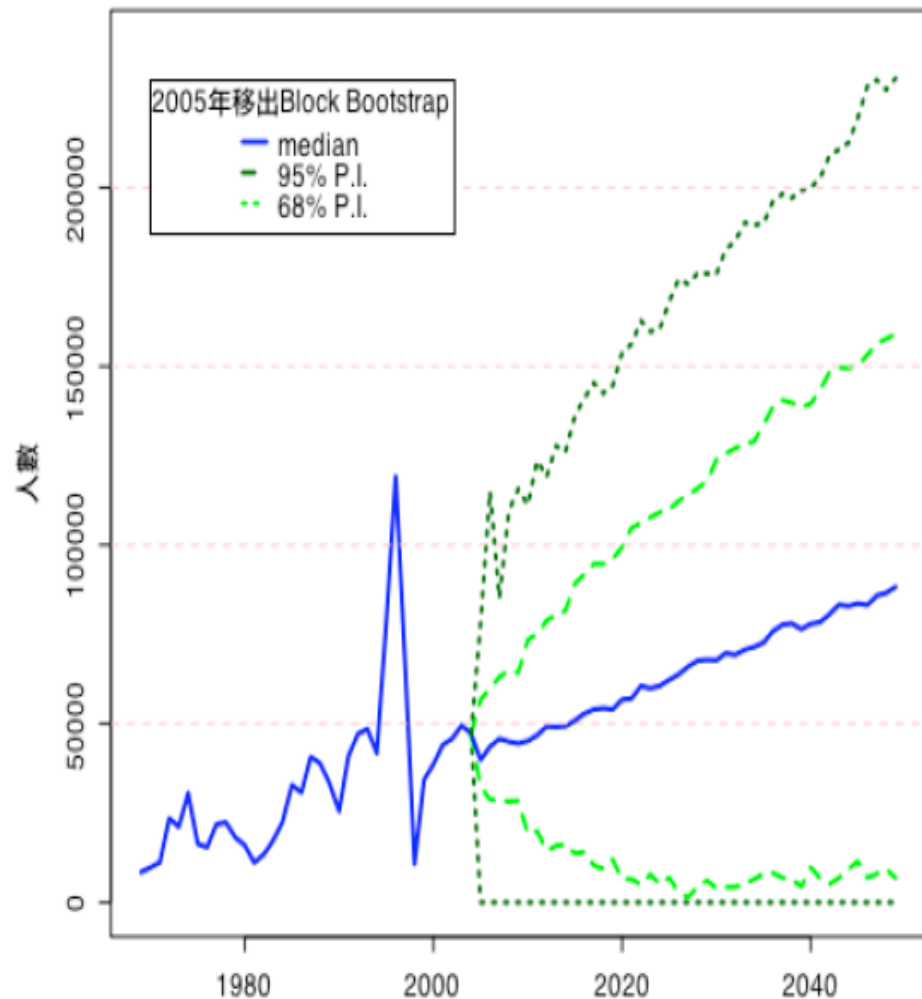


隨機方法推估之呈現：(區塊拔靴法)

遷入台灣人口



遷出台灣人口



■ 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。

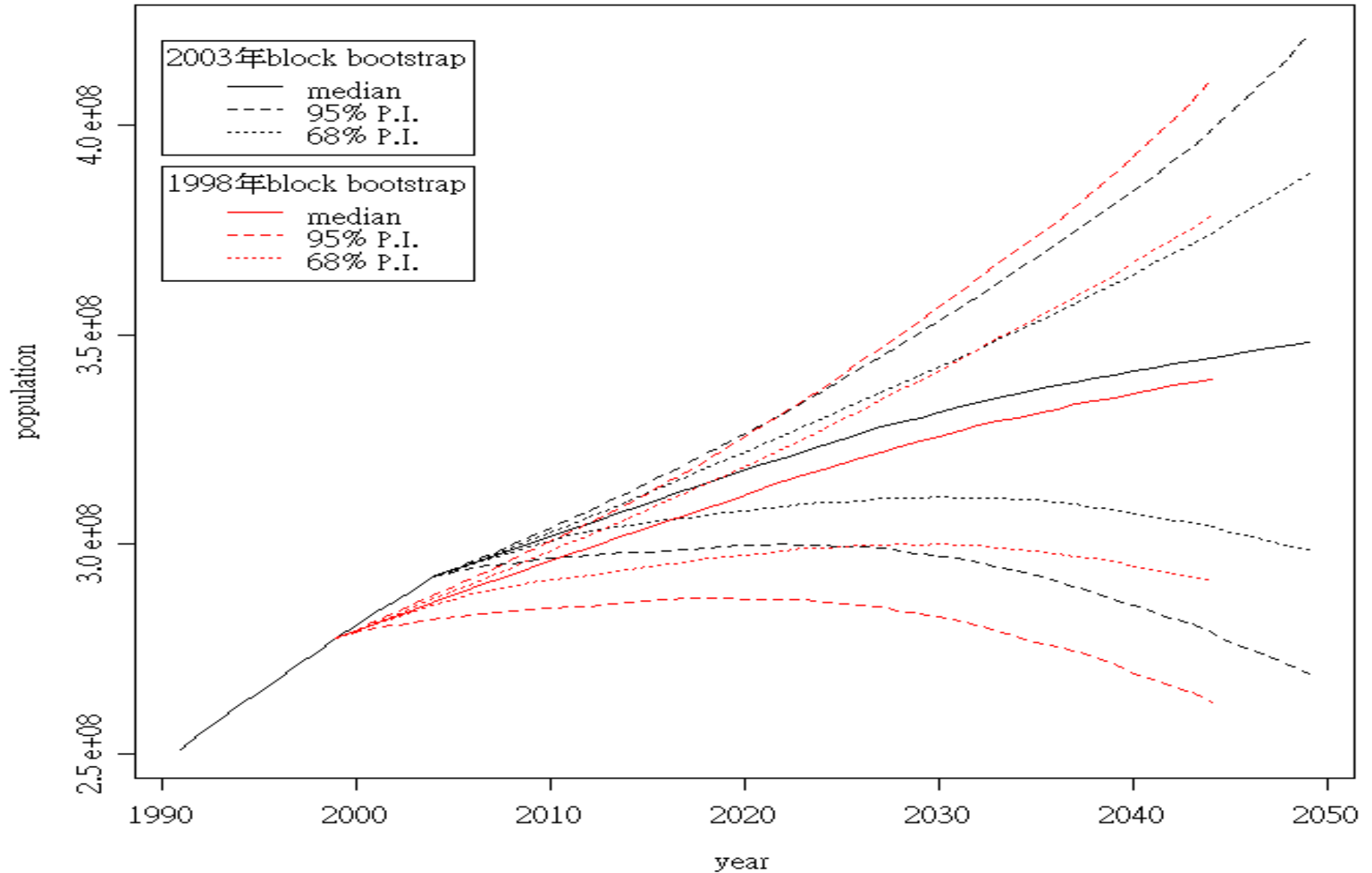
區塊(Block)拔靴法

- 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。)

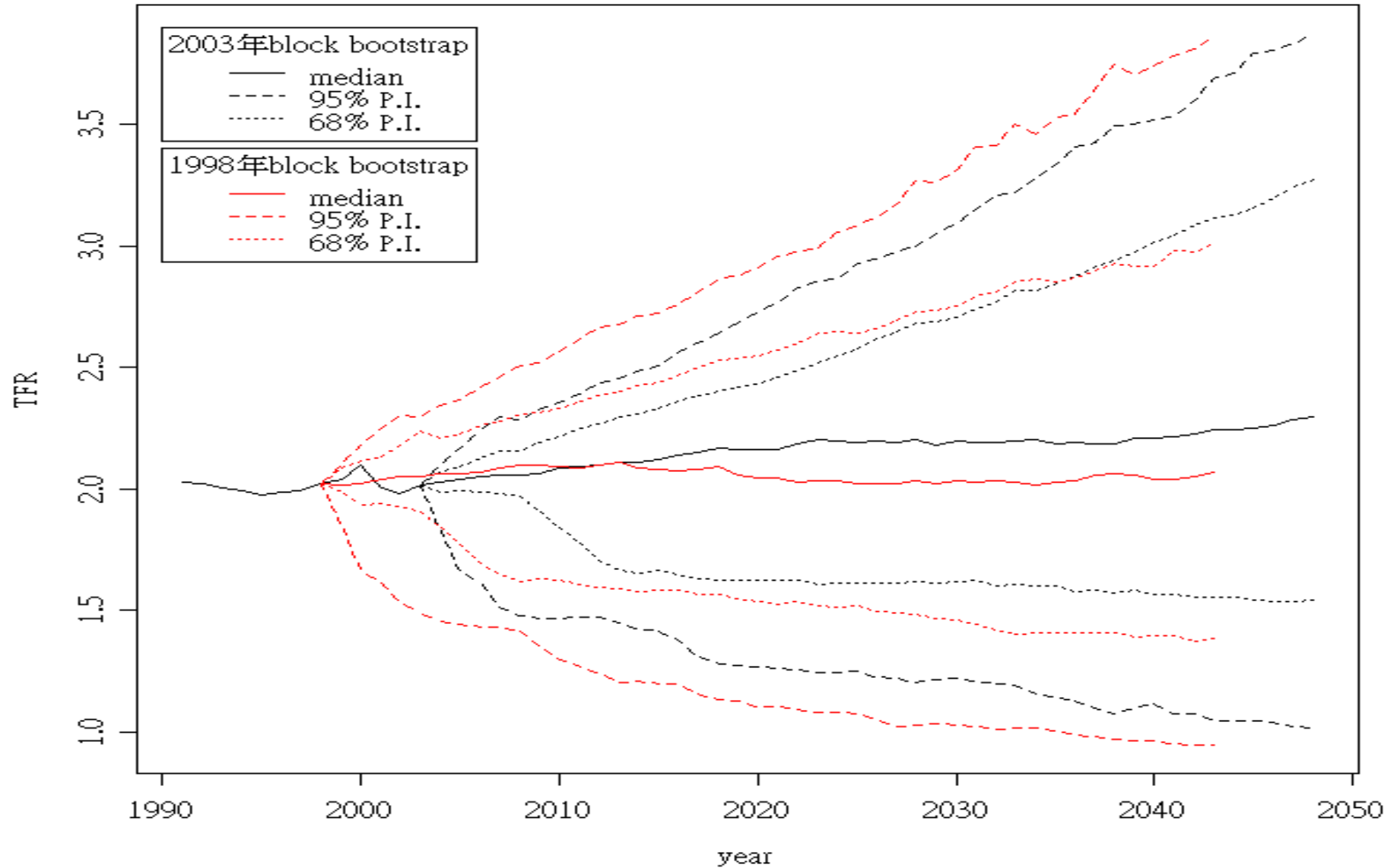
美國總人口預測

USA total population



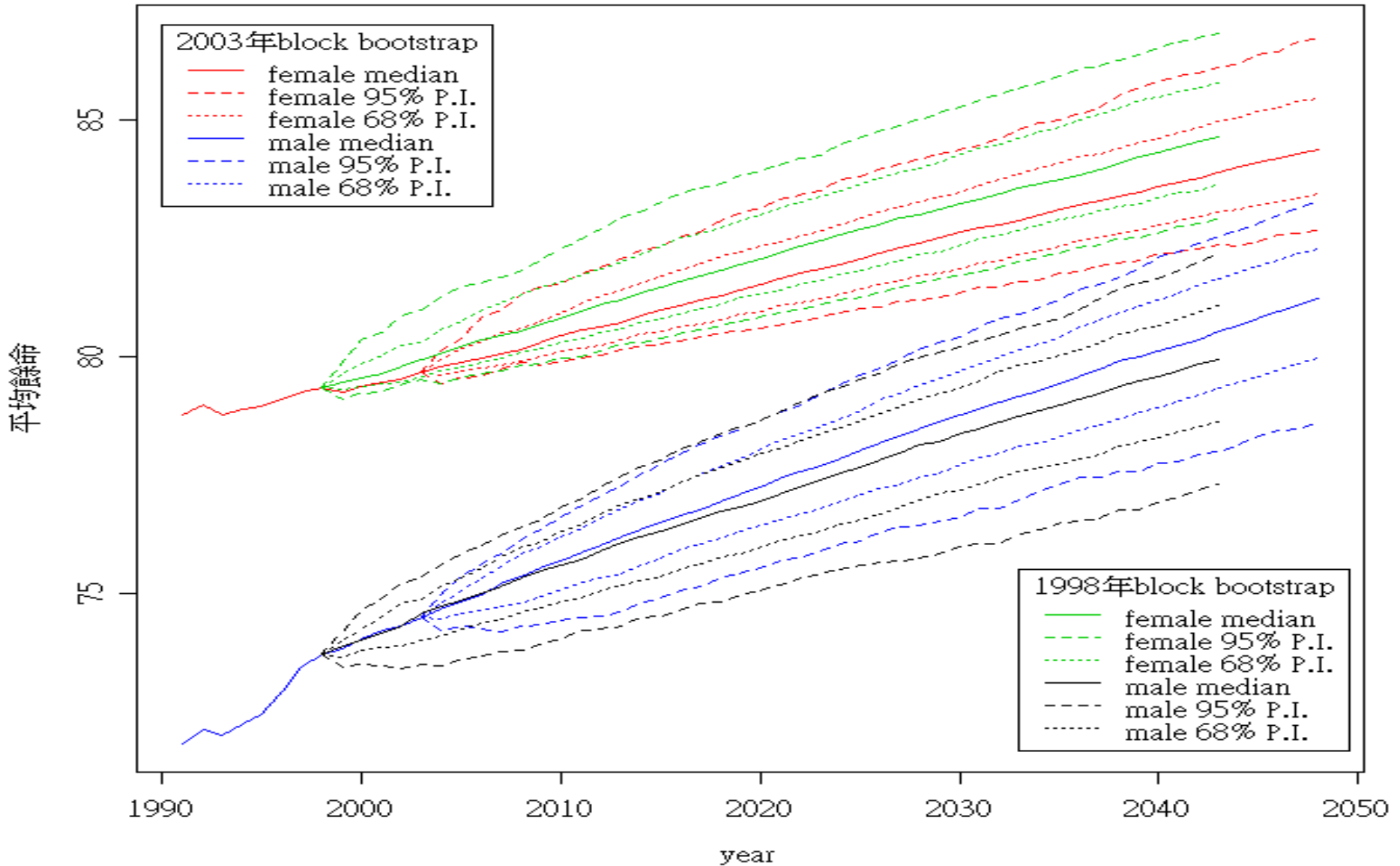
美國總生育率預測

USA TFR



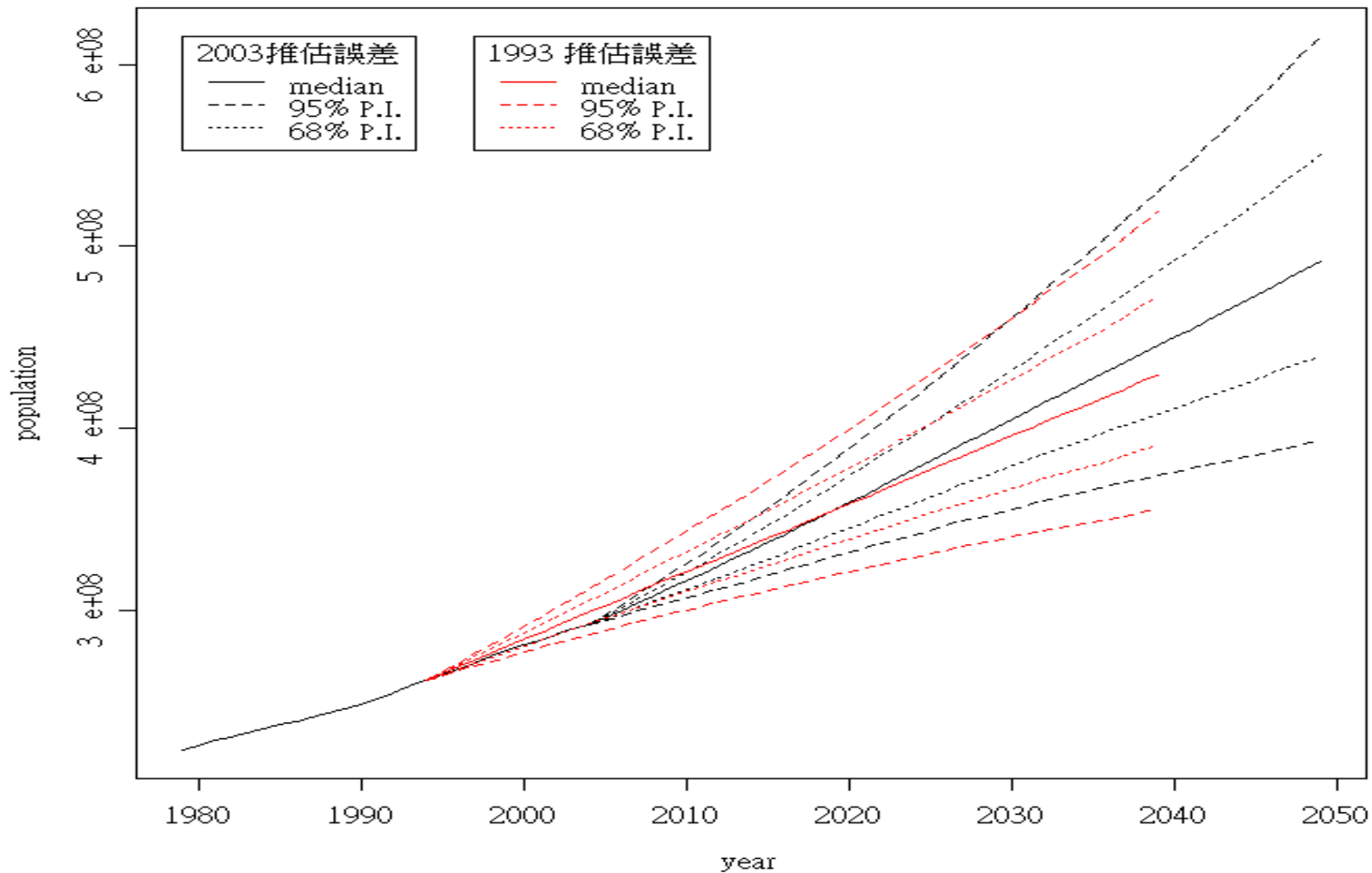
美國零歲平均餘命預測

USA 零歲平均餘命



美國人口推估 (Stoto修正)

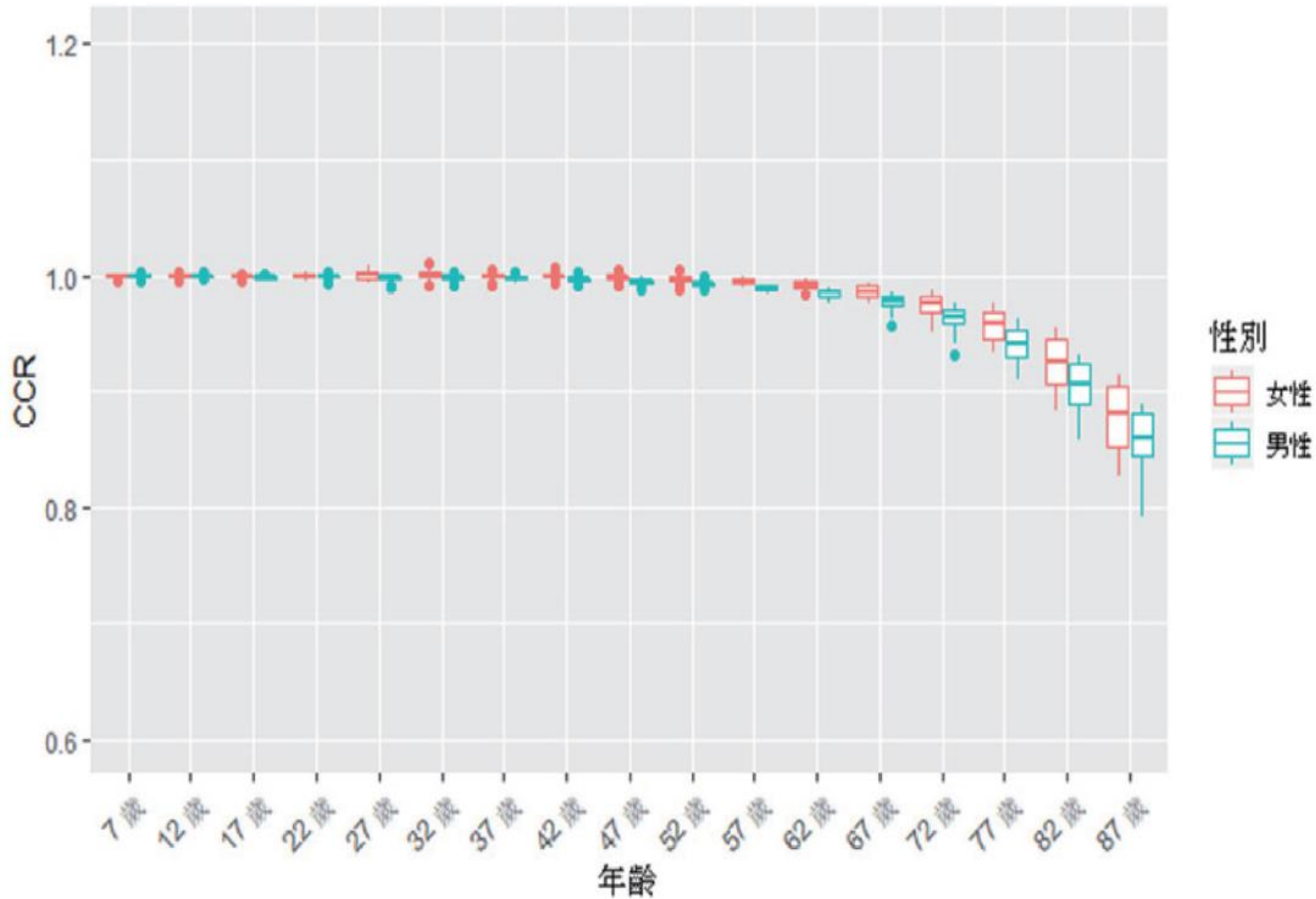
USA total population: 推估誤差 v.s. Block Bootstrap



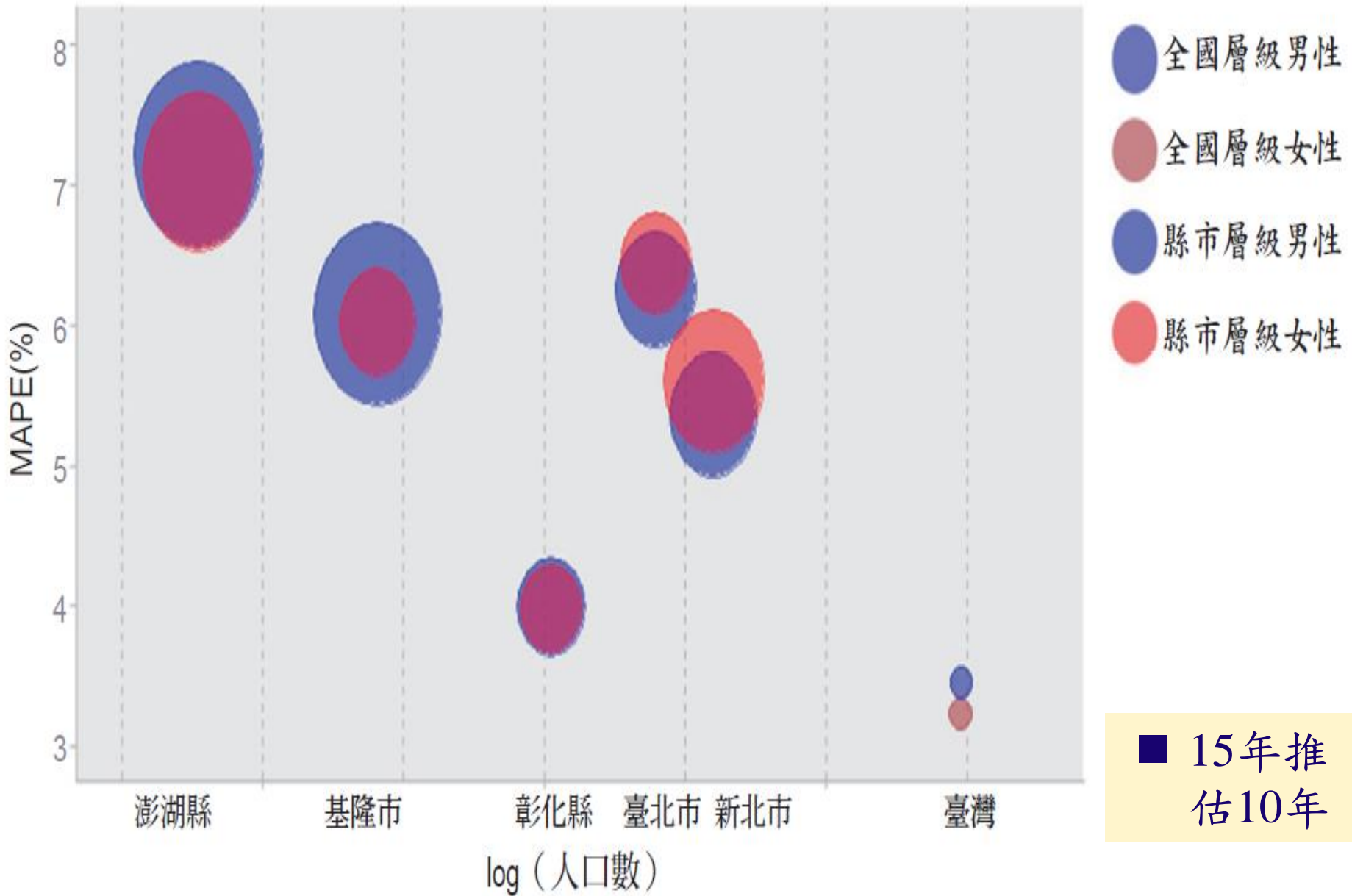
小區域(Small Area)人口推估

- 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。

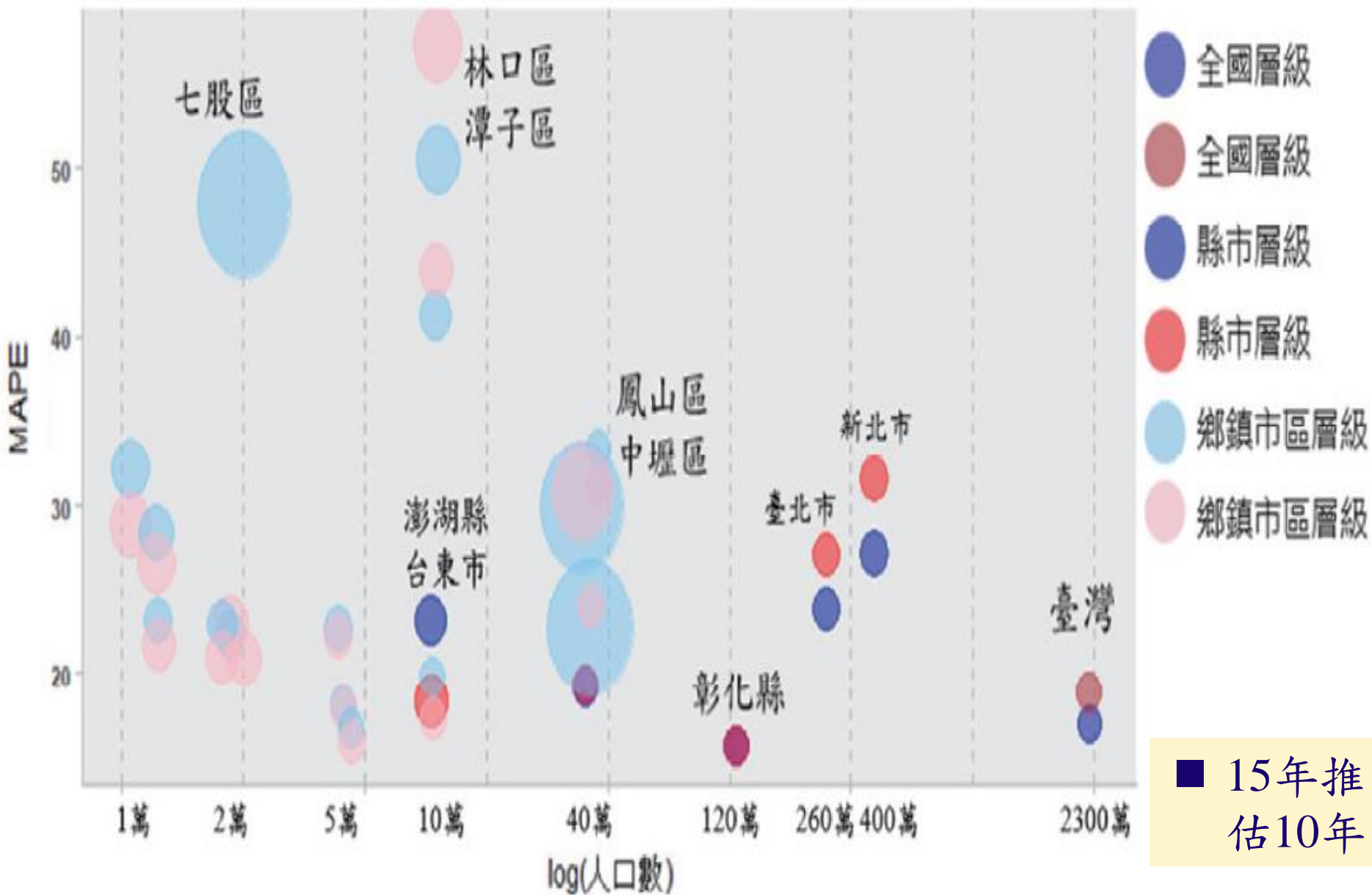
臺灣歷年5~84歲年輪變動比(CCR)



臺灣縣市HP法單齡推估誤差 (1975~2019年)



臺灣HP法五齡推估誤差 (1975~2019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》