

第五章、評估不同 APC 模型估計方法

我們亦透過與理論假設檢查類似的方式，觀察不同估計方法的差異。將模擬的死亡資料代入第二章所使用的各種估計方法，其設定仍然相同，所得到的參數估計值做為參數理論值進行電腦數據模擬。以各方法對台灣死亡資料的參數估計值為參數理論值，進行蒙地卡羅模擬，先以本質估計量的參數估計值為理論值，透過蒙地卡羅方式模擬卜瓦松分配下的死亡人數，再代入各方法估計參數值，比較不同估計方法在整體、年齡組和年代組的模型殘差與模型變異數。為方便討論，設台灣的死亡人數為 O_{ij} 、取對數後台灣的死亡率為 Y_{ij} 、蒙地卡羅的死亡人數模擬值為 O_{ij}'' 、蒙地卡羅取對數的模擬死亡率為 Y_{ij}'' 、各種估計方法估計 Y_{ij} 的估計值為 \hat{Y}_{ij} 以及估計 Y_{ij}'' 的估計值為 \hat{Y}_{ij}'' 。故對數的模擬死亡率估計與對數的台灣死亡率的殘差可寫成兩部分的加總，即 $\tilde{\varepsilon}_{ij} = Y_{ij} - Y_{ij}'' + Y_{ij}'' - \hat{Y}_{ij}'' = \hat{\varepsilon}_{ij} + \hat{\varepsilon}_{ij}''$ ，其中對數的模擬死亡率與台灣的對數死亡率的殘差為 $\hat{\varepsilon}_{ij} = Y_{ij} - Y_{ij}''$ ，對數的模擬死亡率估計值與對數的模擬死亡率之殘差 $\hat{\varepsilon}_{ij}'' = Y_{ij}'' - \hat{Y}_{ij}''$ 。主要目的在比較蒙地卡羅的模擬死亡資料與將模擬資料代入各方法後的估計結果的差異。另外，再以模擬的方式驗證 APC 各種估計方法的參數值結果變化大，但對死亡率的估計結果差異不大。

第一節、模型殘差與模型變異數

討論模型殘差大小、是否為常態分佈及殘差的獨立性。為避免與理論假設檢查部分的殘差與變異數混淆，對數的模擬死亡率與其估計值的差異稱為模型殘差。模型殘差其 $E(Y_{ij}'' - \hat{Y}_{ij}'') = E(\varepsilon_{ij}'') = 0$ ，計算如下： $\hat{\varepsilon}_{ij}'' = Y_{ij}'' - \hat{Y}_{ij}''$ 。無論男性或是女性，在各方法皆出現類似的情形，IE 與 cglim 的平均殘差最小，次為 AR 的平均殘差，序列法的平均殘差最大，其為 IE 與 cglim 的 10 倍。模型變異數，即

$$V(Y_{ij}'' - \hat{Y}_{ij}'') = V(\varepsilon_{ij}'') = \sigma^2, \text{ 其中 } \varepsilon_{ij}'' \sim N(0, \sigma^2), \text{ 計算式如下: } \hat{\sigma}^2 = \sum_{i=1}^{16} \sum_{j=1}^{10} (Y_{ij}'' - \hat{Y}_{ij}'')^2.$$

男性的模型變異數為女性的一半(表 7)，無論男性或女性皆顯示，IE 與 cglim 的模型平均變異數相同且最小，每次的模擬也都非常幾近，再來依序是 AR、序列法 ACP 及 APC，其中以 AR 的平均模型變異數與 IE 及 cglim 方法最接近。檢查整體的模型變異數是否服從常態分配，男性的 p 值介於 0.3~0.6，女性為 0.05~0.55，顯示男性的模型變異數更接近常態分佈。

表 7、平均的模型殘差與模型變異數

估計方法	模型殘差		模型變異數	
	男	女	男	女
IE	-0.00032	-0.00049	0.102701	0.200868
cglim-a	-0.00032	-0.00049	0.102701	0.200868
cglim-p	-0.00032	-0.00049	0.102701	0.200868
cglim-c	-0.00032	-0.00049	0.102701	0.200868
ACP	0.001787	0.005324	0.182034	0.366598
APC	-0.024598	-0.021355	0.769967	0.743065
AR	-0.000497	-0.000721	0.103251	0.202169

將平均殘差以等高線圖呈現，以評斷各方法估計各年齡與年代的情形，即衡量模擬的死亡率與模擬的死亡率估計值的差異，顏色愈深(藍、紫色)表示平均殘差值愈往負的方向，顏色愈淺(膚色)表示愈往正的方向，而黃色與綠色則是接近 0(圖 16~圖 17)。IE、cglim 與 AR 以年齡較輕且年代較長者估計較差，而序列法 ACP 明顯的在年代較輕者估計較差，序列法 APC 則是估計的最差。這結果顯示，除了序列法外，各方法在殘差和變異數的表現具有一致性。

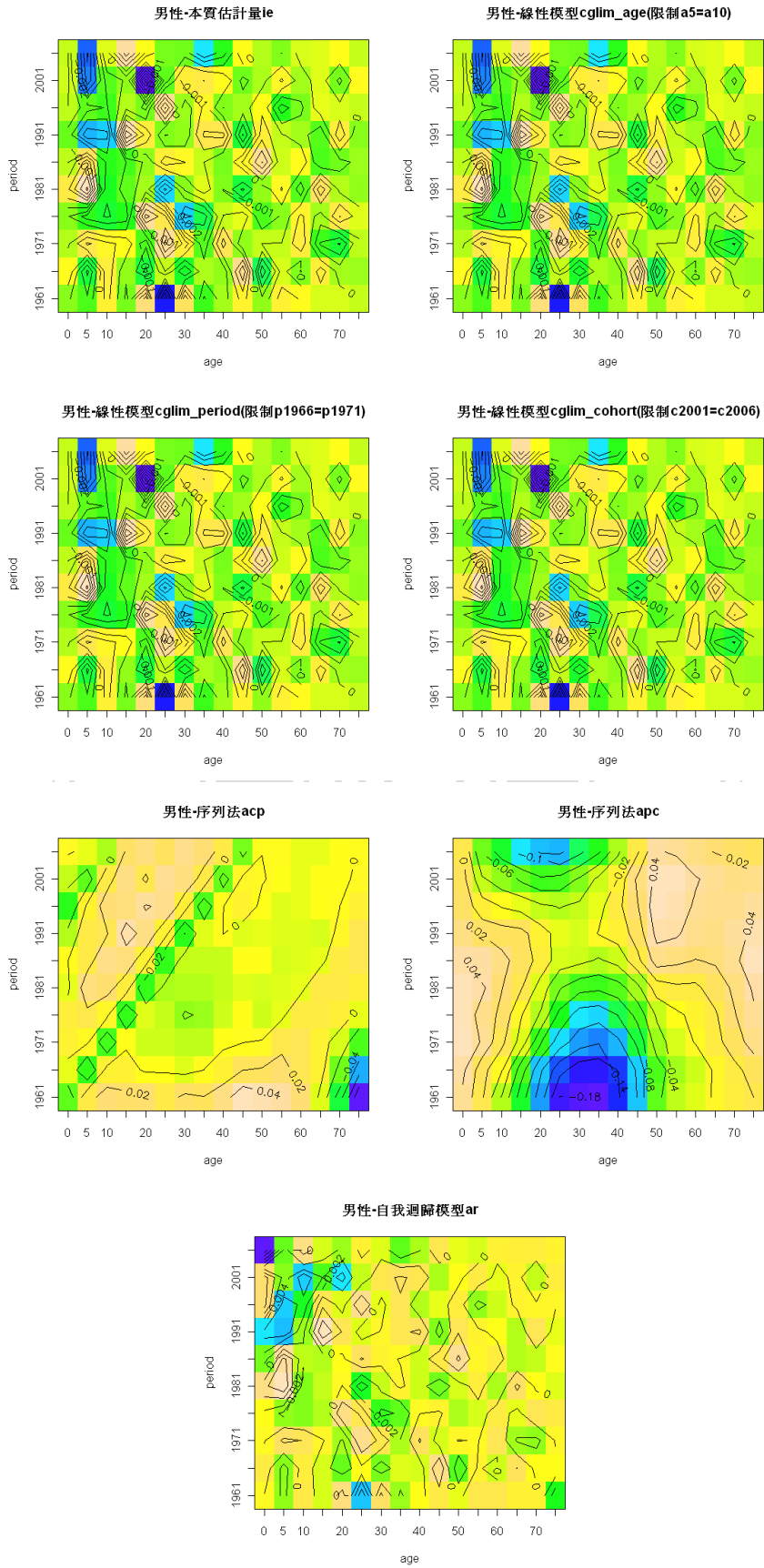


圖 16、男性各方法其平均模型殘差的等高線圖

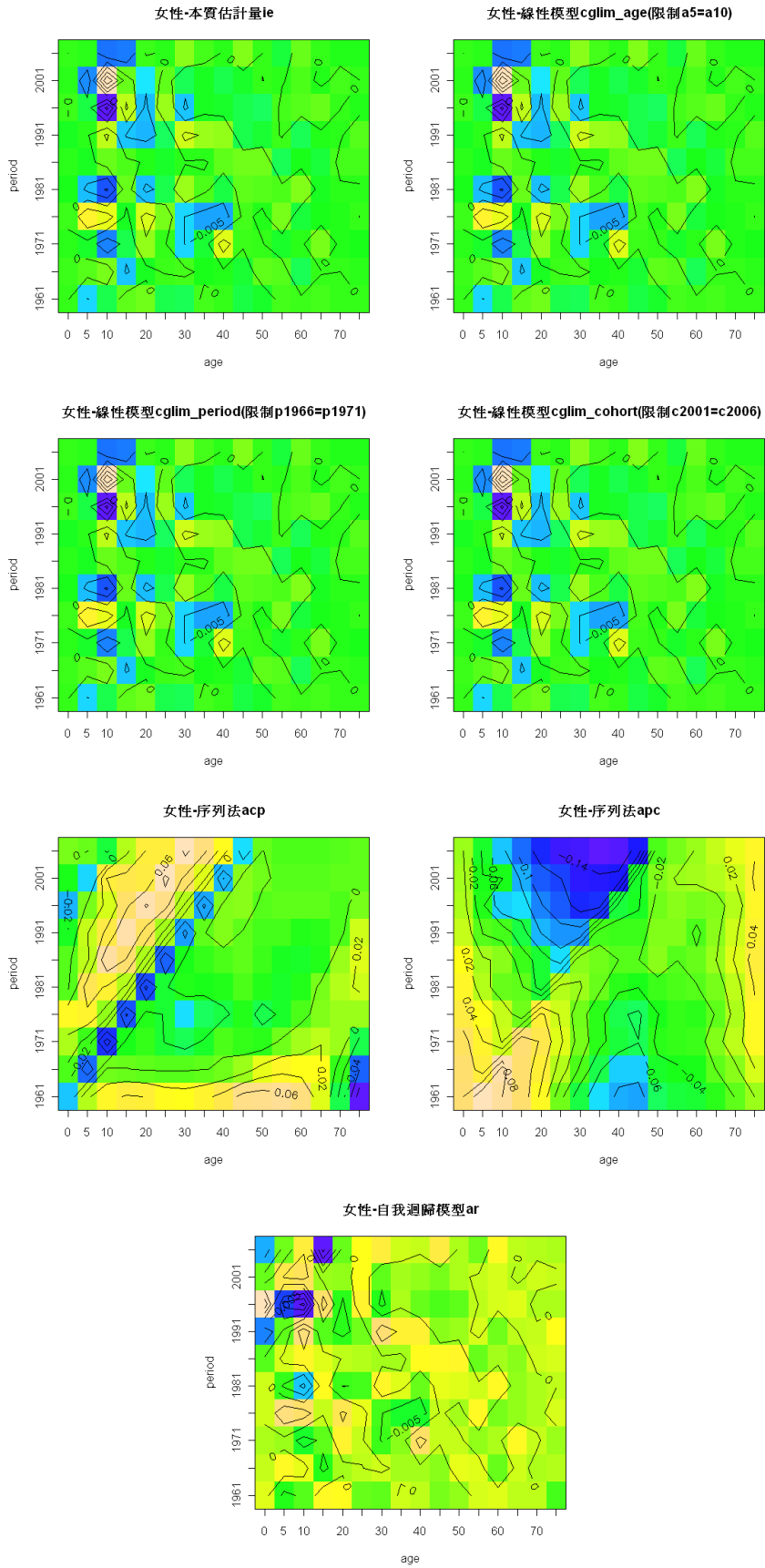


圖 17、女性各方法其平均模型殘差的等高線圖

以估計死亡率的而言，衡量各方法的模擬死亡率估計情形，以平均絕對誤差 (Mean Absolute Percentage Error, MAPE) 測量模擬的死亡率 m_{ij}^m 與其估計值 \hat{m}_{ij}^m 之間的差異。各方法的 MAPE 皆類似，以序列法較大些。模擬結果與第二章台灣死亡資料的 APC 估計結果大小順序類似，序列法稍為估計較差，其他方法皆類似，各方法在估計死亡率的表現很相近。

表 8、各估計方法死亡率的平均估計誤差(MAPE)

%	IE	cglim-a	cglim-p	cglim-c	ACP	APC	AR
男	1.6758	1.6758	1.6758	1.6758	2.3882	5.3493	1.6909
女	2.3473	2.3473	2.3473	2.3473	2.4145	5.1898	2.3699

檢查模型殘差是否為常態分配，顯著水準為 0.05 及 0.10 下的拒絕次數(表 9)，相對於其它估計方法，序列法的拒絕次數最多，故觀察其各方法之殘差的機率密度函數圖形(附錄三)，發現雖然不是符合常態分配，但為鐘形曲線且具對稱的特徵。同樣使用 Up-and-Down 檢定檢查獨立性。殘差 Up-and-Down 檢定在序列法拒絕比例較高，且序列法 APC 是幾乎拒絕，意即不符合殘差不具獨立性，序列法的遞減個數多於遞增個數，IE、cglim 及 AR 不符合獨立的模擬組數明顯較低。意謂不具常態分佈且不具有獨立性，配適能力較差的序列法反而較符合常態假設或是獨立性。

表 9、各估計方法之模型殘差 KS 檢定與 Up-and-Down 檢定拒絕次數

顯著水準	KS 檢定拒絕次數				Up-and-Down 檢定拒絕次數			
	<0.05		<0.10		<0.05		<0.10	
性別	男	女	男	女	男	女	男	女
IE	59	67	84	83	9	8	15	13
cglim-a	59	67	59	83	9	8	15	13
cglim-p	59	67	59	83	9	8	15	13
cglim-c	59	67	59	83	9	8	15	13
ACP	15	20	40	32	25	19	38	28
APC	38	3	73	15	100	83	100	90
AR	56	57	78	79	6	7	12	14

第二節、各年齡組與年代組的模型殘差與模型變異數

我們亦個別討論年齡組及年代組下的模型殘差與模型變異數大小，以及是否為常態分佈。將模型殘差取平均表示(圖 18)，以序列法 APC 偏離 $X=0$ 最遠，其餘方法則類似。年齡部分：無論男性或女性，在 20 歲~50 歲的年齡組對數的模擬死亡率嚴重小於估計值，為高估；年代部分：男性在愈年輕的年代組其高估的情形愈嚴重，女性則相反。

模型變異數為 $\hat{\sigma}_{age}^2 = \sum_{j=1}^{10} (Y_{ij}'' - \hat{Y}_{ij}'')^2, \forall i=1, \dots, 16$ 與

$\hat{\sigma}_{period}^2 = \sum_{i=1}^{16} (Y_{ij}'' - \hat{Y}_{ij}'')^2, \forall j=1, \dots, 10$ 。將模型變異數取平均表示(圖 19)，發現男性和

女性在年齡及年代的變異數表現不同，即不同年齡組和年代組其模型變異數非定值，尤以序列法 APC 的變異數變化最大。

整體來說，無論是年齡或是年代，皆以序列法 APC 變異數最大，次為序列法 ACP。除了序列法 APC 外，年齡方面有異曲同工之妙，皆以較年輕的年齡組之變異數較大，而年代方面較長的年代組有較大的變異數。在男性方面，除序列法外，年齡組大多符合常態分配，較年輕的年代組(1961)與較年長的年代組(1991~2006)其變異數不服從常態分配。在女性方面，除了序列法外，年齡組(15 歲、20 歲與 35 歲)的變異數不服從常態分配，年代組(1961 與 1991)亦不服從常態分配。

對個別年齡組及年代組的模型變異數之常態性檢定結果，發現大致符合常態分配。結果顯示各方法在殘差和變異數的個別表現，除了序列法外具有一致性，並且殘差的估計情形亦類似，序列法具有和其他方法不同的估計結果，可明顯觀察到增加變異的來源來自於那個年齡組或年代組，無論目標是在解釋或是估計死亡率，都會希望殘差能愈小愈好，雖然整體估計表現類似，以年齡組和年代組的殘差表現來看，序列法相對地不穩定。

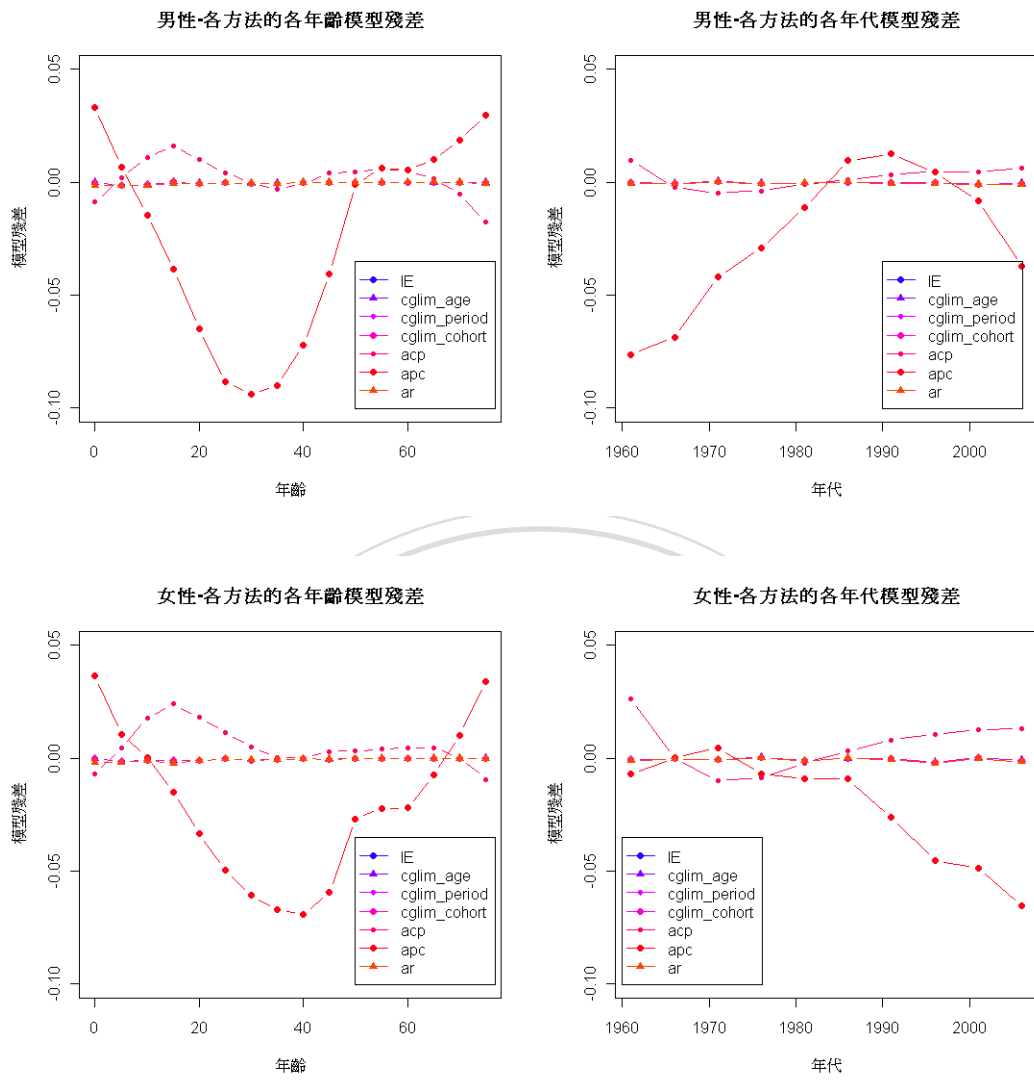


圖 18、年齡組與年代組的平均模型殘差

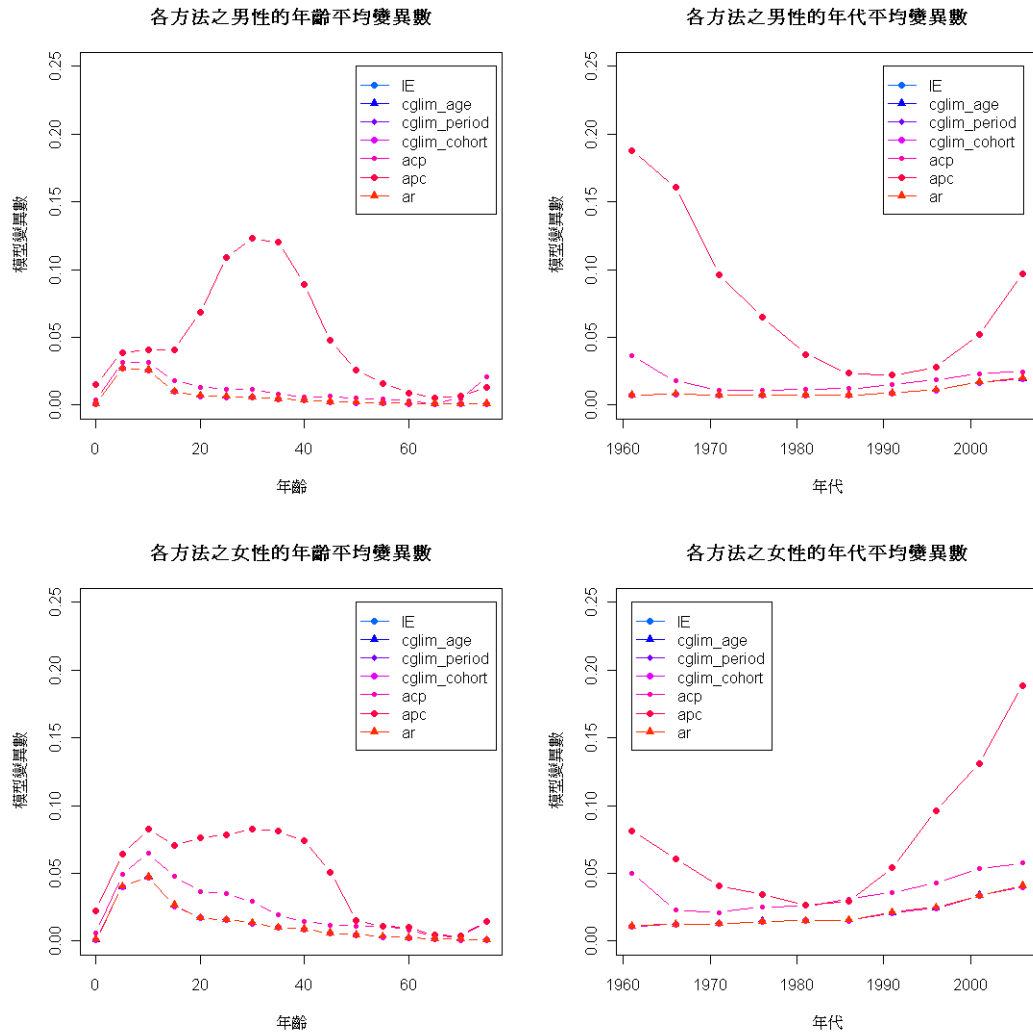


圖 19、年齡組與年代組的平均模型變異數

衡量各方法的在各年齡組與年代組的估計情形，以平均絕對誤差 (Mean Absolute Percentage Error, MAPE) 測量模擬的死亡率 m_{ij}^m 與其估計值 \hat{m}_{ij}^m 之間的差異 (圖 20)。各年齡組與年代組以序列法 APC 的 MAPE 較大，其次是序列法 ACP，但 MAPE 皆類似。由前一節的發現序列法在死亡率估計略遜，本節發現在個別年齡組及年代組在序列法的 MAPE 有顯著落差，其他方法皆類似。整體來說，各方法在估計年齡組與年代組的死亡率表現很相近，序列法確實配適較差。

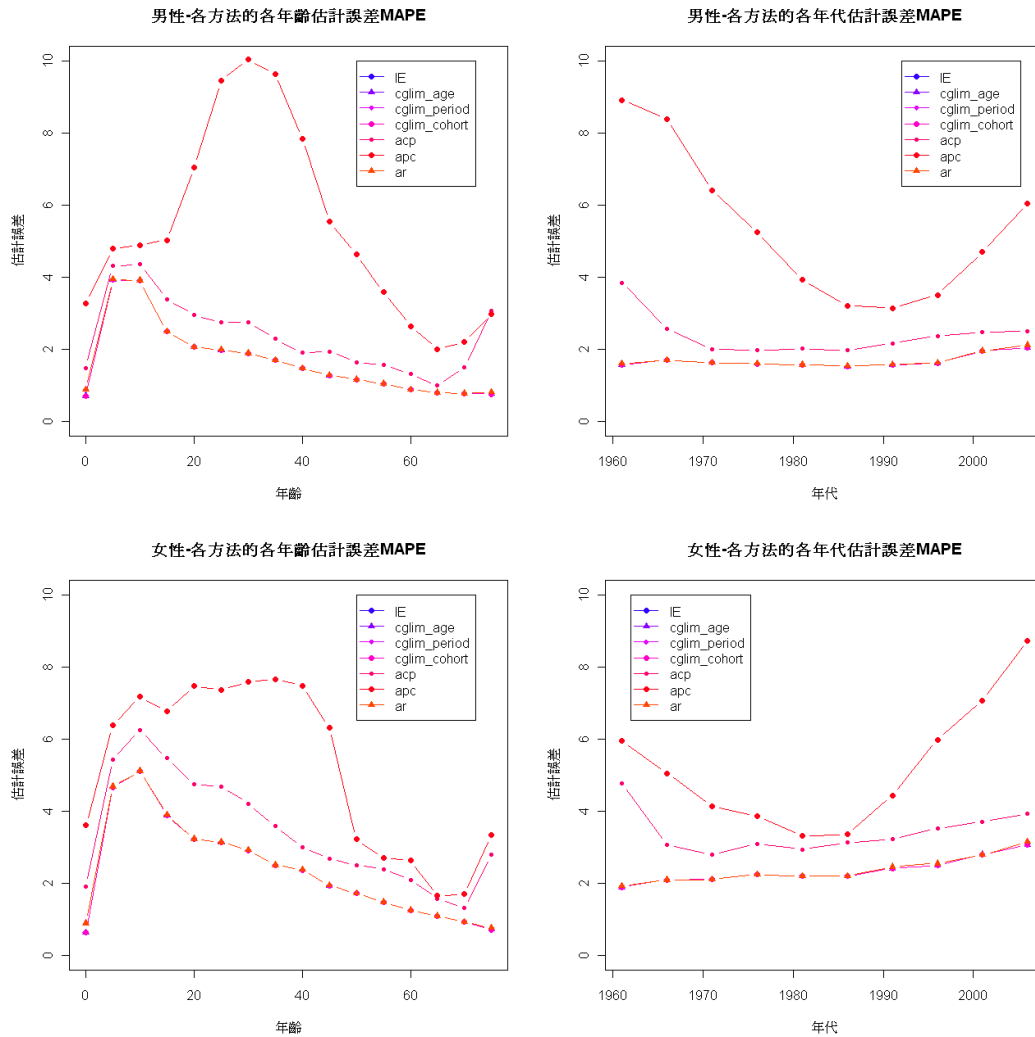


圖 20、年齡組與年代組的平均估計誤差(MAPE)

檢查年齡組及年代組下的模型殘差是否為常態分佈。顯著水準 0.05 和 0.10 下(圖 21)，無論男性或是女性，低年齡組和高年齡組的殘差較不符合常態，相對於年輕的年代組，年長的年代組的殘差明顯不符合常態，且年齡組的殘差符合常態的比例高於年代組。年齡組中，男性的殘符合常態的比例較女性高。年代組中，女性的殘差符合常態的比例較男性高。由於年齡組及年代組的殘差不符合常態比例甚高，從個別的機率密度函數圖可發現，殘差大致是集中型的鐘形分配，IE、cglim 與 AR 較類似，年齡較年代集中，高年齡組的較低年齡組的集中，年代組則是年長的年代組比較分散。相較於前者，序列法屬於分散型的鐘形分配，其中序列法 APC 較 ACP 分散，高年齡組較集中些，序列法的 0 歲組相對於其他方法分散，年代組亦有愈年長的年代愈分散的情形。前一節的整體常態分佈檢查，發

現序列法可能較符合常態分佈，若細部觀察年齡組與年代組的表現，本節發現不符合常態分佈的主要來自於年代而非年齡。

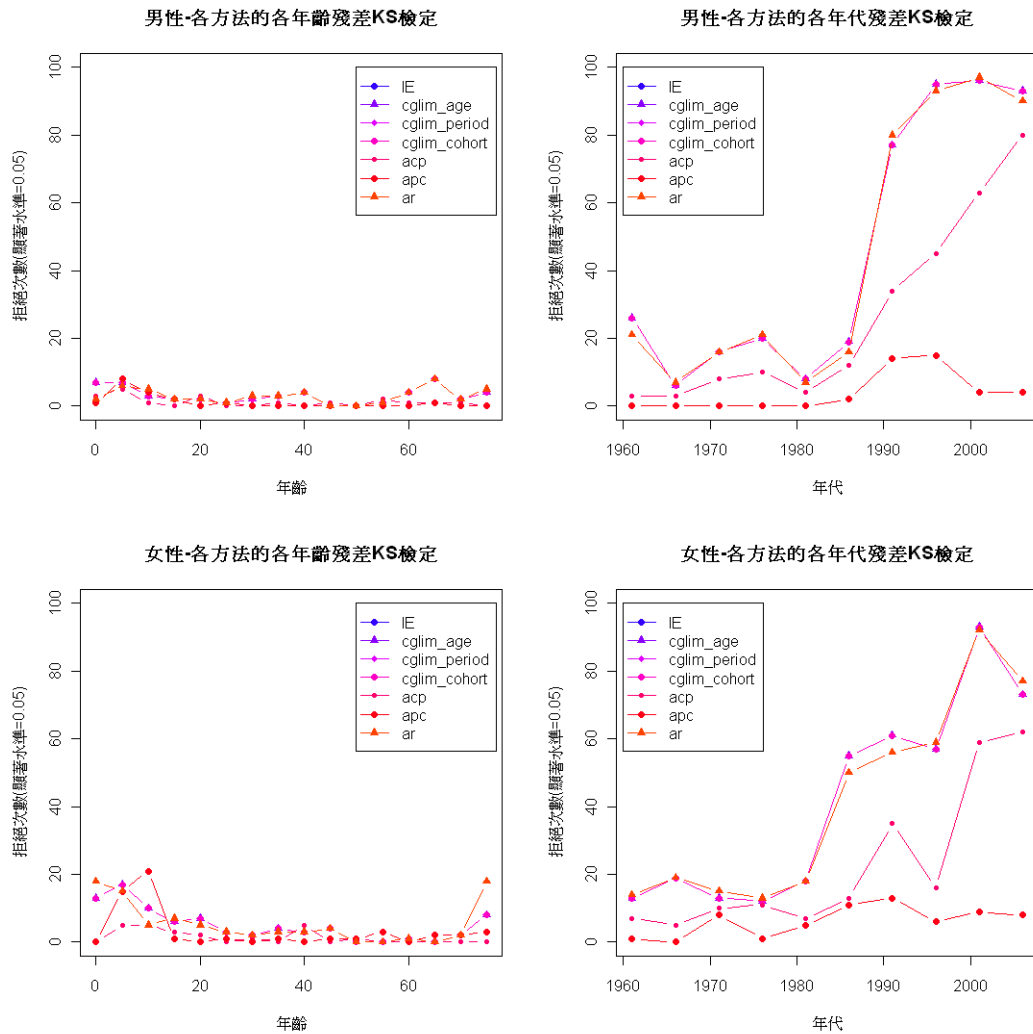


圖 21、年齡組與年代組之 KS 檢定的拒絕次數($\alpha = 0.05$)

同樣使用 Up-and-Down 檢定檢查各年齡組的模型殘差獨立性，即檢查年代間的殘差彼此是否獨立。與檢查整體的模型殘差結果相反，在序列法部分，其每次模擬的 16 個年齡組模型殘差不符合常態的比較相對較低，但有一半左右的模擬組數其年齡組的模型殘差不具獨立性，序列法則是遞增的個數多於遞減的個數。IE、cglim 及 AR 不符合獨立的模擬組數明顯較高，且遞增的個數多於遞減的個數。前一節顯示整體並非相當符合獨立性，本節發現各年齡組的獨立性檢查亦不符合，顯示同一年齡組內，不同年代的殘差具有關聯性。

表 10、各年齡組的模型殘差 Up-and-Down 檢定(拒絕次數)N=10

顯著水準	Up-and-Down 檢定			
	<0.05		<0.10	
性別	男	女	男	女
IE	98	95	98	98
cglm-a	98	95	98	98
cglm-p	98	95	98	98
cglm-c	98	95	98	98
ACP	56	35	68	53
APC	56	54	68	72
AR	97	97	97	98

第三節、參數於空間座標的表現

從第二章可以發現各方法的參數值估計結果差異甚大，但 MAPE 表現類似，本節將以模擬的方式來闡釋這可能的現象。觀察理論值與參數估計值的差異，稱為偏誤(bias)。使用兩種比較方式，其一為王郁萍(2006)所計算的參數值與理論值的差異，另一為比較平均參數估計值與真實值其相對位置的差異。

首先是王郁萍(2006)提及估計序列法 ACP 的參數，先配適 AC 模型，其假設 $\alpha_1 = \gamma_1 = 0$ ，因此 $\alpha_2, \dots, \alpha_{16}$ 與 $\gamma_2, \dots, \gamma_{25}$ 分別為 α_1 與 γ_1 的一階差異，將理論值減去第一個理論值，年齡、年代與世代分別有 15、9、24 個作比較，即 $\alpha_i - \alpha_1, \forall i = 2, \dots, 16$ 與 $\gamma_k - \gamma_1, \forall k = 2, \dots, 25$ 與 $\gamma_k - \gamma_1, \forall k = 2, \dots, 25$ 作為理論值與序列法比較，對參數估計值而言，可表示為 $\hat{\alpha}_i - \hat{\alpha}_1, \forall i = 2, \dots, 16$ 與 $\hat{\gamma}_k - \hat{\gamma}_1, \forall k = 2, \dots, 25$ ，故偏誤表示為 $\hat{\alpha}_i - \hat{\alpha}_1 - (\alpha_i - \alpha_1), \forall i = 2, \dots, 16$ 、 $\hat{\beta}_j - \hat{\beta}_1 - (\beta_j - \beta_1), \forall j = 2, \dots, 10$ 與 $\hat{\gamma}_k - \hat{\gamma}_1 - (\gamma_k - \gamma_1), \forall k = 2, \dots, 25$ ，序列法 APC 皆採用相同比較方式，而 IE 和 AR 採直接相減，為與序列法有相同組數，不比較年齡、年代與世代的第一組，即 $\hat{\alpha}_1 - \alpha_1$ 、 $\hat{\beta}_1 - \beta_1$ 與 $\hat{\gamma}_1 - \gamma_1$ 。基於參數估計值皆代入同一模型估計死亡率，或許應統一比較方式，比較估計的參數值與理論值的相對位置或許可行，與前述序列法參數偏誤相同。

圖 22 為王郁萍(2006)所使用的比較方式，分別為年齡、年代與世代的參數值比較，上為男性，下為女性。直接比較參數估計值與 IE 理論值的差異，IE 為不偏估計，偏誤近為 0，其餘方法皆隨年齡、年代及世代而偏誤變大，主要差異於偏誤增大的速率不同及方向不同，AR 的偏誤對稱於 X 軸。

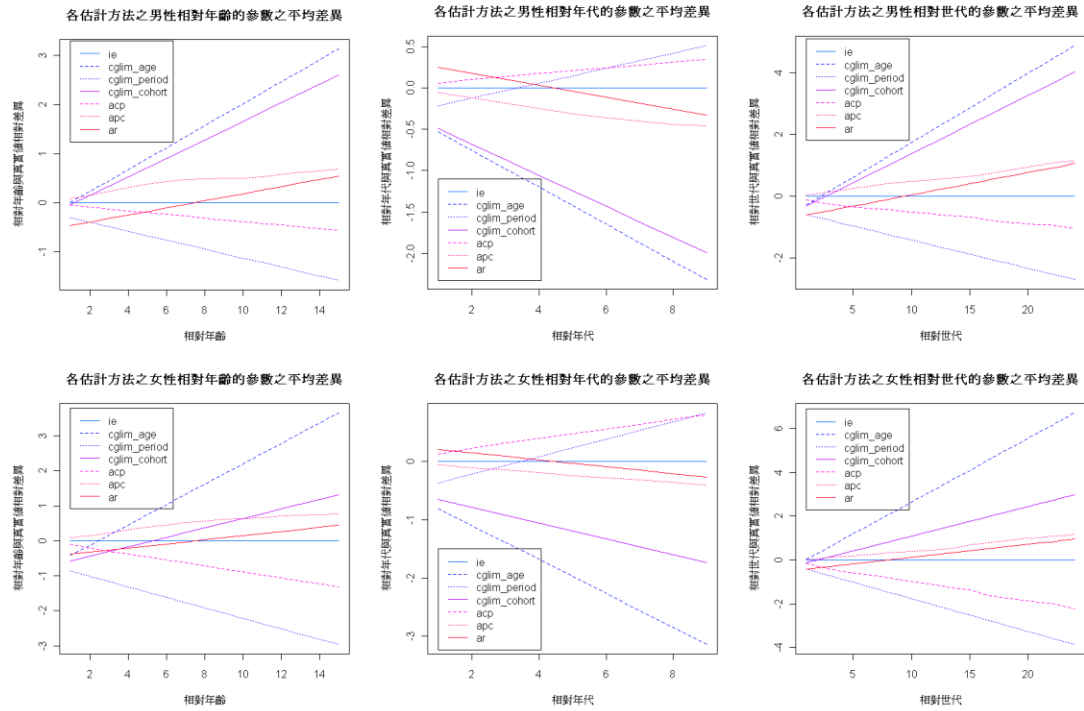


圖 22、年齡、年代與世代參數與理論值比較的相對誤差(王郁萍)

圖 23 為相對位置的比較方式，上為男性，下為女性。以相對位置的差異來看，無論男性或女性，IE 的估計值與理論值的相對位置皆同，其餘方法皆隨著年齡、年代與世代而擴增，尤其是在女性的資料部分，序列法 APC 與 AR 的相對位置較相像。

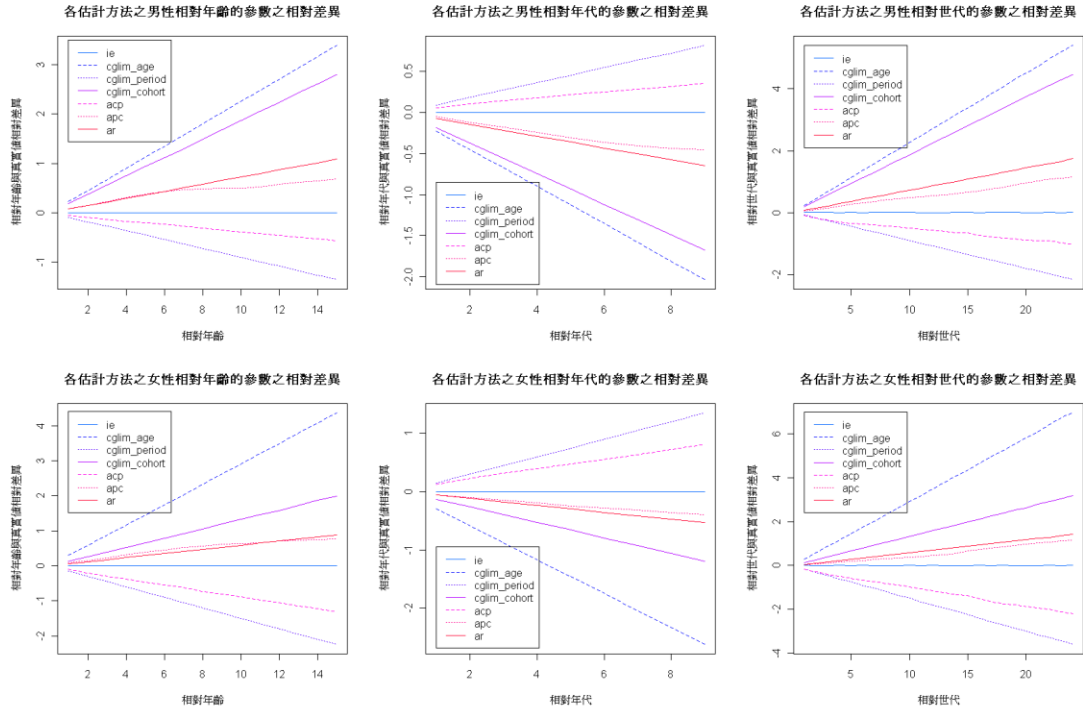


圖 23、年齡、年代與世代參數與理論值的相對位置比較

以本質估計量的參數估計值為理論值，在前述的觀察下，參數理論值設定是否影響估計結果？參數值比較觀點下，IE 確實較好；估計死亡率觀點下，各方法的估計結果差異並不大，所以懷疑 IE 為理論值是否會偏向使 IE 的結果更好。故再以台灣死亡資料代入 AR 方法的參數估計值為參數理論值，產生模擬資料，再代入 IE 及 AR 的方法，以相同的方式比較參數值差異、蒙地卡羅變異數及模型變異數，並找出 IE 和 AR 何者較好。由表 11 發現以無論是以 IE 或是 AR 為理論值其蒙地卡羅變異數類似，且 IE 和 AR 的模型變異數也不受影響，IE 皆比 AR 小一點。為確認理論值設定可能造成的影響，使用序列法為理論值做模擬，比較 IE 與 AR 兩方法，序列法的蒙地卡羅變異數與前述類似，將資料代不同的方法估計的模型變異數皆以 IE 和 AR 最小，與前述的發現有異曲同工之妙。以某方法所產生的參數估計值為理論值模擬死亡資料，再代入該方法求參數估計值，該方法的參數估計值有不偏的性質，其他方法參數估計值則與理論值有相當大的偏誤。

表 11、平均的蒙地卡羅變異數與平均的模型變異數(IE vs AR)

真實參數	本質估計量 IE		自我迴歸 AR		序列法 ACP		序列法 APC	
	男	女	男	女	男	女	男	女
$MC(\hat{\varepsilon}_{xt})$	2.3945	4.2079	2.3818	4.1515	2.6438	4.9802	2.7283	3.7498
$IE(\hat{\varepsilon}_{xt}^n)$	0.1027	0.2009	0.1014	0.2107	0.1057	0.2085	0.0960	0.1895
$AR(\hat{\varepsilon}_{xt}^n)$	0.1033	0.2022	0.1020	0.2123	0.1065	0.2105	0.0970	0.1918
$ACP(\hat{\varepsilon}_{xt}^n)$	0.1820	0.3666	0.1789	0.3769	0.1641	0.3009	0.2090	0.3571
$APC(\hat{\varepsilon}_{xt}^n)$	0.7670	0.7431	0.7626	0.7530	0.8646	0.9291	0.4588	0.4305

進一步，觀察 IE 和 AR 為理論值下，參數估計值與理論值的差異(直接相減)。無論以 IE 或 AR 的參數估計值為理論值，以相同的方法做估計，其參數估計值與理論值的差異趨近於 0，即以 IE(或 AR)的參數估計值為理論值，模擬的死亡資料，再代入 IE(或 AR)方法所得的參數估計值與理論值的差異很小。IE 方法下的參數估計值為理論值，所模擬的死亡資料代入 AR 方法，其參數值與理論值的差異與 AR 為理論值產生資料再代入 IE 方法的參數值與理論值差異，兩種差異值取絕對值後類似，像是在比較相對位置(圖 24)。證實比較參數的不偏估計與參數設定有極大的關聯性，所得到的結果是比較參數的相對位置差異非絕對位置。

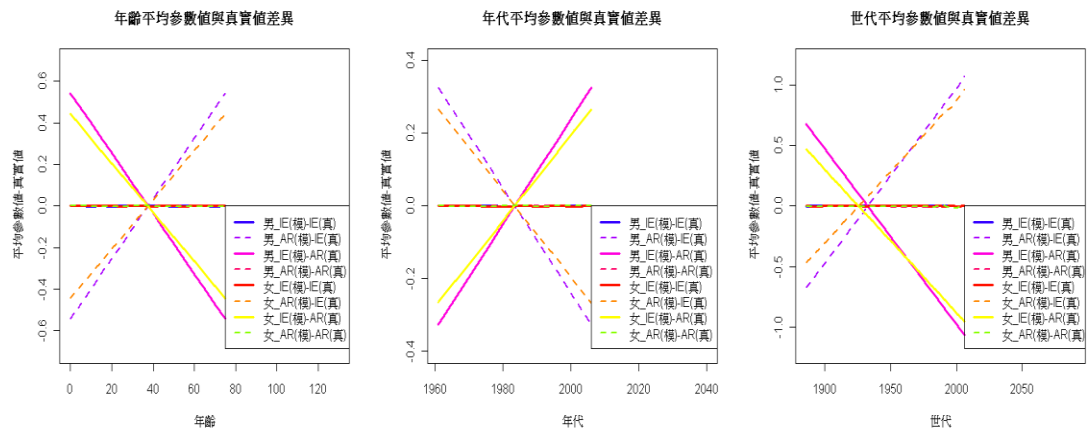


圖 24、年齡、年代與世代之 IE 與 AR 參數估計值比較

由本章的結果，發現各估計方法所估計出來的參數會有相當大的差異，但在死亡率的估計具有一致性的估計結果(MAPE)，這也告訴我們，一開始以 IE 方法的參數估計值為理論值檢驗模型理論假設，其檢驗結果發現理論假設不全然符合，當然再以其他方法的參數估計值為理論值檢驗模型假設，也會發現理論假設不全然符合。