

第二章 研究設計

第一節 研究方法與研究架構

本研究所採取的研究方法是電腦模擬法，以既有的調查研究資料為母體，反覆地樣本抽取，進行參數推估，計算所得之參數平均數與母體參數之間的偏誤和參數變異數，所有抽樣樣本數配對皆以此方式進行，最後比較各抽樣條件的相對偏誤及變異數，數值愈小表示此一抽樣條件較佳，反之，則代表抽樣條件所推估出來的參數較不準確，而為不良之抽樣配對。

電腦模擬法的執执行程序如下，先針對母體原始資料進行整理，後續模擬程序則以撰寫電腦程式執行，主要的軟體為 HLM6.0 版進行多層線性迴歸分析的各项參數估計。估計法的選用如第一章第四節文獻檢閱中所提，一般說來在多層線性模型的估計上有 MLR 和 MLF 二法，本文採用 Browne and Draper(2000)的說法，認為 MLR 的估計通常較 MLF 準確，故在 HLM6.0 的設定上以 MLR 法進行各項參數估計。由於 HLM6.0 版本不僅容許以視窗點選方式進行，亦可以在 DOS 作業系統下以批次檔(batch file)來執行，因此自形成母體資料、抽取樣本、進行參數估計、儲存資料等都使用 C++ 程式語言整合所有模擬過程，詳細流程請見圖一。

開始抽樣前，必須進行母體資料製作，不論是教育學或政治學的原始資料裡，每個群體內的個體樣本數大小都不相同，為了避免抽樣加權問題且讓每個個

體的抽樣機率一致，以隨機抽樣的方式將個體樣本數統一成固定數目¹。筆者針對教育學資料，形成總體樣本 100 個、個體樣本總數 10000 個(100×100)的母體資料，依此一母體資料，另行製作中、高兩種模型解釋力的個體樣本，有了三種個體資料後，再分別進行各群體的 OLS 參數估計，製作出高、中、低三種適合度的總體層次樣本，故共有九個兩層次解釋力不一的母體樣本，可比較不同母體資料特性對於各參數估計的影響。另一方面，從原始的政治學調查研究資料統一形成 1500/150 的母體，根據上述母體，便可算出各國的迴歸參數真值 β_j 、固定效果 $w\gamma_j$ 及隨機效果 u_j 。

其次是抽樣條件的選定，從過去的相關模擬研究中發現，不論樣本組合數為何，總樣本數愈大，所估計出來的參數表現愈好(Mok 1995)，為了控制此變因，不讓總樣本數干擾研究結果，因此本文將研究課題限定於總體個數乘上個體個數的總樣本數不變之情況，評估各項參數在控制總樣本數後的表現。Kreft (1996) 曾建議個體層次及總體層次樣本數都訂在 30 個樣本以上，才能獲得較為穩固的估計，故以此為準，選擇固定總樣本數為 900(30×30)，使得個體與總體樣本數產生五種不同抽樣配對，各自是 10/90、20/45、30/30、45/20、90/10，來檢視個體層次的樣本數較重要，或者是總體層次的樣本數才是研究者不容小覷之處。

第三個步驟是進行模擬程序，按照不同的抽樣條件(群體/個體樣本數組合)

¹ 當母體資料龐大或蒐集困難，但研究又希望節省成本時，多採群集抽樣法(cluster sampling)，先抽取總體層次樣本，進而抽取個體層次樣本，因此常又被稱為次群集抽樣(subsampling)或多階段性抽樣(multistage sampling)，本研究正是以此法進行模擬實驗。群集抽樣法可依群體內個體數目是否相同分為「等量」(equal clusters) 及「不等量」(unequal clusters)兩大類，若資料型態為不平均的群集，按照 Kish(1965: 183)所分析，在選樣上便會產生以下四個問題：第一、為了使每個個體被抽中的機率相同，個體數目的大小勢必將成為決定總體層次樣本選取中的隨機變數，增加了抽樣設計的不確定性；第二、縱使是良好的估計，其比例均數(ratio mean)並不同於母體平均值；第三、變異數估計式不是變異數真值的不偏估計；第四、變異數估計式相當複雜，不易計算。本研究為了簡化研究問題，特將原始資料進一步處理成母體資料，以等量模型(equal cluster model)的方式進行各項實驗。

來反覆抽出 1000 個樣本²，然後分別執行 1000 次的多層線性迴歸分析，算出每個國家在抽中樣本中的迴歸參數值、固定效果、隨機效果等各參數平均值及變異數，在統計術語中，此法稱為再次取樣法(resampling method)。最後則是評估參數表現，通常參數的準確與穩定是研究者的基本要求，因此筆者也將針對此二部分進行討論，由於母體參數已知，參數的準確性可採相對誤差越小越好的標準 (relative bias)，而穩定度使用得以顯示離散程度的變異數來表示。相對偏誤的作法是拿母體參數真值做分母，使用 HLM 所推估出來的參數平均值與母體參數真值之差為分子，即可得偏誤與母體參數之間的比例，經過此一標準化過程，便可進行跨參數、跨模型的比較。迴歸參數的相對偏誤以數學式表示如下：

$$BIAS_{\beta} = \frac{1}{J} \sum_{j=1}^J \left\{ \left| \frac{\bar{\beta}_j - \beta_j}{\beta_j} \right| \right\} \quad [2.1]$$

固定效果的相對偏誤數學式為

$$BIAS_{\gamma} = \frac{1}{J} \sum_{j=1}^J \left\{ \left| \frac{\bar{\gamma}_j W_j - \gamma_j W_j}{\gamma_j W_j} \right| \right\} \quad [2.2]$$

而隨機參數的相對偏誤數學式為

$$BIAS_u = \frac{1}{J} \sum_{j=1}^J \left\{ \left| \frac{\bar{u}_j - u_j}{u_j} \right| \right\} \quad [2.3]$$

[2.1]、[2.2]、[2.3]中，各項樣本平均參數 $\bar{\beta}_j$ 、 $\bar{\gamma}W_j$ 、 \bar{u}_j 是依各群體在 1000 回抽樣樣本裡被抽中的次數加以平均，因此各群體所平均的次數不盡相同，例如第一個群體於 1000 回樣本內被抽中次數為 800，便以此 800 個估計值估算平均數，而第二個群體抽中了 750 次，則平均 750 個估計值。至於變異數，同樣從個別群體的角度出發，個別計算參數離散程度，最後再平均群體總個數。

² 本研究中除以政治學世界價值調查大型資料(1500/150)為母體的實驗外，其餘皆抽取 1000 個抽樣樣本，進行 1000 次參數推估。由於大型資料的世界價值調查較接近實際政治學研究所遭遇之資料型態，筆者為追求更準確可靠的模擬研究結果，故將模擬次數擴大為 10000 次，也就是抽取 10000 個抽樣樣本，進行 10000 次參數估計，而後求其平均之參數值。

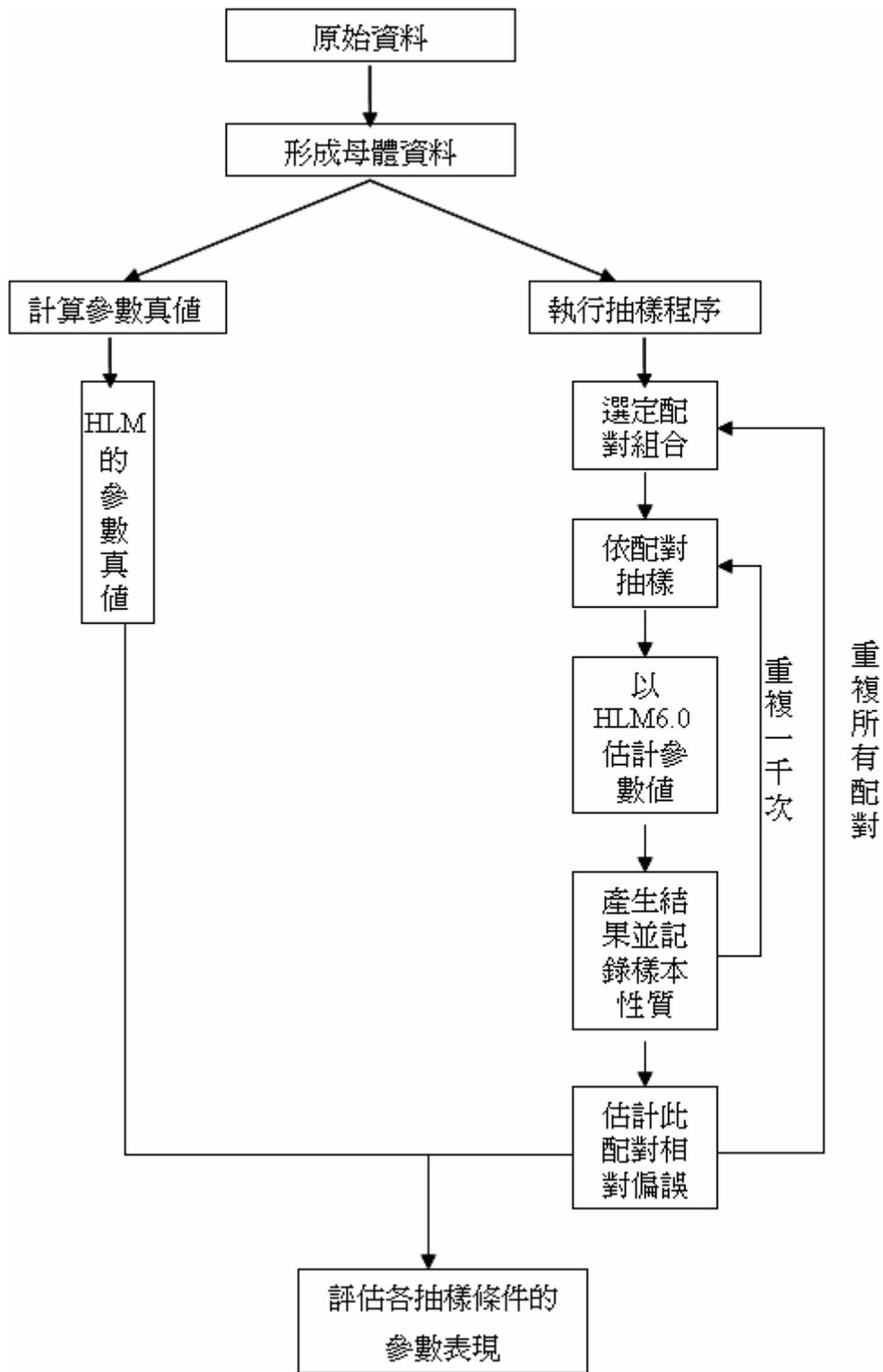


圖 2-1 本模擬研究之流程圖

第二節 資料來源與模型設定

回顧所有研究兩層次迴歸模型樣本配置問題的文獻，大多由教育學者完成，政治學界的方法學者尚未發表任何相關研究成果，筆者認為在沒有針對教育學者所使用的資料進行詳細討論的情況下，冒然地使用其研究結論，實為不妥，因此本研究同時使用二領域既有之調查研究資料，先以教育學資料來驗證之前文獻結論的可靠性，再用政治學界資料加以佐證，方能確保結論之真確。

在教育學資料取用方面，採「台灣高等教育資料系統之建置與運用」研究計畫，由彭森明教授主持，國科會贊助及教育部支持，國立師範大學教學評鑑與發展研究中心所執行，計畫執行時間自民國九十二年七月至民國九十五年八月止。此計畫內容相當豐富，筆者擷取其中 92 學年度大一問卷、92 學年度大三問卷以及 94 學年度大一問卷等以學生為本位的研究資料，依學生科系代碼為分類依據，整理歸併之後，實際進行模擬研究的科系個數有 111 個、學生總人數為 25763 名。

在模型設定上，過濾三份問卷共有的問題，最終使用一個個體層次自變項及一個總體層次自變項的簡單模型，以「學生修習學分數」為依變項，「平均每週工作時數」做為個體層次自變數，總體層次自變項則是「由父母或親戚支付生活費之金額」³。假設學生修習學分數與平均每週工作時數為負相關，二者為相互排擠之效應，而常數項與此共變皆可由群體的父母或親戚支付生活費之金額平均來解釋。所有變項放入模型時，全經標準化處理，減去變項均值，為求討論方便，以數學模型表示其關係為

³ 本文著重於方法論上的探討，至於實際的教育學理論不在討論之列，有興趣之讀者敬請自行參閱教育學研究等相關文獻。

$$\begin{aligned}
Y_j &= \beta_{0j} + \beta_{1j}(X_j - \bar{X}_j) + e_j, e_j \sim N(0, \psi) \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}(W_j - \bar{W}) + u_{0j}, u_{0j} \sim N(0, \tau_0) \\
\beta_{1j} &= \gamma_{10} + \gamma_{11}(W_j - \bar{W}) + u_{1j}, u_{1j} \sim N(0, \tau_1)
\end{aligned}
\tag{2.4}$$

政治學調查資料採用的是「世界價值調查」(World Value Survey)，此研究資料是由美國密西根大學 Ronald Inglehart 教授主持，總共有四波調查，時間分別為 1981、1990、1995-1997、1999-2000 年，筆者合併四波 202 個地區或國家樣本，扣除地區樣本以及遺漏值過多的 67 個國家樣本，實際進行模擬研究的國家個數為 135 個。針對每個國家樣本，以心理計量量表方式製作「政治參與度」、「政治信賴感」、「道德價值觀」、「宗教虔誠度」、「社會容忍度」等五個變項，各變項皆由調查中的四或五項問題所組成。

我們採用「全資訊項目因素分析」(Full-Information Item Factor Analysis)來進行二分項目的因素分析，測試各變項的「單面向性」(unidimensionality)假設，剔除因素負荷過低的項目，最後所形成的結果如表 2-1 所示，而遺漏值的回答則以相近的回應模式(response pattern)估計之⁴。至於國家層次的變數，選取人均國內生產毛額來詮釋各國彼此之間的差異，此變項是參考世界發展指標(World Development Indicators)⁵中所公布的國家數據，按照世界價值觀各波的調查時間所建立。

⁴ 關於詳細做法，參見 Huang(2004), p.282。

⁵ 世界銀行(World Bank)每年定期發表世界發展指標，提供了社會與經濟的第一手數據資料，從 1960 年追蹤至今，世界銀行亦提供線上資料庫供讀者查閱：<http://devdata.worldbank.org/data-query/>，本文主要引用資料庫中的國家人均生產毛額。

表 2-1 心理量表所形成之個體層次自變項問卷項目與因素分析結果

變數	題號	題目敘述	難度	萃取	成份(1)
政治參與度	V134	我曾經或我願意簽署請願書。	-.432	.590	.768
	V135	我曾經或我願意參與罷工活動。	.289	.671	.819
	V136	我曾經或我願意前往合法的示威。	-.045	.692	.832
	V137	我曾經或我願意參與不合法的罷工。	.791	.672	.820
	V138	我曾經或我願意佔領建築物或工廠。	1.122	.567	.753
政治信賴感	V148	我對軍隊有信心。	-.226	.237	.487
	V152	我對警察有信心。	-.14	.386	.621
	V155	我對立法機關有信心。	.226	.596	.772
	V156	我對文官體系有信心。	.132	.575	.758
道德價值觀	V208	同性戀的合理性程度，請以 1~10 表示之。	-.746	.367	.606
	V210	墮胎的合理性程度，請以 1~10 表示之。	-.651	.635	.797
	V211	離婚的合理性程度，請以 1~10 表示之。	-.33	.643	.802
	V212	安樂死的合理性程度，請以 1~10 表示之。	-.364	.428	.654
	V213	自殺的合理性程度，請以 1~10 表示之。	-1.092	.450	.671
宗教虔誠度	V185	受訪者提及清真寺是他們非常信任的社會組織之一。	.399	.515	.717
	V186	不管我去不去清真寺禮拜，我自認是虔誠的信徒。	-.719	.740	.860
	V191	我相信上帝。	-1.130	.930	.964
	V196	請問上帝對您生活中的重要程度，請以 1~10 表示之。	-.643	.838	.915
	V197	受訪者認為他們從宗教上獲得慰藉與力量。	-.650	.893	.945
社會容忍度	V68	我不介意與前科犯成為鄰居。	.066	.459	.677
	V69	我不介意與不同種族的人成為鄰居。	-1.031	.183	.428
	V70	我不介意與酗酒者成為鄰居。	.205	.439	.662
	V71	我不介意與情緒不穩定者成為鄰居。	-.270	.463	.680
	V73	我不介意與移民或外國勞工成為鄰居。	-.966	.240	.490

應用軟體：TESTFACT 4.0

政治學資料的模型設定方面，個體層次依變項採用的是「政治參與度」，其餘的變項則為個體層次自變項。各自變項和依變項的共變都假設了現代化理論的邏輯：凡是現代化所代表的特質，包括政治信賴感高、道德價值觀弱、宗教虔誠度低、社會容忍度高，都會有較高的政治參與度。在國家層次模型上，由於國內

生產毛額被定義為一個國家在一年內所生產產品和貨物的總值，可藉此窺探各國國力強弱，進而企圖解釋政治參與度和各自變項間的關係會因人均國內生產毛額而有波動，相同地，將各變項減去其均值，故納入模擬實驗的兩層次模型即為：

$$\begin{aligned}
 pp_j &= \beta_{0j} + \beta_{1j}(pc_j) + \beta_{2j}(mo_j) + \beta_{3j}(re_j) + \beta_{4j}(sd_j) + e_j, e_j \sim N(0, \psi) \\
 \beta_{0j} &= \gamma_{00} + \gamma_{01}(W_j - \bar{W}) + u_{0j}, u_{0j} \sim N(0, \tau_0) \\
 \beta_{1j} &= \gamma_{10} + \gamma_{11}(W_j - \bar{W}) + u_{1j}, u_{1j} \sim N(0, \tau_1) \\
 \beta_{2j} &= \gamma_{20} + \gamma_{21}(W_j - \bar{W}) + u_{2j}, u_{2j} \sim N(0, \tau_2) \\
 \beta_{3j} &= \gamma_{30} + \gamma_{31}(W_j - \bar{W}) + u_{3j}, u_{3j} \sim N(0, \tau_3) \\
 \beta_{4j} &= \gamma_{40} + \gamma_{41}(W_j - \bar{W}) + u_{4j}, u_{4j} \sim N(0, \tau_4) \quad [2.5]
 \end{aligned}$$

其中代號使用分別是 *pp* 政治參與度、*pc* 政治信賴感、*mo* 道德價值觀、*re* 宗教虔誠度、*sd* 社會容忍度。

第三節 研究限制

在過往的複層次迴歸模型研究裡，每當討論至樣本組合問題時，常常會涉及「調查成本」(cost)、「設計效果」(design effect)及「檢定力」(power)三項概念，毫無疑問地，三者與抽樣、多層次資料架構、參數估計等息息相關，因為我們之所以必須自母體中抽取具代表性的樣本進行推論是因為成本考量，為了節省成本採行多階段抽樣，加上研究對象的巢狀結構，迫使研究者必須考慮同一群體內個體間的相似程度，設計效果問題隨之發生，另外，當研究者在檢定研究假設時，錯誤地拒絕或錯誤地接受虛無假設都是犯了推論上嚴重的謬誤，因此檢定力是很

重要的一環。然而，本文礙於某些因素而未將上述概念納入模擬實驗，筆者將分別詳述如下。

一、調查成本

不可否認的，如果在總樣本數一定下，群體樣本數越大，通常調查成本也會隨之增高，其關係式如下：

$$Nc(n+t) \leq K \quad [2.6]$$

N 是群體樣本數， n 是每個群體內所抽取的個體樣本數， c 則為每抽一個個體層次樣本所必須花費的成本， t 指多一個群體樣本比多一個個體樣本在成本上是幾倍，且以 K 表示總預算的限制。由式子[2.6]來看，在一定的總預算限制下， N 與 n 勢必成反比，而若維持總樣本數一定， N 必然會有上限，特別在 c 特別大的情況下。

任何研究都侷限在資源有限的框架內，抽樣問題上的限制主要在於時間與經費，Snijders and Bosker(1993)的文章中便是利用數學推導，將樣本數帶入二層次線性迴歸模型參數估計式，再與成本公式[2.6]結合，試圖找出一定成本內的最適樣本配置，而 Cohen(1998)也是採用類似做法，將 Hansen(1953)所推導出用於二層次模型的成本公式帶入變異數估計式，以極小化變異數為目標而得最佳樣本數組合數學式。值得注意的是，兩位作者都非常仰賴二層次模型的參數估計式與成本公式，希望利用數學邏輯推演來得到最適樣本數組合，但其所用的參數估計式與成本公式皆以具漸近性的大樣本為前提而成立，與本文想要研究的小樣本參數特性有所出入，因此筆者不以公式推演，而用模擬實驗法企圖歸納出最適樣本組合的一般性原則。

況且，既然是模擬研究，本文希望盡可能地測試所有 (N, n) 配對，所以不對 c 與 t 設上限，這麼一來放寬成本限制，僅控制樣本數組合此一變因，使迴歸參

數表現的評估更單純化，才能夠徹底地分析樣本數配置，避免多重操縱下的問題，當日後抽樣設計者遭遇到成本限制時，便可奠基在一般性原則下做調整，因此暫不討論成本問題。

二、設計效果

多層模型的出現是因為研究者所關心的現象具有層級性結構，同一個群體中的個體彼此之間存在著相似性，這個事實違反了受訪者必須獨立單一的前提，故不適用單層線性模型，然而，個體之間的相似性我們可以用組內相關係數 (intraclass correlation, ICC) 來表示，常以 ρ 代表，組內相關係數是一比例數，說明總體層次能夠解釋總變異的量有多少。統計學界對於組內相關係數的討論不在少數，故發展出眾多根據不同模型而來的估算公式，Shrout and Fleiss(1979) 整理了各個組內相關係數公式，相互比較，提供研究者參考。

我們設想一個簡單的二層次迴歸模型，不論是個體層次或總體層次都不放入任何自變項，以數學式表示為

$$Y = \beta_0 + e, \quad e \sim (0, \sigma_e^2) \quad [2.7]$$

$$\beta_0 = \gamma_{00} + u_0, \quad u_0 \sim (0, \sigma_{u_0}^2) \quad [2.8]$$

將[2.7]、[2.8]合併成

$$Y = \gamma_{00} + u_0 + e \quad [2.9]$$

從[2.9]得知，這個模型並沒有任何自變項來解釋依變項 Y 的變異，故依變項的變異於是分成了兩個部分：個體層次誤差 σ_e^2 及總體層次誤差 $\sigma_{u_0}^2$ ，就此模型而言，組內相關係數便被定義為

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} \quad [2.10]$$

我們可自這個簡單的模型中理解組內相關係數的意含，假設總體層次自變項能完全解釋迴歸係數 β_0 ，總體層次變異為零，此時總變異完全來自個體層次，相關係數隨之較小，推論在同一群體內的兩個個體彼此之間的差異性大，反之，當個體層次自變項完全解釋依變項，總變異便完全來自總體層次，組內相關係數大，故同一群體內的兩個個體彼此之間相似性高。

通常教育學者會強調「巢狀資料」的產生是為了減少抽樣成本的緣故，所以如果組內相關係數越高，個體彼此之間有越高的相似性時，代表設計效果(design effect)越強，實質樣本數就越少，三者之間的關係如下

$$Nn_{\text{effective}} = \frac{Nn}{\text{design effect}}, \text{ design effect} = 1 + (n-1)\rho \quad [2.11]$$

$Nn_{\text{effective}}$ 是實質總樣本數， ρ 是組內相關係數，設計效果指的就是 $1 + (n-1)\rho$ 的部分，由[2.11]可知，當個體樣本數或組內相關係數變大，實質樣本數就會降低，若我們能知道或合理地猜想組內相關係數，那麼兩層次樣本數及有效總樣本數關係便能得知，但是一旦對研究對象毫無資訊或錯估組內相關係數時，有效樣本數的推測很可能產生偏差。

Muthen and Satorra(1995)曾認為組內相關係數的大小會影響參數估計的準確性(參見 Hox and Maas 2001)，而 Mass and Hox(2005)將組內相關係數納入模擬研究條件之一，把組內相關係數分為 0.1、0.2、0.3 三個層級，與樣本數組合同為控制變因，但其研究結論指出組內相關係數所造成的差異可說是微乎其微，加上本模擬研究的資料來源是以既有的調查訪問為基礎，故無法控制事前檢定力、組內相關係數等抽樣條件，僅能由事後分析的結果觀察組內相關係數和參數估計之間的關連性。

三、檢定力

一般說來，統計檢定主要在控制虛無假設為真(H_0)卻錯誤地拒絕虛無假設的風險，稱為 Type I 誤差(α)，然而 Cohen(1988;1992)極力主張，當我們在對於迴歸係數進行檢定時，除了應該注意 Type I 誤差之外，也應該注意 Type II 誤差(β)的問題，統計檢定力的概念便是由 Type II 誤差而起。

以檢定力來說， $1-\beta$ 代表在對立假設為真的前提下拒絕虛無假設的機率，換句話說，迴歸係數不顯著時，有可能是因為檢定力不足，而不代表此迴歸係數真正沒有影響力，Snijders and Bosker (1999:142)提供一個通用的檢定力公式，如下所示

$$\frac{\gamma_{es}}{s.e.(\hat{\gamma})} \approx (z_{1-\alpha} + z_{1-\beta})$$

移項後得
$$z_{1-\beta} = \frac{\gamma_{1 \cdot es}}{s.e.(\hat{\gamma}_1)} - z_{1-\alpha} \quad [2.12]$$

γ_{es} 是有效程度(effect size)， $s.e.(\hat{\gamma})$ 為 γ 的標準誤，由[2.12]中得知，只要知道數學式裡的其中三個數值，便能求得第四個數值。有效程度意指一個現象或效果存在的程度，也可理解為「估計虛無假設錯誤的程度」，在 Z 檢定中，以虛無假設及對立假設平均值的相對差之絕對值來作為指標，由於平均值是有單位的，故要除以標準差給予標準化。這裡我們一樣舉簡單的例子加以說明，比方說有一模型式為

$$Y = \gamma_0 + u \quad [2.13]$$

γ_0 是常數項， u 為誤差項，倘若我們覺得誤差項的變異 σ_u^2 可以進一步被自變項 x_1 解釋，所以[2.13]的變異為

$$V(Y) = V(\gamma_0 + \gamma_1 x_1 + u_1)$$

經過化簡可得
$$V(Y) = \gamma_1^2 V(x_1) + V(u_1) = \sigma_u^2$$

假設 x_1 可以解釋10%的 σ_u^2 ，於是假設檢定便可寫成

$$H_0 : \gamma_1 = 0$$

$$H_1 : \gamma_1 = 0.1\sigma^2$$

按照上述有效程度定義，此檢定的有效程度即為

$$\gamma_{1\cdot es} = \sqrt{\frac{0.1\sigma_u^2}{V(x_1)}} \quad [2.14]$$

將[2.14]帶入[2.12]就可以知道 $1-\beta$ 大小。

通常在事後分析上，我們早已決定 α 的水準，會知道 $z_{1-\alpha}$ 和估計值 γ_{es} ，然後由 $s.e.(\hat{\gamma}_1)$ 來算出檢定力好不好，Cohen(1988, 1992)建議以0.8為標準(相當於 $\beta=0.2$)。就研究設計的角度而言，於實際進行抽樣前先預設檢定力水準，然後算出 $s.e.(\hat{\gamma}_1)$ 應該是多少，進而求出合理的樣本數大小，不失為一個決定最適樣本的好方法，所以有許多學者把檢定力概念帶入多層模型樣本數配置的探討(Raudenbush 1997)。

可是，我們所關切的是各個迴歸參數估計的準確和穩定，並不是檢定力的強弱，因為通常在進行事前的檢定力評估時，我們必須對於有效程度的大小做出定義(即 γ_{es})，但任何的定義都相當武斷，況且，本文的意圖也不在測試任何一個自變項在實際的理論意義上的顯著與否，所以檢定力的考量就不在思考之列，重點仍在對於迴歸參數估計的準確性和穩定性上。

