

多階層取樣單位異質性 對科學調查之母體推論研究

蔡良庭* 楊志堅**

摘要

本研究以 Bootstrap 重複取樣程序，評估分層隨機取樣 (stratified random sampling) 及規模等比率取樣 (probability proportional to size)，對於調查研究的母體推論影響。調查研究的資料常包含多層次的結構，而分析必須搭配適當的取樣權重，才能正確推論母體特性 (蔡良庭、楊志堅，2008)。權重的計算又隨不同取樣設計而計算方式不同，但此權重計算的差異對於母體特性的推論，卻最常被忽略，也需更進一步驗證。本研究以數值模擬方法並搭配臺灣社會變遷調查的實徵資料分析，探討不同權重的計算對於母體推論的影響。結果顯示，若忽略了取樣單位間的樣本數差異，採用分層隨機取樣設計時，取樣數愈多，母群體特性的推論愈不精確。規模等比率的取樣設計及權重計算能提供更精準的母體推論。

關鍵字：取樣設計、取樣權重、確認性因素分析、分層隨機取樣、規模等比率取樣

-
- * 國立臺中教育大學教育測驗統計研究所認知神經計量實驗室博士後研究員。
 - ** 國立臺中教育大學教育測驗統計研究所認知神經計量實驗室教授，住址：403 臺中市西區民生路 140 號 教育測驗統計研究所，電話：04-22183523；傳真：04-22183520，電子郵件：Noahyang@ntcu.edu.tw。

備註：2012 年 8 月 17 日收稿；2013 年 1 月 30 日接受。

Heterogeneity of Multi-level Primary Sampling Units on Population Inference in Scientific Research

Liang-Ting Tsai,* Chih-Chien Yang**

ABSTRACT

In this study, the Str. RS (stratified random sampling) and PPS (probability proportional to size) sampling procedures were used to evaluate the heterogeneity of multi-level primary sampling units on population inference in social science research. When a social science survey has complexly designed sampling frames, the Str. RS and PPS are often applied. However, the effects of different sampling weights within these two methods on the inference of population characteristics are often neglected. A numerical simulation study and a real data analysis on Taiwan Social Change Survey data with a further extension confirmatory factor analysis model based on Yang and Tsai (2008) were proposed in this study. Independent variables manipulated in this study include the sampling designs, data type (continuous or categorical), sampling size, and heterogeneity of PSU. The results suggest the PPS sampling design can provide a more precise parameter estimate of CFA models in a complexly designed sampling frame survey, whether the data type is continuous or categorical. We summarize the findings and recommend the PPS multistage procedure based on the bootstrap method that can be used in practical social science survey applications.

Keywords: sampling design, sampling weight, confirmatory factor analysis, stratified random sampling, probability proportional to size

-
- * Postdoctoral Fellow, Cognitive NeuroMetrics Laboratory, Graduate Institute of Educational Measurement & Statistics, National Taichung University of Education, Taichung
 - * * Professor, PI, Cognitive NeuroMetrics Laboratory, Graduate Institute of Educational Measurement & Statistics, National Taichung University of Education

壹、緒論

科學的調查研究資料因為複雜取樣 (complex sampling) 或分層取樣 (stratified sampling)，而常具有多階層的資料結構，在這情形下，多階層模式 (multi-level model) 或階層線性模式 (hierarchical linear modeling) 是應用於分析此類資料非常普遍的方法 (周玉慧、朱瑞玲，2008；熊瑞梅、張峰彬、林亞鋒，2010)。進行多階層結構資料分析時，必須搭配適當的取樣權重 (sampling weights)，才能正確推論母體的特性 (Kaplan and Ferguson, 1999；楊志堅、蔡良庭，2008；蔡良庭、楊志堅，2008)。不同的取樣設計伴隨不同的權重計算方式，但此不同權重計算對於母體特性的推論影響，卻最常為研究者所忽略，也需更進一步驗證。尤其當科學調查研究包含多階層的複雜資料結構時，所涉及的多階層取樣調查就相對的重要，而分層隨機取樣 (stratified random sampling, Str. RS) 及規模等比率取樣 (probability proportional to size, PPS) 設計與推論母體總數之間的關係，就顯得格外重要。因此本研究延伸蔡良庭與楊志堅 (2008) 的確認性因素分析 (confirmatory factor analysis, CFA) 為母體模式，以 Bootstrap 重複取樣 (resampling) 程序，評估不同權重計算方式的 Str. RS 及 PPS 取樣設計，對於推論母體特徵的影響，並以臺灣社會變遷調查第五期第一波資料 (張荳雲、廖培珊主編，2008) 為例子，說明不同取樣設計對於科學調查研究之母體推論影響。

能否精確的推論母體特性，是科學調查研究必須具備的條件之一 (Lohr, 2010; Stapleton, 2008; Wolter, 2007)。推論精確與否，除了觀察變項統計量的估計之外，觀察變項統計量之標準誤 (standard error)

的估計也是探討的重要指標 (Lee and Forthofer, 2006; Lohr, 2010; Rust and Rao, 1996; Stapleton, 2008; Wolter, 2007)。影響上述兩個指標能否正確推論母體特性的原因，例如：取樣設計的選擇、權重計算、取樣數多寡及其代表性、乃至最後資料分析的統計模式及參數估算法等，都是重要因素。尤其在取樣權重的應用，相關研究 (例如：Kaplan and Ferguson, 1999；楊志堅、蔡良庭，2008；蔡良庭、楊志堅，2008) 指出，進行科學調查的資料分析時，須搭配使用「適當」的取樣權重，才能正確推論母體的特性。

忽略取樣權重將導致母體推論不精確 (Kaplan and Ferguson, 1999；楊志堅、蔡良庭，2008；蔡良庭、楊志堅，2008)。蔡良庭與楊志堅 (2008) 指出，進行 CFA 模式分析時，即使增加取樣樣本數，相關情形並未獲得改善。同理，分層層別間變異性愈大，推論愈不準確；若考慮權重時，則分層層別間變異性對於估計準確性影響並不大。楊志堅與蔡良庭 (2008) 將權重概念應用於 Likert 問卷的類別型資料測量恆等性 (measurement equivalence/invariance, ME/I) 檢定，發現權重對於檢定效果扮演相當重要的影響，例如：考慮權重的情況下，檢定正確性隨取樣數增加而上升，且優於未考慮權重的檢定效果。另外，Kaplan and Ferguson (1999) 的研究結果亦顯示：若忽略權重的計算，將導致 CFA 模式的參數估計產生偏誤 (bias)，且隨分層變異性愈大而偏誤增大。楊志堅等人 (2009)、Asparouhov (2005; 2006)、Grilli and Pratesi (2004) 及 Stapleton (2006) 等研究，陸續將權重的概念應用於多層次模式、潛在類別分析 (latent class analysis) 等模式，也有類似的研究發現。然上述研究多僅針對單一階層及單一權重計算的取樣設計進行分析，並未針對科學調查研究中，常遇到的多階層複雜取樣設計 (complex sampling designs) 資料進行探討。且 Lohr (2010)

指出，實際進行調查研究的工作者，在面對多階層複雜資料結構時，常在多階層取樣程序的複雜性或忍受採用單一階層的隨機取樣所產生的母體推論偏誤間進行取舍。因此，多階層資料的不同取樣設計對推論母體特性的正確性，則需更進一步的驗證。

具有多階層架構的大型或跨國際科學調查研究是很普遍也很重要的，例如：TIMSS (trends in international mathematics and science study)、PISA (programme for international student assessment)、PIRLS (progress in international reading literacy study) 等，常包含多個國家、地區或是學區；國內的調查研究，例如：臺灣青少年成長歷程研究 (Taiwan youth project, TYP)、臺灣地區社會變遷基本調查 (Taiwan social change surveys, TSCS)、臺灣教育長期追蹤資料庫研究 (Taiwan educational panel study, TEPS) 等，亦包含多個縣市，縣市內又包含鄉鎮或多個學區的複雜取樣設計結構。為了使取樣樣本更有效代表階層在整個母體所扮演的角色，常以不同的取樣設計進行樣本抽取，例如：以分層隨機取樣或規模等比率取樣等多階段的設計取代簡單隨機抽樣 (simple random sampling, SRS)。

多階層取樣設計的推論母體總數易受基本取樣單位 (primary sampling units, PSU) 內不等比率的樣本數影響，而導致錯誤的研究結果推論。多階層的取樣方法中，取樣權重隨取樣設計有不同計算方式 (Scheaffer et al., 2006)，例如：多階段的複雜取樣設計中，基本取樣單位內的樣本總數常不相同，以 Str. RS 進行第一階段取樣單位取樣，則最後母體中的每一個樣本被抽取的機率並不相同，這也是因 Str. RS 強調取樣單位的變異性，而不強調取樣單位的樣本數與母體的等比率，因此可能產生不等比率的樣本數，也導致實際的實驗設計時，加入取樣權重的推論樣本數結果就不必然與母體的總數相同。同理，若以 PPS

進行分層資料抽取，不僅強調取樣單位的樣本數與母體的等比率，且母體中每個樣本被抽取機率相同，最後加入取樣權重的樣本數推論總和與母體的總數相同。因此在進行資料分析時，極有可能因忽略樣本數與母體的等比率而導致即使考慮了取樣權重的重要性，亦產生錯誤的研究結果推論，而此部分卻也最常為研究者所忽略，且目前有關取樣權重的相關文獻也較少關注於不同分層取樣設計所造成的差異，對母體推論正確性的實徵數據，導致進行實際科學調查研究時，常因沒有明確的依據可遵循，僅憑經驗法則或以最方便取樣的方式進行資料蒐集。

綜合上述，本研究以 CFA 模型為架構，採數值模擬的實驗方式進行並以實徵調查資料分析輔以說明，模擬實驗設計中包含多種不同的取樣數、取樣單位變異性程度及資料類型，並運用 Bootstrap 重複取樣程序，探討 Str. RS 及 PPS 兩種取樣設計在 CFA 模式中，推論母體參數估算的正確性，以提供未來科學調查研究在進行取樣設計時能有更明確的參考依據。進行複雜資料結構的分析時，是否考慮取樣單位與母體的等比率對於母體特性推論的影響，以做實徵科學調查研究參考有其必要性。尤其 CFA 模式常為調查研究者應用來進行量表編製的項目分析與效度評量的建構，因此分析結果的精確性在量表的理論發展與測量上極為重要，顯見其更有探討必要性。本文的安排：除了首先介紹研究動機及目的外，以下依序為調查研究的參數及參數估計標準誤估算方法、Bootstrap 在不同取樣設計的執行程序、實際的研究進行步驟與研究結果，最後並以臺灣社會變遷調查第五期第一波資料為例子，說明不同取樣設計對於參數估計的影響，並驗證電腦模擬研究之結果，最後為本研究的結論及建議。

貳、調查研究的參數及參數估計標準誤估算方法

衡量調查研究取樣資料的觀察變項測量值的參數估計精確與否，除了觀察變項統計量的估計之外，觀察變項估計的標準誤也是探討的重要指標。目前可應用於估算參數估計標準誤的有泰勒線性法 (Taylor linearization method, TLM) 的非重複取樣 (non-resampling) 方法及 jack-knife repeated replication (JRR)、Bootstrap 等重複取樣方法兩類。對於非線性的統計分析，例如：參數比率 (ratio)、迴歸參數 (regression parameter) 或是相關係數 (correlation coefficients) 等的參數變異數估計，除了 TLM 估計方法之外，JRR、Bootstrap 等重複抽樣方法也已被廣泛使用 (Babu and Singh, 1983; Kovačević et al., 2006; Sahinler and Topuz, 2007; Wu, 1986)。雖然 TLM 法適用於一般性的取樣設計，但對於不同的參數估計的標準誤估算必須採用不同的公式，相對於重複取樣方法只需利用單一的計算公式，即可針對不同的統計量進行估計標準誤計算，顯然 TLM 並不是一個便利的方法。此外，重複取樣程序的優點是無須特別對母體作預先性的分配假設，因而取代了統計理論的分析及繁複的推論。同時，根據 Andersson et al. (1987)、Kovačević et al. (2006) 的研究指出 JRR、Bootstrap 等重複取樣方法比 TLM 的非重複取樣方法在多階層模式分析時，能提供更精確的參數估計。

根據 Stapleton (2008) 應用重複取樣程序於 SEM (structural equation modeling) 的結構模式 (structure model) 參數估計，雖然 Bootstrap 取樣的參數估計其標準誤估計之相對偏誤 (relative bias) 略高於 JRR 的研究結果。但在實質的應用上，JRR 受限於分層中的取樣單位個數；個數較少時，則重複取樣次數也相對的變少，如此便可能影響參數估

計的準確性。相對的，Bootstrap 取樣程序對於科學調查資料的分析而言，是具有彈性且有效的方法（Chernick, 2007; Davison and Hinkley, 1998; Efron and Tibshirani, 1993; Lahiri, 2003; Stapleton, 2008），並可由研究者自行決定取樣次數。同時，許多研究（Davison et al., 2003; Kaufman, 1998; Lahiri, 2003; Rust and Rao, 1996; Stapleton, 2008）證實，複雜取樣設計的科學調查資料庫應用中，Bootstrap 程序可應用在多階段取樣中的任一階段。然而，以 Bootstrap 進行取樣時，要抽取多少個取樣單位或是樣本來代表整個分層的資料，才能得到最佳的母體特性推論，一直是研究者關心的議題。根據 Rust and Rao（1996）及 Stapleton（2008）建議，以分層中抽取 $(J-1)$ 個取樣單位來代表 J 個取樣單位的分層資料或是抽取 $(I-1)$ 個樣本取代 I 個樣本的取樣單位，都是一個有效的取樣方式。

參、Bootstrap 在不同取樣設計的執程序

在具有分層結構特性的複雜、大規模科學調查研究，多以兩階段或多階段的取樣設計取代簡單隨機取樣。本研究採用的分層隨機取樣及規模等比率取樣即屬於多階段取樣中的一種。研究中先抽取大的調查單位（例如：PSU），後續再從大單位中抽小單位，如包含有更小單位則再進行更小單位的抽取。以下簡單說明兩種取樣設計的方法，後續以更詳細的實例說明 Bootstrap 重複取樣程序在這兩種不同權重計算的取樣設計的差異。

分層隨機取樣先將母群體資料依某些相異特徵或研究需要分為不重複的組別，也就是分層（蔡良庭、楊志堅，2008），然後於每分層中再進行獨立的簡單隨機取樣，此方法稱為分層隨機取樣。不但可以更

經濟且提高估計的精準度，而分層時又以同層之內同質性為最大，異層之間同質性最小為原則，使層內的資料愈一致，估計的標準差愈小，則取樣誤差也愈小（Scheaffer et al., 2006；蔡良庭、楊志堅，2008），然而分層隨機取樣強調分層的變異性，而常忽略分層的樣本數與母體的等比率，因而可能產生取樣的不等比率樣本數，而導致加權的樣本數結果不必然等同於母體數。規模等比率取樣的目的是使母體中的個體不論屬於哪一分層，最終被抽取的機率相同。即修改分層隨機取樣中的簡單隨機取樣為每一階段各單位的抽中機率與該單位內的樣本數大小成比例。經由此程序取樣後，母體的每一個樣本最後都具有相等的被抽中機率，經由加權程序後的樣本數結果等同於母體總數。

Bootstrap 取樣步驟是在相同的資料集進行重複取樣，每次取樣後必須調整資料的取樣權重以反應分層內的樣本數。以二階段取樣且抽取 $(J-1)$ 個取樣單位為例，其中 J 為取樣單位總數，抽取樣本的取樣權重計算如公式 (1)：

$$w_{hji}(r) = w_{hji} \times \frac{n_h}{n_h - 1} m_j(r) \quad (1)$$

其中， w_{hji} 為第 h 個分層第 j 個取樣單位內樣本 i 的原始權重， $w_{hji}(r)$ 則為第 r 次 Bootstrap 取樣的樣本權重， n_h 為第 h 個分層中取樣單位的個數， $m_j(r)$ 代表第 r 次 Bootstrap 取樣中，第 j 個取樣單位被隨機抽中的次數。舉例來說：若有 2 個分層 (S_1 及 S_2)，每分層中各包含 4 個取樣單位（分別為 $P_{11}, P_{12}, P_{13}, P_{14}, P_{21}, P_{22}, P_{23}, P_{24}$ ），假設 2 個分層中各有一個取樣單位 (P_{11} 及 P_{21}) 樣本數為 60，其餘為 30 個樣本，則母體總數為 300。若第一階段從每分層中抽取 2 個取樣單位；第二階段再從已抽取的 4 個取樣單位中各隨機抽取 3 個樣本，共計抽取 12

個樣本進行資料分析。

第一步驟：分層取樣中，假設第一階段採用 Str. RS 分別從 S_1 中隨機抽中 P_{12} 及 P_{13} ， S_2 中隨機抽中 P_{23} 及 P_{24} ，第二階段再從已抽取的 4 個取樣單位中各隨機抽取 3 個樣本。每個抽取樣本依取樣機率倒數計算樣本的原始權重（例如： P_{12} 中抽取的 3 個樣本，原始權重 $w_{hji}(r) = 1/\frac{2}{4} \times \frac{3}{30} = 20$ ）。詳細結果如表 1 中 W_R 一欄所示。

第二步驟：在分層 S_1 中以 Bootstrap 方法反覆隨機抽取 (2-1) 個取樣單位代表分層 S_1 中的樣本。同理，分層 S_2 重複上述程序。因每分層中隨機抽取 1 個取樣單位代表分層資料，則抽中的取樣單位內樣本權重必須將原始權重乘以 2 倍 ($w_{hji} \times \frac{2}{2-1} \times 1 = 2w_{hji}$)，其餘樣本權重則為 0。表 1 中， W_{Boot1} 代表第 1 次 Bootstrap 重複取樣的權重計算結果， $W_{Boot2} \sim W_{Boot6}$ 則依此類推為第 2 至第 6 次的重複取樣計算權重。

表 1 Str. RS 設計之 Bootstrap 重複取樣權重計算實例

stratum	PSU	subject	W_R	W_{Boot1}	W_{Boot2}	W_{Boot3}	W_{Boot4}	W_{Boot5}	W_{Boot6}
S_1	P_{12}	1	20	40	40	0	0	0	40
S_1	P_{12}	2	20	40	40	0	0	0	40
S_1	P_{12}	3	20	40	40	0	0	0	40
S_1	P_{13}	4	20	0	0	40	40	40	0
S_1	P_{13}	5	20	0	0	40	40	40	0
S_1	P_{13}	6	20	0	0	40	40	40	0
S_2	P_{23}	7	20	0	40	40	0	40	40
S_2	P_{23}	8	20	0	40	40	0	40	40
S_2	P_{23}	9	20	0	40	40	0	40	40
S_2	P_{24}	10	20	40	0	0	40	0	0
S_2	P_{24}	11	20	40	0	0	40	0	0
S_2	P_{24}	12	20	40	0	0	40	0	0

從表 1 中發現，取樣的 12 個樣本不論是原始權重或是 Bootstrap 重複取樣權重總和皆為 240，與取樣樣本的權重總和理應等於母體總數 300 並不相符。

從上述實例，若第一步驟的分層取樣忽略取樣單位內的樣本數差異，則原始權重及 Bootstrap 重複取樣權重無法真實反應母體總數。而 PPS 取樣設計將分層取樣的第一階段修改為分層不相等機率取樣，第二階段仍維持簡單隨機取樣，使母體中的每個樣本被抽中的機率相等，不受取樣單位內樣本數不相同影響。延續上述例子，分層 S_1 中的取樣單位樣本數分別為 60、30、30、30。因此第一階段取樣中，以每個取樣單位中的樣本數除以 S_1 的總樣本數為取樣機率，可由公式 (2) 表示：

$$\pi_{hj} = \frac{n_{hj}}{\sum_{j=1}^J n_{hj}} \quad (2)$$

其中 π_{hj} 為第 h 個分層中第 j 個取樣單位被抽中的機率， n_{hj} 代表第 h 個分層中第 j 個取樣單位的樣本數。因此分層 S_1 中四個取樣單位被抽中的機率分別為 0.4、0.2、0.2 及 0.2，而第二階段以簡單隨機取樣方式進行取樣單位內的樣本抽取。因第一階段的取樣已修改為不相等機率取樣，則原始權重的計算必須進行修正。假設 $\pi_{i|hj}$ 代表第 h 個分層中第 j 個取樣單位內樣本 i 的抽中機率，且 K_h 為分層 h 中取樣的取樣單位個數，則分層不相等機率取樣的原始樣本權重可修正為：

$$w_{hji} = \frac{1}{\pi_{hj} \times \pi_{i|hj} \times K_h} \quad (3)$$

分層不相等機率取樣的原始權重計算結果如表 2 所示，例如： P_{12} 中抽取的 3 個樣本，原始權重 $w_{hji} = 1 / \frac{30}{150} \times \frac{3}{30} \times 2 = 25$ 。依據原始原權重，則 Bootstrap 重複取樣權重計算結果如表 2 中 $W_{Boot1} \sim W_{Boot6}$ 所示。從表 2 發現，不論是原始權重或是重複取樣的權重總和都與母體總數相同。

表 2 PPS 設計之 Bootstrap 重複取樣權重計算實例

stratum	PSU	subject	W_R	W_{Boot1}	W_{Boot2}	W_{Boot3}	W_{Boot4}	W_{Boot5}	W_{Boot6}
S_1	P_{12}	1	25	50	50	0	0	0	50
S_1	P_{12}	2	25	50	50	0	0	0	50
S_1	P_{12}	3	25	50	50	0	0	0	50
S_1	P_{13}	4	25	0	0	50	50	50	0
S_1	P_{13}	5	25	0	0	50	50	50	0
S_1	P_{13}	6	25	0	0	50	50	50	0
S_2	P_{23}	7	25	0	50	50	0	50	50
S_2	P_{23}	8	25	0	50	50	0	50	50
S_2	P_{23}	9	25	0	50	50	0	50	50
S_2	P_{24}	10	25	50	0	0	50	0	0
S_2	P_{24}	11	25	50	0	0	50	0	0
S_2	P_{24}	12	25	50	0	0	50	0	0

Bootstrap 的重複取樣次數可由研究者自行決定，假設重複取樣 R 次，並經由 R 次模式分析後，其參數估計的標準誤可由公式 (4) 計算得到 (Stapleton, 2008; Wolter, 2007)：

$$SE_{\hat{\theta}_{boot}} = \sqrt{\frac{\sum_{r=1}^R (\hat{\theta}_r - \hat{\theta})^2}{R-1}} \quad (4)$$

其中 $\hat{\theta}_r$ 為第 r 次重複取樣的參數估計值， $\hat{\theta}$ 則為原始取樣資料參數估計值， R 為重複取樣次數。

肆、研究設計及執行步驟

本研究主要探討不同權重的取樣設計對於科學調查研究的母體推論正確性影響。為使實驗符合調查實徵資料分析情境，研究中，以 CFA 為母體資料分析模式，在不同的實驗條件，再依不同取樣數進行樣本抽取，最後以 QPML (quasi pseudo maximum likelihood) (Asparouhov, 2005; Asparouhov et al., 2004) 估算法進行模式的參數估算。

研究中設計母體資料為包含二階層結構的複雜取樣設計。母體包含有 6 個分層，每分層內含有 8 個取樣單位 ($P_{h1} \sim P_{h8}$)，為凸顯 Str. RS 及 PPS 兩種取樣設計的權重計算差異，其中 $P_{h1} \sim P_{h2}$ 的取樣單位樣本分別為相對較少的 600 及 800； $P_{h3} \sim P_{h6}$ 屬於大樣本，分別為 3,500、4,000、4,000 及 3,500； $P_{h7} \sim P_{h8}$ 亦屬小樣本人數，分別為 800 及 600，合計每分層中包含 17,800 筆樣本，而母體總計有 106,800 筆資料。

本研究的 CFA 模式，設計五個觀測變項 (observed variable) 及單一潛在變項 (latent variable) 的測量模式 (measurement model)。其架構如下：

$$y_i = \lambda_i \eta + \varepsilon_i \quad (5)$$

其中 i 為觀察變項個數， η 為單一潛在變項， λ_i 為因素負荷量， ε_i 為測量誤差 (measurement errors)。研究中每一取樣單位的資料都是設定五個觀測變項 (y_1, \dots, y_5) 的連續資料，當觀測變項為類別型資料時，設定為五等第計分之 Likert 問卷資料，因此所有觀測變項都有五個類別，

並由四個閾值 (threshold) 將所產生的五個觀測變項的連續資料進行區隔, 閾值之設定參照 Kim and Cohen (1998) 的實徵研究 (詳如表 3)。另, 測量誤差不論於連續資料或類別資料都是設定為平均數 0, 標準差為 1 的標準常態分配。

表 3 類別型資料之觀察變項閾值參數設定值

觀測變項	閾值 (threshold)			
	τ_1	τ_2	τ_3	τ_4
y_1	-0.35	0.67	0.97	1.94
y_2	0.18	0.90	1.29	1.94
y_3	-0.37	0.03	0.91	2.29
y_4	-0.56	-0.13	0.80	2.22
y_5	-0.38	0.49	1.04	2.33

依據上述, 研究模式設定如圖 1 所示。為探討取樣單位變異對於母體推論的影響, 本研究以 CFA 模式中, 各取樣單位的因素負荷量差異定義為取樣單位變異, 實驗設計時, 僅固定取樣單位在第二個因素負荷量具有差異, 模式中其餘所有取樣單位的第一個因素負荷量設定為 1 ($\lambda_1=1$), 第三至第五設定為 0.5 ($\lambda_3=\lambda_4=\lambda_5=0.5$)。研究設計有四種取樣單位變異, 以 $P_{h1}\sim P_{h2}$ 、 $P_{h3}\sim P_{h6}$ 及 $P_{h7}\sim P_{h8}$ 的第二個因素負荷量分別為 0.1、0.5 及 0.9 為變異最大, 本研究中以 d 代表不同因素負荷量差距, 此設計記錄為 $d=0.4$, 另三種設計分別為 (0.2、0.5、0.8)、(0.3、0.5、0.7) 及 (0.4、0.5、0.6), 記錄為 $d=0.3$ 、 $d=0.2$ 及 $d=0.1$ 。根據研究 (Kaplan and Ferguson, 1999; 蔡良庭、楊志堅, 2008) 指出, CFA 模式的分析中, 若不同分層中的觀測變項因素負荷量相同, 則參數估計的精確度並不因各分層中的權重不相同而受影響, 因此本研究

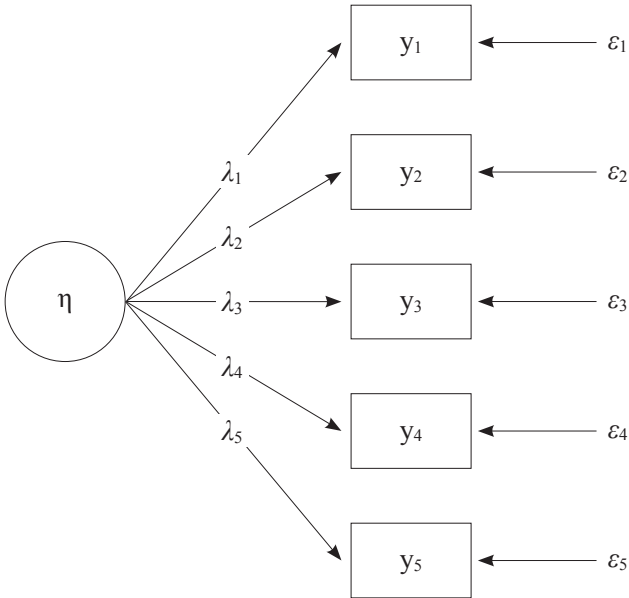


圖 1 潛在變項模式路徑圖

將不探討第三至第五個因素負荷量的估計精準度。

上述設定，可利用 AMOS (Arbuckle, 2006)、Mplus (Muthén and Muthén, 1998-2011)、LISREL (Jöreskog and Sörbom, 1996) 等相關軟體協助進行資料產生。後續實驗中的取樣及權重計算則依以下二個步驟進行：

步驟一：分層取樣的第一階段採用 Str. RS 及 PPS 兩種取樣設計於每一分層抽取 4 個取樣單位，第二階段再從抽取的 4 個取樣單位中，分別隨機抽取樣本數 10、50、100 及 200。因此當取樣單位中隨機抽 10 個樣本數時，總取樣數為 $10 \times 4 \times 6 = 240$ 。其餘取樣數分別為 1200、2400 及 4800。取樣後依據不同取樣設計計算樣本原始權重。

步驟二：利用 Bootstrap 取樣程序，於每一分層中以取後放回的方式，一次隨機抽取 1 個取樣單位，反覆抽取 3 次共計抽取 3 個取樣單位，抽取的取樣單位則利用原始權重及公式 (1) 計算重複取樣權重。此外，在每一次的分層取樣中重複 100 次 Bootstrap 的重複取樣及權重計算。

依據上述的模式設計及取樣程序，本研究共計有 2(資料型態)×2(取樣設計)×4(取樣單位變異)×4(抽取樣本數)=64 種情境，每一種情境下重複 500 次 (replication) 的實驗，每一次的實驗都必須進行 100 次的 Bootstrap 取樣及模式估計，每一次的估算採用不同的 Bootstrap 樣本及其取樣權重，以探討模式的參數估算及參數估計標準誤的準確性。

經由取樣程序及模式估算後，依據公式 (6) 至公式 (9) 的定義計算參數估計值及參數估計標準誤的相對偏誤及 MSE (Stapleton, 2008; Wolter, 2007)，以衡量參數估計及估計標準誤的準確性。

$$B(\hat{\theta}_p) = \frac{\overline{\hat{\theta}_{p_i}} - \hat{\theta}_p}{\hat{\theta}_p} \times 100 \quad (6)$$

$$MSE(\hat{\theta}_p) = \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_{p_i} - \hat{\theta}_p)^2 \quad (7)$$

$$B(s\hat{e}_{\hat{\theta}_p}) = \frac{\overline{s\hat{e}_{\hat{\theta}_{p_i}}} - s\hat{e}_{\hat{\theta}_p}}{s\hat{e}_{\hat{\theta}_p}} \times 100 \quad (8)$$

$$MSE(s\hat{e}_{\hat{\theta}_p}) = \frac{1}{500} \sum_{i=1}^{500} (s\hat{e}_{\hat{\theta}_{p_i}} - s\hat{e}_{\hat{\theta}_p})^2 \quad (9)$$

其中 $\hat{\theta}_p$ 是第 p 個估計的母體參數值， $\overline{\hat{\theta}_{p_i}}$ 為 500 次重複實驗的第 p 個參數估計平均值， $s\hat{e}_{\hat{\theta}_p}$ 為母體參數值 $\hat{\theta}_p$ 的估計標準誤， $s\hat{e}_{\hat{\theta}_{p_i}}$ 代表 i 次取樣估計後的第 p 個參數估計標準誤。

伍、研究結果

本節主要將模擬實驗的結果以圖、表呈現整體的趨勢外，並依連續資料及類別資料分別說明本研究之結果。

一、連續資料

連續資料的設定下，以 CFA 模式估算後之因素負荷量估計值及估計標準誤相對偏誤作為比較基準，數值愈接近 0 表示估計正確性愈高，並輔以 MSE 補充說明。以下分別依取樣設計、取樣數及取樣單位變異等因子進行結果說明。實際的參數估計相對偏誤及 MSE 如表 4 所示。

參數估計部分，在 PPS 及 Str. RS 取樣設計之下，相對偏誤及 MSE

表 4 連續資料之參數估計相對偏誤及 MSE

		Difference of factor loading (d)							
		PPS				Str. RS			
n		0.4	0.3	0.2	0.1	0.4	0.3	0.2	0.1
Relative Bias (%)	240	0.20	0.12	0.33	0.05	0.09	-0.61	0.56	0.06
	1200	0.07	0.09	0.07	0.05	0.95	0.09	0.37	0.36
	2400	-0.06	0.08	0.03	0.00	-0.27	-0.28	0.39	-0.15
	4800	-0.06	0.03	0.02	-0.01	0.07	-0.34	0.39	-0.34
MSE	240	.0032	.0022	.0020	.0018	.0026	.0026	.0022	.0026
	1200	.0014	.0009	.0005	.0005	.0008	.0008	.0006	.0005
	2400	.0011	.0007	.0004	.0002	.0005	.0004	.0003	.0003
	4800	.0010	.0006	.0004	.0001	.0003	.0003	.0002	.0001

雖有不同但差異極小且都趨近於 0，顯示不同取樣設計對於參數估計影響極小。同理，在不同取樣數及取樣單位變異之下，PPS 取樣的參數估計相對偏誤約介於 -0.06% 至 0.33%，Str. RS 取樣的參數估計相對偏誤約介於 -0.61% 至 0.95%。依據 Hoogland and Boomsma (1998) 建議的參數估計其標準誤估計之相對偏誤絕對值應小於 5% 為標準，則兩種取樣設計的估計結果都在標準之內。MSE 則隨取樣數增加而降低，雖然樣本數 240 時，高於其他結果，但差距不大。

參數估計標準誤部分，數值模擬結果如表 5 所示。在不同的取樣數之下，以 PPS 進行取樣，則參數估計標準誤的相對偏誤介於 -6.60% 至 9.64% 之間，依據 Stapleton (2002) 及 Hoogland and Boomsma (1998) 建議的參數估計標準誤相對偏誤絕對值應小於 10% 為標準，則都在設定的標準之內，不因取樣數的變化而有特定趨勢。若以 Str. RS 為取樣

表 5 連續資料之參數估計標準誤相對偏誤及 MSE

		Difference of factor loading (<i>d</i>)							
		PPS				Str. RS			
<i>n</i>		0.4	0.3	0.2	0.1	0.4	0.3	0.2	0.1
Relative Bias (%)	240	-3.06	3.64	4.28	-1.43	-13.84	-0.16	8.70	10.79
	1200	2.23	3.53	9.64	-6.60	-38.59	-28.92	-19.11	0.37
	2400	8.75	1.89	1.54	5.45	-40.68	-39.62	-18.70	3.24
	4800	-2.47	6.11	-5.03	-3.68	-39.98	-35.77	-37.02	-9.47
MSE	240	.0001	.0001	.0001	.0001	.0005	.0002	.0002	.0002
	1200	.0001	.0000	.0000	.0000	.0005	.0002	.0001	.0000
	2400	.0001	.0000	.0000	.0000	.0003	.0002	.0000	.0000
	4800	.0000	.0000	.0000	.0000	.0003	.0002	.0001	.0000

設計，則參數估計標準誤相對偏誤明顯表現不佳，甚至達 40% 的低估。例如：取樣單位變異 0.4 時，隨取樣數從 240 增加至 4800，則相對偏誤分別為 -13.84% 、 -38.59% 、 -40.68% 及 -39.98% 。將標準誤相對偏誤繪製成圖 2，X 軸分別代表不同的取樣數及取樣設計，Y 軸為參數估計標準誤相對偏誤。從圖中可同時看出兩種取樣設計的相對偏誤變化情形，在 PPS 取樣設計，相對偏誤多介於正負 10% 之內，但 Str. RS 的情況下，大多隨取樣數增加而參數估計標準誤愈低估。

從圖 2 中同時可看出不同取樣單位變異之間的變化，圖中的每一折線代表不同的取樣單位變異。PPS 取樣設計時，相對偏誤在不同的取樣單位變異下並無固定趨勢，但在 Str. RS 取樣設計時，參數估計其標準誤估計之相對偏誤隨差距增加而愈嚴重低估。參數估計標準誤的 MSE 表現，從表 5 中可得知其 MSE 之間雖有差異，但差異並不大。

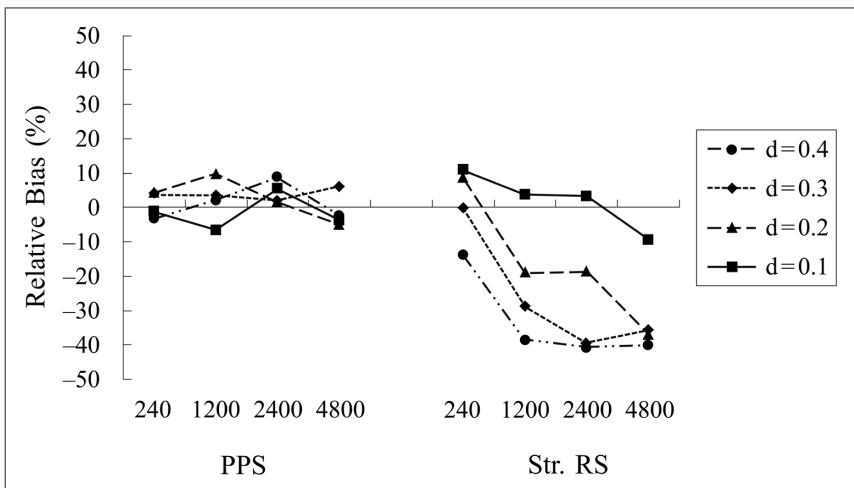


圖 2 連續資料之參數估計標準誤相對偏誤折線圖

二、類別資料

類別資料的設定下，參數估計值的相對偏誤及 MSE 如表 6 所示。參數估計部分，在 PPS 及 Str. RS 取樣設計之下，相對偏誤及 MSE 雖有不同但差異極小且都趨近於 0，PPS 取樣的參數估計相對偏誤約介於 -0.2% 至 0.05%，Str. RS 取樣的參數估計其標準誤估計之相對偏誤約介於 -1.32% 至 0.96%，其相對偏誤絕對值皆小於 5% 的標準，顯示不同取樣設計對於參數估計影響極小。

參數估計標準誤部分，數值模擬結果如表 7 所示。不同的取樣數之下，以 PPS 進行取樣，參數估計標準誤相對偏誤介於 -9.76% 至 8.61% 之間，絕對值小於 10% 的標準，且不因取樣數的增加或減少而有固定趨勢，顯示 PPS 的取樣設計，取樣樣本數對於參數估計標準誤

表 6 類別資料之參數估計相對偏誤及 MSE

		Difference of factor loading (d)							
		PPS				Str. RS			
n		0.4	0.3	0.2	0.1	0.4	0.3	0.2	0.1
Relative Bias (%)	240	-0.20	-0.01	0.00	0.04	0.13	0.93	-0.08	-1.32
	1200	-0.03	0.02	-0.03	0.01	0.66	0.96	0.69	-1.12
	2400	-0.03	-0.02	-0.02	0.03	0.69	0.65	0.29	-1.21
	4800	-0.01	0.00	0.05	-0.01	0.41	0.09	0.05	-1.26
MSE	240	.0038	.0037	.0033	.0034	.0047	.0042	.0044	.0038
	1200	.0014	.0009	.0008	.0006	.0010	.0011	.0008	.0008
	2400	.0011	.0008	.0005	.0003	.0007	.0004	.0005	.0004
	4800	.0008	.0005	.0004	.0002	.0003	.0004	.0002	.0002

表 7 類別資料之參數估計標準誤相對偏誤及 MSE

		Difference of factor loading (d)								
		n	PPS				Str. RS			
			0.4	0.3	0.2	0.1	0.4	0.3	0.2	0.1
Relative Bias (%)	240	-3.04	-6.82	1.72	-9.76	4.56	12.20	11.11	10.59	
	1200	-5.76	-3.36	0.29	2.96	-21.54	-18.02	-4.23	4.99	
	2400	-4.09	-6.97	6.29	3.75	-34.61	-24.84	-11.96	9.18	
	4800	3.51	8.61	-2.40	8.16	-35.26	-39.52	-16.73	-1.13	
MSE	240	.0002	.0002	.0001	.0002	.0004	.0004	.0004	.0006	
	1200	.0000	.0000	.0000	.0000	.0002	.0001	.0001	.0001	
	2400	.0000	.0000	.0000	.0000	.0002	.0001	.0000	.0000	
	4800	.0000	.0000	.0000	.0000	.0002	.0002	.0000	.0000	

相對偏誤影響不大。以 Str. RS 為取樣設計，則參數估計標準誤相對偏誤明顯的隨取樣數增加而愈表現不佳。例如：取樣單位變異 0.3 時，取樣數從 240 增加至 4800，相對偏誤分別為 12.20%、-18.02%、-24.84% 及 -39.52%，絕對值都超過設定的標準。將標準誤相對偏誤繪製成圖 3，X 軸分別代表不同的取樣數及取樣設計，Y 軸為參數估計標準誤相對偏誤。圖中可同時看出兩種取樣設計的相對偏誤變化情形，在 PPS 取樣設計，相對偏誤多介於正負 10% 之內，但 Str. RS 的情況下，隨取樣數增加而參數估計標準誤愈低估。

從圖 3 中同時可看出不同取樣單位變異之間的變化，圖中的每一折線代表不同的取樣單位變異。PPS 取樣設計時，相對偏誤在不同的取樣單位變異下並無固定趨勢，但 Str. RS 取樣設計時，除了取樣數 240 之外，參數估計其標準誤估計之相對偏誤隨差距增加而愈嚴重低估。參數估計標準誤的 MSE 表現，從表 7 中可得知彼此之間雖有差

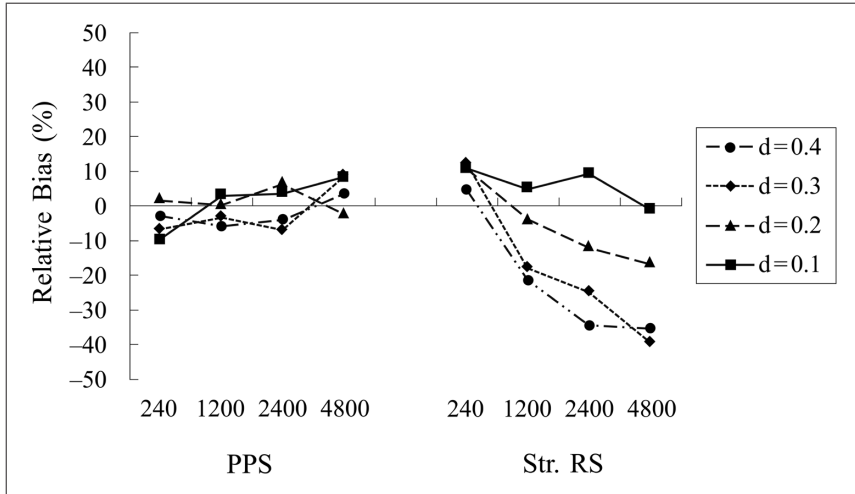


圖 3 類別資料之參數估計標準誤相對偏誤折線圖

異，但差異並不大。綜合上述，顯示參數估計標準誤受到 Str. RS 取樣設計的影響。

陸、實際資料之應用分析

以下以 2007 年臺灣社會變遷調查第五期第一波之實徵資料（張荳雲、廖培珊主編，2008），說明忽略取樣單位內樣本數不相同，而採取不同取樣方法對於參數估計精準度的影響。本研究根據周玉慧與朱瑞玲（2008）利用臺灣社會變遷調查的資料探討臺灣民眾的心理需求、疏離感及身心困擾程度的變遷，其研究指出，臺灣社會變遷調查第五期第一波的綜合問卷中，包含五個題目：覺得許多事情對您來說是個負擔、覺得對自己失去信心、覺得生活毫無希望、覺得緊張不安，無法放輕鬆、覺得家人或親友會令您擔心等，可用來代表受試者的情

緒困擾潛在變項。此結果符合本研究設定的一個潛在變項由五個觀測變項所測量得到的研究架構。表 8 所呈現的是依此模式所估算母體及不同性別在各觀測變項與潛在變項間之因素負荷量，其中以覺得對自己失去信心與情緒困擾間之因素負荷量 (λ_2) 在不同性別間的差距為 0.289 (1.466-1.177)，與模擬研究設計中取樣單位變異為 0.3 相近，因此以下將以不同性別的第二個因素負荷量為研究探討變項。

本資料總母體數為 2,146 人，依照不同性別分層，並以臺灣地區城鄉層級作為基本取樣單位，其不同分層及取樣單位之樣本數如表 9

表 8 情緒困擾驗證性因素分析結果

Factor loading	Total (n=2,146)	Gender	
		Male (n=1,073)	Female (n=1,073)
λ_1	1.000	1.000	1.000
λ_2	1.323	1.466	1.177
λ_3	2.006	2.181	1.826
λ_4	0.714	0.713	0.716
λ_5	0.548	0.624	0.482

表 9 不同分層及 PSU 樣本數分佈

PSU	Male (n=1,073)	Female (n=1,073)
核心都市	255	243
一般都市	286	281
新興市鎮	235	267
傳統市鎮	106	97
一般鄉鎮	139	138
高齡、偏遠	52	47

所示，其中以高齡、偏遠的取樣單位人數最少，因此本研究在取樣單位的取樣部分，從每個分層的 6 個取樣單位中，各抽出 3 個，兩個分層共計有 6 個取樣單位被選取，每個取樣單位分別再抽取 10、20 及 40 人，合計總抽樣數為 60、120 及 240 人。在固定此取樣人數後，以分層隨機取樣及規模等比率取樣進行樣本抽取。研究中 Bootstrap 重複取樣則選擇與模擬研究設計相同的 100 次，以進行參數估計準確性的探討。

實徵資料的分析結果如表 10 所示，可以看出第二個因素負荷量部分，其參數估計值及估計標準誤的 MSE，在兩種不同取樣方法下，皆隨抽樣數增加而略微下降，但差距不大。參數估計值相對偏誤部分，在 PPS 及 Str. RS 取樣設計下，參數估計值的相對偏誤雖有不同，但差異極小且趨近於 0，都在 Hoogland and Boomsma (1998) 建議的參數估計相對偏誤絕對值應小於 5% 的標準之內。參數估計標準誤的相對偏誤部分，在 PPS 的取樣設計時，不論樣本數為何，都維持在標準誤相對偏誤絕對值應小於 10% 的標準，而在 Str. RS 的設計中，雖然在取

表 10 不同取樣方法之實徵資料驗證性因素分析結果

λ_2	n	Parameter		Standard Error	
		PPS	Str. RS	PPS	Str. RS
Relative Bias (%)	60	-0.23	0.15	-3.51	4.21
	120	-0.25	0.13	-3.28	5.22
	240	-0.18	0.12	-4.23	10.34
MSE	60	0.0054	0.0058	0.0004	0.0006
	120	0.0042	0.0055	0.0002	0.0002
	240	0.0036	0.0045	0.0002	0.0001

樣數 60 及 120 時，都在標準之內，但隨取樣數增加，相對偏誤也會愈嚴重，當取樣數達 240 時，相對偏誤即達 10.34%。

柒、結論與建議

科學調查研究的不同取樣設計、取樣數多寡、取樣單位異質性等，對於研究資料之母體推論正確性均有重要的影響。本研究透過 CFA 模式，利用 Bootstrap 取樣程序分別於連續資料及 Likert 五點量表計分的類別資料、不同取樣數及取樣單位異質性的各情境下，探討 PPS 及 Str. RS 取樣設計的權重計算對於母體特性的推論正確性影響，並以臺灣社會變遷調查第五期第一波資料為例，說明兩種取樣設計對於母體推論的影響。整體而言，在連續及類別兩種不同的資料型態，參數估計的相對偏誤都在正負 5% 之內，也就是不同權重的取樣設計、取樣數及取樣單位異質性等因素，對於推論母體 CFA 模式的參數並無影響，而此結果亦與 Stapleton (2008) 的研究結果相同。但參數估計標準誤相對偏誤在這些因素上則有明顯的差異。

進行實徵科學調查研究時，除了考慮取樣樣本代表性之外，總希望在最有效且經濟的情況下，以最少的取樣數正確推論母體特性。根據本研究結果，在 PPS 設計時，參數估計標準誤的相對偏誤，雖然隨取樣數不同而有差異，但差距不大且趨近於 0，顯見在 PPS 取樣設計能提供穩定的母體參數估算而不受樣本數影響。但在 Str. RS 的取樣設計，取樣數增加而造成參數估計標準誤愈嚴重低估。此結果是一值得注意的現象。根據相關研究（例如：Asparouhov, 2005; Asparouhov et al., 2004; Kaplan and Ferguson, 1999; Scheaffer et al., 2006；楊志堅、蔡良庭，2008；楊志堅等人，2009；蔡良庭、楊志堅，2008）指出，從

有限母體進行取樣的分析中，加入「適當」取樣權重的目的在於實際反應母體總數。據此推測，在 Str. RS 取樣設計下，取樣數的增加反而使得參數估計標準誤的相對偏誤愈偏離正確範圍的原因，為多階段結構的複雜取樣設計中，Str. RS 的取樣程序，將每一階段取樣都以簡單隨機取樣方式進行，忽略了取樣單位內的樣本數不一致，導致加權總和不同於母體總數，亦使得原始權重及後續 Bootstrap 權重計算出現不精確的情況。相反的，PPS 取樣設計將每階段的取樣修改為分層不相等機率取樣，也就是分層或是取樣單位被抽中的機率與該單位內的樣本數大小成比例，經由此程序取樣後，母體中每一個樣本最後被抽中機率皆為相等。此設計也修正了 Str. RS 中的取樣權重不等於母體總數問題。而此差異也反應在母體 CFA 模式的推論，導致參數估計標準誤的相對偏誤，且取樣數增加反而使母體推論的錯誤更加嚴重，還不如在正確取樣設計下，使用較小取樣數並配合適當取樣設計及權重。此結果值得進行科學實徵調查研究的工作者一些參考，不能因 Str. RS 在執行程序上相較於 PPS 來得容易，而忽略「不適當」取樣權重對於母體推論 CFA 模式所造成的影響，導致無法正確的推論母體特性，反而有更高的機會造成推論錯誤。

取樣單位變異的影響部分，相關研究 (Kaplan and Ferguson, 1999 ; 楊志堅、蔡良庭, 2008 ; 楊志堅等人, 2009 ; 蔡良庭、楊志堅, 2008) 指出，進行複雜資料結構的科學調查研究時，母體內各取樣單位變異程度對權重的影響力有重要的貢獻。本研究以不同取樣單位間因素負荷量差距代表母體內各子群的異質性，探討不同權重計算的取樣設計在不同異質性情況下，對於母體 CFA 模式的推論影響。綜合來說，PPS 取樣設計時，參數估計及估計標準誤的相對偏誤隨異質性的大小不同而有差異，但絕對值皆低於 5% 及 10% 的標準 (Hoogland and

Boomsma, 1998; Stapleton, 2002)。顯見 PPS 取樣設計能提供精確且穩定的參數估算而不受異質性的影響。在 Str. RS 的取樣設計下，參數估計相對偏誤都在設定標準之內，但只要異質性大於 0.2，估計標準誤相對偏誤即有可能超過設定標準，顯見在 Str. RS 取樣設計時，參數估計標準誤易受取樣單位變異影響。

綜合上述的結果及討論，進行科學調查研究時，儘管相關的統計學教科書或有關取樣的書籍都指出 PPS 及 Str. RS 取樣設計的差異，但有關兩種不同權重計算的取樣設計對於母體 CFA 模式估算的影響，則需更進一步的驗證。而本研究透過電腦數值模擬方式及社會變遷調查研究的資料分析，發現採用 Str. RS 取樣設計時，取樣數愈多模式參數估計愈不精確的現象。Lohr (2010) 指出，實際進行調查研究的工作者，在面對多階層的複雜資料結構時，常在多階層的取樣程序複雜性或忍受採用單一階層的隨機取樣會產生資料的估計相對偏誤間進行取捨。而本研究即便已經採用分層而非單一階層的隨機取樣，在實驗設計下，最大達到約 40% 的參數估計標準誤低估。因此，此估計的相對偏誤值得實務工作者注意。在面對複雜資料結構或多階段取樣設計的調查資料時，研究者欲應用 CFA 模式進行資料分析，建議採用 PPS 取樣設計抽取資料並計算權重才能達到較準確且穩定的母體特性推論。

參考資料

- Andersson, Claes, Gösta Forsman, and Jan Wretman
1987 “Estimating the Variance of a Complex Statistic: A Monte Carlo Study of Some Approximate Technique,” *Journal of Official Statistics* 3(3): 251-265.
- Arbuckle, James L.
2006 *Amos 17.0 User's Guide*. Chicago: SPSS Inc.

- Asparouhov, Tihomir
2005 “Sampling Weights in Latent Variable Modeling,” *Structural Equation Modeling* 12(3): 411-434.
2006 “General Multi-Level Modeling with Sampling Weights,” *Communications in Statistics Theory and Methods* 35(3): 439-460.
- Asparouhov, Tihomir, Linda K. Muthén, and Bengt O. Muthén
2004 *Stratification in Multivariate Modeling*. Retrieved May 17, 2006, from <http://www.statmodel.com/mplus/examples/webnotes/MplusNote921.pdf>.
- Babu, G. Jogesh and Kesar Singh
1983 “Inference on Means Using the Bootstrap,” *The Annals of Statistics* 11(3): 999-1003.
- Chernick, Michael R.
2007 *Bootstrap Methods: A Guide for Practitioners and Researchers*. New York: Wiley.
- Davison, Anthony C. and David V. Hinkley
1998 *Bootstrap Methods and Their Application*. Cambridge: Cambridge University Press.
- Davison, Anthony C., David V. Hinkley, and G. A. Young
2003 “Recent Developments in Bootstrap Methodology,” *Statistical Science* 18(2): 141-157.
- Efron, Bradley and Robert J. Tibshirani
1993 *An Introduction to the Bootstrap*. New York: Chapman & Hall.
- Grilli, Leonard and Monica Pratesi
2004 “Weighted Estimation in Multilevel Ordinal and Binary Models in the Presence of Informative Sampling Designs,” *Survey methodology* 30: 4-14.
- Hoogland, Jeffrey J. and Anne Boomsma
1998 “Robustness Studies in Covariance Structure Modeling: An Overview and a Meta-Analysis,” *Sociological Methods and Research* 26(3): 329-367.
- Jöreskog, Karl G. and Dag Sörbom
1996 *LISREL8: User's Reference Guide*. Chicago: Scientific Software International.
- Kaplan, David and Aaron J. Ferguson
1999 “On the Utilization of Sample Weights in Latent Variable Models,” *Structural Equation Modeling* 6(4): 305-321.
- Kaufman, Steven
1998 “A Bootstrap Variance Estimator for Systematic PPS Sampling,” *Proceedings for the Section on Survey Methods, American Statistical Association*, pp. 769-774.

- Kim, Seock-Ho and Allan S. Cohen
 1998 "Detection of Differential Item Functioning under the Graded Response Model with the Likelihood Ratio Test," *Applied Psychological Measurement* 22(4): 345-355.
- Kovačević, Milorad S., Huang Rong, and Yong You
 2006 "Bootstrapping for Variance Estimation in Multi-Level Models Fitted to Survey Data," *Proceedings for the Section on Survey Methods, American Statistical Association*, pp. 3260-3269.
- Lahiri, Partha
 2003 "On the Impact of Bootstrap in Survey Sampling and Small-Area Estimation," *Statistical Science* 18(2): 199-210.
- Lee, Eun S. and Ronald N. Forthofer
 2006 *Analyzing Complex Survey Data*. Newbury Park, CA: Sage.
- Lohr, Sharon L.
 2010 *Sampling: Design and Analysis* (2nd ed.). Pacific Grove, CA: Duxbury Press.
- Muthén, Linda K. and Bengt O. Muthén
 1998-2011 *Mplus User's Guide*. Los Angeles: Muthén & Muthén.
- Rust, Keith F. and J. N. K. Rao
 1996 "Variance Estimation for Complex Surveys Using Replication Techniques," *Statistical Methods in Medical Research* 5(3): 283-310.
- Sahinler, Suat and Dervis Topuz
 2007 "Bootstrap and Jackknife Resampling Algorithms for Estimation of Regression Parameters," *Journal of Applied Quantitative Methods* 2(2): 188-199.
- Scheaffer, Richard L., III William Mendenhall, R. Lyman Ott, and Kenneth G. Gerow
 2006 *Elementary Survey Sampling*. Boston: Duxbury.
- Stapleton, Laura M.
 2002 "The Incorporation of Sample Weights into Multilevel Structural Equation Models," *Structural Equation Modeling* 9(4): 475-502.
 2006 "An Assessment of Practical Solutions for Structural Equation Modeling with Complex Sample Data," *Structural Equation Modeling* 13(1): 28-58.
 2008 "Variance Estimation Using Replication Methods in Structural Equation Modeling with Complex Sample Data," *Structural Equation Modeling* 15(2): 183-210.
- Wolter, Kirk M.
 2007 *Introduction to Variance Estimation* (2nd ed.). New York: Springer-Verlag.

Wu, C. F. Jeff

1986 “Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis,”
The Annals of Statistics 14(4): 1343-1350.

周玉慧、朱瑞玲

2008 〈變遷中的臺灣民眾心理需求、疏離感與身心困擾〉，《臺灣社會學刊》41:
59-95。

張苙雲、廖培珊（主編）

2008 《臺灣社會變遷基本調查計畫第五期第三次調查執行報告》。臺北：中央研
究院社會學研究所。

楊志堅、蔡良庭

2008 〈評估取樣權重於檢定 Likert 問卷之測量恆等性研究〉，《中華心理學刊》
50(3): 257-269。

楊志堅、蔡良庭、楊志強

2009 〈以 LVQ-ESW 推估受訪者未知抽樣權重之研究〉，《中華心理學刊》51(3):
47-65。

熊瑞梅、張峰彬、林亞鋒

2010 〈解嚴後民眾社團參與的變遷：時期與世代的效應與意涵〉，《臺灣社會學
刊》44: 55-105。

蔡良庭、楊志堅

2008 〈取樣權重值於應用 SEM 模式分析之參數估算正確性研究〉，《教育與心理
研究》31(1): 155-177。