

驗證性因素分析：小樣本參數 估計方法之模擬比較

蔡 坤 宏*

摘 要

本文主要在比較驗證性因素分析中三種常用估計方法在樣本小時的估計結果。因解析上的困難，本文以Monte Carlo法模擬進行分析。根據兩個大小不同模式的模擬結果發現：樣本大小在200以下時，加權最小平方法（WLS）的估計精確程度普遍不如最大概似估計法（ML），且隨樣本大小的減少益加明顯。一般化最小平方法（GLS）在樣本較小時，亦有類似的情形。然而，在模式評估上，GLS顯然地較ML或GLS容易拒絕虛無假設，這種情形在較大的模式中更為明顯。此外，ML亦比GLS易於產生型一的錯誤決策。

本文之發現對於實證研究而言，有著重要的涵意：當實證分析的樣本大小在200以下時，WLS並非是估計CFA模式的好方法。甚至最常用的ML在模式較大時，都可能不是較佳的選擇。另外，需注意的是：雖然GLS是一種較佳的估計選擇，但樣本大小在50以下時，其模式評估過於保守。

*作者為本校企管所畢，現任職中華經濟研究院。

壹、前言

驗證性因素分析法（Confirmatory Factor Analysis, CFA）是一種藉測量指標（measurement indicators）探討潛伏變項（latent variables）間關係之統計分析方法。伴隨著電腦程式（如LISREL VIII、CALIS、EQS等）的發展，CFA目前已廣泛地應用於社會科學各領域中。近年來，學術期刊中經常可發現CFA的應用研究。

應用上，CFA有著不同的估計方法可供選擇。常用的方法包括最大似估計法（Maximum Likelihood Estimation, ML）、一般化最小平方法（Generalized Least Square, GLS）及加權最小平方法（Weighted Least Square, WLS）三種。基本上，這些方法都是加權估計法，但加權的設定內容並不相同，因而估計結果亦會有所不同。不同的程度將因樣本分佈及大小而有差異。根據Browne（1982；1984）的研究顯示，如果樣本分佈峰度（kurtosis）高時，ML估計之容易產生型一的錯誤（type I error），WLS則不受影響。

儘管如此，理論上WLS估計的精確度受到樣本大小的影響（Bollen 1989, P.432），因此，樣本太小可能會影響WLS相對ML或GLS的估計結果。究竟影響情形如何？因以往相關研究多集中於ML估計法（如Anderson & Gerbing 1984; Boomsma 1985; Babakus、Ferguson, JR、Joreskog 1987; Cudeck & Browne 1983; Marsh、Balla & McDonald 1988）或polychoric相關矩陣上各種估計法的比較（如Rigden & Ferguson 1991）。過去並沒有文獻對這個問題做過深入探討，對WLS的小樣本估計特性並不太清楚。就理論而言，有其研究的價值性。再者，採用實證分析常有樣本數並不夠大（多在200以小）之情形（如陳正昌 民80；王健全、陳厚銘 民83）。因此，了解小樣本WLS相對於ML及GLS的的估計，亦有提供實證研究上的參考性。

是以，本文之主要目的即在比較WLS、ML及GLS三種估計方法在小樣本時估計CFA模式的結果。具體而言，本文是想了解WLS是否會因小樣本的影響，而使得估計不如ML或GLS，甚至較容易產生錯誤的決策，冀能使過去相關的研究發現更加一般化，並能提供實證

上參考之依據。因以往的研究結果皆來自單一大小的模式（註一），使得結論的效度（validity）有待考驗，故本文同時設定兩個大小不同的CFA模式進行探討。除前言外，將依序說明CFA模式的估計，研究設計及分析結果，最後，則說明研究發現的涵意及提出建議。

貳、模式估計方法及相關文獻之探討

要了解WLS、ML及GLS的估計式，當然先要說明CFA的基本模式。因此，本節先簡單介紹CFA的基本模式，再討論估計方法及一些相關的研究結果。

一、CFA之基本模式

社會科學的研究，有時會遇到某些變項是無法觀察到的，如人格特質、生活型態、創造力等等，但這些變項卻往往反映到一些可間接測量的指標上，而指標間由於測量同一或類同樣念（construct），常常存在高度相關。這並非一般迴歸計量模式可以處理，然而，線性結構關係模式（Linear Structural Relations Model, LISREL）卻是一種可行的分析模式。但實證上並非總是應用到LISREL的整體模式，有時僅使用到LISREL的測量模式（measurement model）（註二），即CFA模式。

理論上，CFA模式將測量結果分解成由潛伏變項解釋及測量誤差兩部分。一般而言，如果模式中有P個測量指標，且包含S個潛伏變項，模式常以符號表示成（Long 1983, P.57）：

$$(1) \underline{X} = \Lambda_x \underline{\xi} + \underline{\delta}$$

模式中， \underline{X} 為（ $P \times 1$ ）測量指標向量； $\underline{\xi}$ 是（ $S \times 1$ ）之向量，表示測量指標之潛伏變項； $\underline{\delta}$ 為（ $P \times 1$ ）之測量誤差向量； Λ_x 是反映測量指標之因素負荷量矩陣，維數為（ $P \times S$ ）。假設 $\underline{\delta}$ 及 $\underline{\xi}$ 的期望值為0且兩者不共變，則相對式(1)的共變模式可寫為：

註一：FA模式的大小（model size）是指模式中參數相對的多寡而言。

註二：LISREL的整體模式包括測量模式與結構方程模式（structural equation model）兩部分，CFA模式即是其中的一個部分。

$$(2) \Sigma = \Lambda \Phi \Lambda' + \Theta$$

式中， Σ 為測量指標 X 之共變矩陣； Φ 表示潛伏變項 ξ 之共變矩陣； Θ 則是測量誤差 δ 之共變矩陣。 Σ 、 Φ 、 Θ 三者的維數分別為 $(P \times P)$ 、 $(S \times S)$ 及 $(P \times P)$ 。WLS、ML及GLS即是以式(2)為基礎，建立配適函數(fitting function)進行估計。

二、參數估計方法及相關發現

CFA常用的參數估計法為WLS、ML及GLS三種，各有其配適函數形式。相對上，WLS估計的配適函數雖然不同ML或GLS，卻更加一般化。一般而言，WLS的配適函數是誤差加權的二次式(quadratic form)(Bollen 1989, P.425)(註三)：

$$(3) F = (\underline{s} - \underline{\delta}(\underline{\theta}))W^{-1}(\underline{s} - \underline{\delta}(\underline{\theta}))$$

式中， \underline{s} 、 $\underline{\delta}$ 分別是樣本與母體之共變矩陣上(或下)三角的元素以行向量形式堆成的向量， $\underline{\theta}$ 為參數向量。相對式(1)而言， \underline{s} 及 $\underline{\delta}$ 的元素個數皆為 $[1/2 P(P+1)]$ ； W 則為 $[1/2P(P+1)] \times [1/2P(P+1)]$ 之加權矩陣。

根據Browne(1982; 1984)的研究，如果 W 等於漸近樣本共變矩陣(asymptotic covariance matrix)其一致性估計量(consistent estimator)，則WLS合計量在母體任何分佈下漸近有效(asymptotically efficient)，而且樣本大小減1與配適函數值的乘積漸近 χ^2 分佈。此時，參數 $\underline{\theta}$ 的合計漸近共計為：

$$(4) \text{ACOV}(\hat{\underline{\theta}}) = N^{-1} \left\{ \left(-\frac{\partial \underline{\delta}(\underline{\theta})}{\partial \underline{\theta}} \right) \Sigma_{ss}^{-1} \left(-\frac{\partial \underline{\delta}(\underline{\theta})}{\partial \underline{\theta}} \right) \right\}^{-1}$$

式中， Σ_{ss} 為 \underline{s} 之漸進共變矩陣， N 為樣本大小。矩陣展開中的元素可寫為：

註三：Brown(1974)認為式(2)之共變結構式可寫成如一般化的加權迴歸式， $\underline{S} = \underline{\delta}(\underline{\theta}) + \underline{e}$ ，式(3)即是這個式子的一般化最小平方法的估計式。在一般迴歸式中，加權矩陣元素為樣本間的共變，當然，因共變結構式的樣本為共變矩陣中的元素，故式(3)的 W 為樣本共變矩陣之加權矩陣。

$$(5) \text{ACOV}(S_{ij}, S_{gh}) = N^{-1}(\delta_{ijgh} - \delta_{ij}\delta_{gh})$$

式中， δ_{ijgh} 為第四階動差 (the fourth-order moment)。如果母體分佈為常態，則 $\delta_{ijgh} = \delta_{ij}\delta_{gh} + \delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg}$ ，式(4)可進一步簡寫成：

$$(6) \text{ACOV}(S_{ij}, S_{gh}) = N^{-1}(\delta_{ig}\delta_{jh} + \delta_{ih}\delta_{jg})$$

此時，相對於式(4)的配適函數為 (Bollen 1989, P.428)：

$$(7) F = \frac{1}{2} \text{tr} \{S - \Sigma(\underline{\theta})\} V^{-1} \}^2$$

式中， S 、 Σ 分別為樣本及母體之共計矩陣，維數皆為 $(P \times P)$ ； V 則是 $(P \times P)$ 的加權矩陣。比較式(1)及式(7)可知，式(7)是式(1)在常態假設下的特殊情況。當 V 以 S 取代時，式(7)即是 GLS 配適函數。而如果 V 為 Σ 時，當樣本大小增加，式(7)的加權變換估計即將產生與 ML 估計相同的結果 (Jennrich & Lee 1979)。事實上，樣本大時，即使以固定加權 (即指 S) 用的 GLS 估計，結果亦會近同於 ML (Browne 1974)。

從上述各估計方法間關係的說明可以了解，模式估計結果的差異來自於分佈假設及樣本大小。當分佈為常態時，理論上，式(5)至式(7)顯示 WLS 可回到 GLS 或 ML 的近似式，因此，其與 GLS 的估計結果應相同；而與 ML 的估計結果應近似一樣。然而，當樣本小時， Σ 與 S 往往不一致，ML 與 GLS 估計因此有所差異。同時，小樣本之情形下，式(5)、式(6)中 δ 的估計亦會影響 W 的估算，進而影響 WLS 估計的精確度。對參數估計而言，這些對估計結果可能影響不大，但會隨著樣本的減少而趨於明顯。Rigdon & Ferguson (1991) 的研究結果顯示，樣本大小在 100 時，WLS 相對較 ML 或 GLS 的估計精確。但要注意的是：這個研究的資料尺度為順序尺度 (ordinal scale)，而矩陣類型為 polychoric 相關係數。此外，Anderson & Gerding (1984)、Boomsma (1985) 皆未進行各估計法上的比較。

就模式評估來說，理論上，樣本小時，可能會使模式估計之整體評估傾於保守 (即 χ^2 值較小)。但另一方面，因樣本小時的配適較差，換言之，配適值會較大 (即 χ^2 值較

大)，而使得模式評估傾向拒絕虛無假設(null hypothesis)。引申而言，當模式配適改善的情形沒有與樣本的增加幅度一致時（即樣本的增加數無法足以改善配適情形時），會發生樣本大時，發生錯誤決策的機會反而較高。但根據Cudeck & Browne(1983)、Marsh、Balla & McDonald(1988)的研究顯示：ML的估計下，當模式設定與真實模式相同時，各樣本大小水準下之 χ^2 值並不會有明顯的不同。然而，Cudeck & Browne(1983)、Marsh等人（1988）並未探討其他估計方法下的情形。

對WLS、GLS或ML而言，這些影響方面皆相同，但影響程度可能不同。要了解這些問題，理論上應以分析性推導的方式進行探討。但是，從上列各式的說明中，可知在數學解析上相當困難，故本文將以模擬方法為之。

參、研究設計

因本文以模擬方式進行探討，是以在說明模擬結果前，需先說明模擬設定的模式、因子水準及資料產生、與評估準則。

一、模式設定

根據Anderson & Gerbing(1984)、Gerbing & Anderson(1985)、Boomsma(1982;1985)的研究，樣本小於100而且因素中測量指標僅有兩個時，ML估計往往無法收斂。為了避開這個問題，本文設定的CFA模式中，包括兩個相關因素，每一因素各包括三個測量指標。相對於式(2)中各矩陣的參數值設定如下：

$$\Lambda' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$\Phi = \begin{bmatrix} 1.0 & \\ 0.3 & 1.0 \end{bmatrix}$$

$$\Theta = \text{Diag}(0.2 \quad 0.2 \quad 0.2 \quad 0.2 \quad 0.2 \quad 0.2)$$

表中顯示，樣本大小包括5個水準，依序為50,75,100,150及200。對CFA模式的實證估計而言，這些樣本並不大。本文主要目的在觀察樣本大小變動對模式估計及評估的影響。估計方法則包括WLS、ML及GLS三種，利用這些估計法估計模式參數。值得說明的是，估計方法影響評估資料的產生，而操弄資料模擬的因子則是樣本大小水準。依據設定的模式及樣本大小方設定水準，在各估計方法下，經由LISCOMP(Muthen 1988)模擬程式，分別隨機產生100組常態分佈資料。繼而，利用三種估變法分別進行估計，最後，則依據估計及模式評估結果進行分析。

三、評估準則

Jennrich & Lee(1979)、Anderson & Gerbing(1984)、Gerbing & Anderson(1985)、Boomsma(1982;1985)的研究中曾列舉許多評估準則，但有些準則如起始值(starting value)、收斂次數、運算時間的設定等，對實證研究者而言，並不太重要。因為目前的電腦程式比以前改善許多。因此，本文僅選擇一些重要且常用的準則，這些評估準則包括參數估計的精確性及模式評估的正確性。

參數估計精確性的準則上，指的是衡量參數估計結果與實際參數值的差異程度。差異程度愈小，表示該方法之估計結果接近真實值。反之，則表示估計值離真實值較遠。一般而言，常用衡量估計結果的指標為均方誤差(mean squared error, MSE)，這個誤差值可分解成估計變異(variance)及估計偏誤(bias)。前者反映估計值變動的大小，後者則表示參數估計平均值與真實值的差異。理論上，LISREL的參數估計量具有一致性(consistency)(Browne 1982)，因此，無論對WLS、ML或GLS而言，其均方誤值隨著樣本的增加而變小。換言之，估計值會隨機收斂至真實值。本文藉用這個指標主要在觀察小樣本(50至200)下，WLS、MS及GLS估計結果的相對精確性。而評估相對精確性的方法為Friedman檢定(Randle & Wolfe 1979, PP.401-403)。並以H-W法(Hollander & Wolfe 1973)進行多重比較。基本上，這些方法皆屬無母數之統計方法，評估上並不需要任何分佈假設。

模式評估的正確性則指在正確樣式設定下估計參數所複製之共變矩陣與樣本共變矩陣差異的大小。差異愈小，表示複製之共變矩陣與樣本共變矩陣一致性很高。5%的顯著水準

下，雙因素模式的自由度為8，其 χ^2 值應不大於15.51；三因素模式的自由度為24， χ^2 值應不大於36.42。大於這個值，模式評估便產生錯誤決策。如果發生錯誤決策的次數多於模擬次數的5%（以本文而言，即5次）時，一般皆認為其估計方式在此種情況下使用可能不適當。此外，理論上，模擬的 χ^2 值分佈隨著模擬次數的增加，其平均值會呈現常態分佈，即 $\sqrt{n}(\bar{\chi^2} - \nu) \xrightarrow{d} N(0, 2\nu)$ ， n 為模擬次數， ν 為自由度，藉此評估模擬的 χ^2 值是否偏高或偏低。申言之，即使評估的決策正確，亦有可能是其 χ^2 值偏低所致。

肆、模擬結果分析

根據雙因素模式及操弄因子水準模擬產生的資料，經由各種估計方法的分析結果彙整於附表一、附表二及附表三，參數估計差異比較之檢結果則列於表二。而來自三因素模式及操弄因子水準模擬產生的資料，於各估計方法下的分析結果列於附表四、附表五及附表六（註四），參數估計差異比較之檢定結果則列於表三，雙因素模式及三因素模式之評估結果則列在表四中。

一、參數估計之精確性

就參數估計的精確性而言，從表二及表三中可以看出，當樣本大小為50或75時，5%顯著水準下，檢定結果皆拒絕虛無假設。顯示，WLS、ML及GLS三者整體上在參數估計之精確度上有明顯高於ML及GLS。然而，當樣本大小在100以上時，雙因素模式及的檢定結果顯示，WLS、ML及GLS在估計偏誤及變異上皆未有顯著的不同；三因素模式的檢定結果則表示ML在估計偏誤明顯低於GLS，與WLS則無顯著的差異；而WLS的估計變異則顯著地高於ML及GLS。但樣本大小為200時，三者間之估計變異的差異已不明顯。綜言之，WLS或GLS在小樣本時（100以下），參數估計精確度的確不如ML。但樣本大小100以上時，WLS及GLS相對ML估計的精確度視模式大小而定。這顯然與Rigdon & Ferguson(1991)在polychoric

註四：附表四、五、六中僅列出與附表一、二、三相對之參數估計部分，目的為相互比較及節省篇幅。

表二 各估計方法估計模擬之差異比較(雙因素模式)

樣大 本小	方 法 準 則	Friedman 檢定	H-W多重比較
50	估計 偏誤	16.77*	ML>WLS
	估計 變異	16.77*	ML>WLS
75	估計 偏誤	12.08*	ML>WLS
	估計 變異	25.69*	ML>WLS GLS>WLS
100	估計 偏誤	0.46	—
	估計 變異	1.08	—
150	估計 偏誤	3.23	—
	估計 變異	2.92	—
200	估計 偏誤	2.00	—
	估計 變異	4.15	—

註：“*”表示在5%的顯著水準下拒絕虛無
假設；“>”表示顯著低於或小於；“—”
表示無顯著差異。

驗證性因素分析：小樣本參數估計方法之模擬比較

表三 各估計方法估計模擬之差異比較(三因素模式)

樣大 本小	方 法 準 則	Friedman 檢定	H-W多重比較
50	估計 偏誤	0.01<0.05	ML>WLS
	估計 變異	0.13>0.05	—
75	估計 偏誤	9.69*	ML>GLS ML>WLS
	估計 變異	20.46*	ML>WLS GLS>WLS
100	估計 偏誤	8.77	ML>GLS
	估計 變異	19.08	ML>GLS GLS>WLS
150	估計 偏誤	7.54*	ML>GLS
	估計 變異	11.69*	ML>GLS GLS>WLS
200	估計 偏誤	4.54	—
	估計 變異	7.54*	ML>WLS

註1：樣本大小50之水準下，採二項式檢定，欄中為機率值而非 χ^2 值；ML>GLS表示ML的誤差較小。

註2：餘同表二註。

相關矩陣上進行的模擬發現不同（樣本大小100）。

進一步，從附表中可以更地瞭解差異顯著的來源。當樣本大小為50時，雙因素模式的模擬結果顯示，ML參數估計的平均值與實際值的誤差（估計偏誤）均低於GLS或WLS。這種現象在 δ 及 Φ 參數上更加明顯。當樣本大小為75時，情形仍相同。三因素模式除在樣本大小50無法進行WLS估計外，樣本大小為75時，WLS或GLS在 δ 及 Φ 估計偏誤上仍然傾向高於ML，甚而，在樣本大小為100時，結果仍是如此。顯示，前述樣本大小為50或75時差異檢定的顯著性，明顯地來自 δ 及 Φ 部分。在估計變異上，雙因素模式的模擬結果皆顯示，樣本大小為50時，WLS的估計變異大於ML及GLS，而ML在 λ 參數的估計變異又低於GLS。這些發現在樣本大小為75時，大致相同。到樣本大小為100時，這種現象已不明顯。但三因素模式的模擬結果則顯示，在樣本大小為200時，ML在 λ 參數上的估計變異仍傾向低於GLS或WLS，這是差異檢定顯著性之主要來源。

從上面的分析結果發現：樣本愈小時，WLS、GLS或ML估變的精確性的確有不同。估計上，ML較GLS或WLS更為精確。

二、模式評估之正確性

在模式估計完成後，首先要注意的並非是參數估計結果的精確與否，而是模式整體評估的結果，而這個結果將反應在 χ^2 值上。雙因素模式的模擬結果，從表四（第一欄）中可以看出，無論在那一個樣本水準下，WLS產生的 χ^2 值均明顯高於ML及GLS，而以GLS最大。當樣本大小為50時，WLS的 χ^2 值為9.33，ML及GLS則分別為9.18與7.09。而當樣本大小為200時，WLS的 χ^2 值是8.15，ML及GLS則分別是8.09與7.08。其中，WLS在樣本大小為150時， χ^2 值仍明顯地偏高；而ML或GLS則沒有這種情形。

就錯誤決策的發生情形來看，在樣本大小50時，WLS產生的錯誤次數有8次，ML也是8次，GLS並沒有發現。當樣本增至75時WLS的錯誤次數有13次，GLS與ML則皆是4次。當樣本增至150時，WLS發生的次數為仍然偏高，相對地ML與GLS仍未高於臨界次數。樣本大小為200時，WLS降至7次，ML與GLS則仍低於5次。顯示，雖然WLS發生的錯誤次數隨著樣本大小的增加而有減少的情形，但這種情形從樣本50增到75時並不成立。這意謂著：樣

驗證性因素分析：小樣本參數估計方法之模擬比較

本從小慢慢增大的過程，配適的情形並未有明顯改善，樣本的增加反而造成了 χ^2 值的偏高。此外，即使樣本大小增至200時，WLS仍然可能會造成較高的錯誤決策。相對而言，ML除了在小樣本時，可能會產生較高的錯誤評估外，樣本增加時，這種情形即減少至臨界範圍下。對GLS來說，在樣本大小各水準下，較不易產生錯誤決策。但值得說明的是，GLS雖然在樣本大小50時沒有產生錯誤的決策，但其評估值卻明顯地過於保守（ χ^2 值偏低）。

表四 模式評估之模擬結果

樣本大小	估計方法	雙因素模式		三因素模式	
		χ^2 值	錯誤次數	χ^2 值	錯誤次數
50	WLS	9.33 ¹ (4.53)	8	—	—
	ML	9.18 ¹ (4.88)	8	25.97 ¹ (7.89)	10
	GLS	7.09 ² (3.84)	0	20.58 ² (5.16)	0
75	WLS	9.38 ¹ (4.61)	13	40.75 ¹ (15.18)	60
	ML	9.07 ¹ (3.97)	4	27.17 ¹ (8.20)	14
	GLS	7.44(3.97)	4	22.51 ² (5.85)	3
100	WLS	9.09 ¹ (4.93)	10	33.65 ¹ (10.82)	37
	ML	8.67 ¹ (4.01)	4	26.17 ¹ (8.12)	8
	GLS	7.60(3.46)	2	23.11(5.94)	2
150	WLS	8.42 ¹ (5.42)	10	29.78 ¹ (9.57)	24
	ML	8.21(4.08)	4	25.67(8.10)	12
	GLS	8.23(3.08)	2	24.85(5.64)	4
200	WLS	8.15(4.24)	7	29.08 ¹ (7.74)	12
	ML	8.09(3.62)	3	25.54(8.11)	7
	GLS	7.80(3.35)	0	24.29(5.93)	3

註：1.5%顯著水準下， χ^2 值明顯偏高。

2.5%顯著水準下， χ^2 值明顯偏低。

上述的模擬發現與三因素模式的結果並不全相同。WLS在樣本大大為200時， χ^2 值仍然偏高，ML則與雙因素模式結果相同，樣本大小在100（含）以下時， χ^2 值有偏高的現象。在錯誤決策上，當樣本增加到200時，WLS與ML產生錯誤的次數雖然已分別降到12次及

7次，但仍然高於臨界次數，而GLS在各樣本水準下，並未產生過高型一的錯誤決策。顯示，模式擴大後，WLS與ML產生錯誤的次數隨之增加。對WLS而言，這種現象尤其明顯，即使樣本大小為200時，情況依然如此。這與Rigdon & Ferguson(1991)在polychoric相關矩陣上的發現不同（註五）。

此外，從表四亦可看出，樣本大小各水準下，三種估計方法在雙因素模式產生的 χ^2 值於並沒有明顯的不同；但三因素模式中，WLS在樣本大小在100以下與100以上卻有所差異。顯示Cudeck & Browne(1983)、Marsh等（1988）在ML的研究發現並無法概括到WLS。

綜合前面的評估可以得知，在參數估計上，WLS傾向高於GLS或ML。當樣本逐漸減少時，這種現象會較明顯。而且，其間的差異亦會較大。模式評估上，WLS發生的錯誤機會則明顯相對高於ML或GLS。這些發現解釋了Bollen（1989，P.432）的實例說明。模式擴展後，這種現象更為明顯。而且，ML亦比GLS更容易產生錯誤的決策。但是，值得注意的是，儘管GLS在雙因素模式樣本大小50或三因素模式樣本大小100（含）以下，較ML及WLS不會產生型一錯誤，但其產生的 χ^2 值很可能亦不正確。換言之，此時其樣本大小減1（N-1）與配適值（F）的乘積尚未近似 χ^2 分佈。

伍、結論與建議

本文主要在探討小樣本下WLS、ML、GLS三種CFA估計方法的相對特性。結果發現：

(1)參數估計上，由於受到樣本較小時加權矩陣W估計不穩定及不精確的影響，WLS估計的精確性不如ML。對GLS而言，亦是如此。

(2)模式評估方面，WLS傾向產生較高型一的錯誤決策，即使樣本大小為200，亦是高於ML或GLS。而ML則只有在樣本較小（50）時，才會產生較高的錯誤決策。當模式較大時，WLS產生錯誤決策的情形將更加嚴重，樣本增至200時，錯誤決策產生的機會仍然偏高。對

註五：Rigdon & Ferguson(1991)在polychoric相關的研究發現WLS的評估結果較佳。但事實上，這有兩個問題要注意。一者，相關矩陣型態上的ML及GLS模式評估結果不正確（Joreskog & Sorbom 1989，P.31）。二者，PRELIS7在WLS的加權估計上太過簡化，因此，其結果並不一定正確。

驗證性因素分析：小樣本參數估計方法之模擬比較

ML而言，即使樣本大小為200，其錯誤決策發生的機會仍然高於GLS。而GLS雖然不易產生錯誤決策，但其在樣本較小下產生的 χ^2 值並不可靠。

對實證研究而言，這些發現有著重要的涵意：(1)WLS雖然有著分佈自由及有效的估計特性，但估計卻較不精確。甚者，在模式評估上當樣本大小200以下時，使用WLS之模式評估結果很可能不正確。因此，小樣本下的CFA估計，WLS並非一種好方法。(2)對ML而言，雖然其參數估計相當精確，但模式評估可能會產生較高型一的錯誤決策。因此ML非是小樣本時估計CFA模式的良好選舉。(3)小樣本下的CFA模式估計最好使用GLS，而且根據蔡坤宏（民83）的模擬研究顯示，GLS在非常態小樣本下（N=100）具有強韌性（robustness）。但仍需注意的是，當樣本較小時（N=50），GLS產生的 χ^2 值傾向保守，其評估結果很可能不正確。申言之，樣本太小時，使用 χ^2 進行整體評估並無意義，此時，其評估指標應以其他配適指標為之較適合。

本文的發現主要建立在常態分佈的假設上，然而，實務上的資料分佈常會無法滿足這假設，故改變這種假設對研究必有其需要性，值得加以探討。再者，本文說明的模式是特定的CFA模式，雖然其存在的發現對其他同類模式，甚至整體LISREL模式而言，皆會有類似現象存在。但因模式的不同，意謂著參數估計數目的多寡，對三種估計法之模式評估的影響程度，可能亦因之有所差別，擴大模式到整體模式的比較研究似乎是需要的。此外，GLS在小樣本時產生的 χ^2 值明顯傾向低估，雖然不易造成型一錯誤，但其是否因此而產生較高的型二錯誤，亦是值得探究的課題。

附表一 WLS參數估計之模擬結果（雙因素模式）

參數	50		75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	1.019	0.132	1.010	0.093	0.996	0.079	0.994	0.063	1.011	0.052
$\lambda_{31}(1.0)$	1.018	0.146	1.021	0.089	0.998	0.078	1.004	0.056	1.001	0.048
$\lambda_{52}(1.0)$	1.024	0.128	0.989	0.098	1.006	0.077	1.002	0.065	0.998	0.057
$\lambda_{62}(1.0)$	1.012	0.127	0.986	0.107	1.005	0.083	1.004	0.070	0.999	0.053
$\delta_{11}(0.2)$	0.171	0.068	0.184	0.050	0.181	0.041	0.183	0.038	0.194	0.030
$\delta_{22}(0.2)$	0.183	0.065	0.179	0.058	0.195	0.042	0.195	0.035	0.190	0.030
$\delta_{33}(0.2)$	0.174	0.067	0.176	0.053	0.185	0.051	0.185	0.034	0.199	0.037
$\delta_{44}(0.2)$	0.180	0.070	0.187	0.057	0.192	0.039	0.187	0.039	0.193	0.036
$\delta_{55}(0.2)$	0.179	0.068	0.179	0.049	0.183	0.048	0.198	0.040	0.195	0.029
$\delta_{66}(0.2)$	0.170	0.071	0.182	0.058	0.186	0.042	0.195	0.037	0.191	0.035
$\Phi_{11}(1.0)$	0.937	0.269	0.965	0.211	0.972	0.176	0.979	0.141	0.977	0.135
$\Phi_{21}(0.3)$	0.317	0.219	0.294	0.118	0.300	0.123	0.281	0.104	0.297	0.078
$\Phi_{22}(1.0)$	0.921	0.278	0.987	0.219	0.981	0.193	0.945	0.140	0.987	0.130

註：各參數估計之均方誤差(MSE)可以由估計之變異數及估計值與實際值之差異推得。

附表二 ML參數估計之模擬結果（雙因素模式）

參數	50		75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	1.007	0.104	1.013	0.081	1.014	0.070	0.996	0.059	1.006	0.051
$\lambda_{31}(1.0)$	0.012	0.090	1.002	0.078	0.996	0.060	0.991	0.056	1.001	0.047
$\lambda_{52}(1.0)$	0.992	0.105	1.004	0.082	0.994	0.080	1.000	0.058	1.001	0.049
$\lambda_{62}(1.0)$	1.008	0.095	1.000	0.079	0.993	0.070	0.991	0.056	0.999	0.048
$\delta_{11}(0.2)$	0.192	0.066	0.190	0.051	0.196	0.043	0.200	0.035	0.193	0.034
$\delta_{22}(0.2)$	0.182	0.059	0.195	0.049	0.196	0.047	0.194	0.033	0.199	0.031
$\delta_{33}(0.2)$	0.204	0.060	0.192	0.049	0.198	0.041	0.198	0.039	0.196	0.031
$\delta_{44}(0.2)$	0.196	0.065	0.200	0.048	0.192	0.047	0.201	0.034	0.200	0.032
$\delta_{55}(0.2)$	0.201	0.069	0.204	0.053	0.204	0.047	0.197	0.039	0.200	0.033
$\delta_{66}(0.2)$	0.179	0.063	0.190	0.052	0.194	0.044	0.200	0.041	0.195	0.033
$\Phi_{11}(1.0)$	0.991	0.254	1.004	0.187	0.966	0.174	1.013	0.146	0.982	0.117
$\Phi_{21}(0.3)$	0.309	0.153	0.305	0.140	0.270	0.101	0.306	0.091	0.305	0.082
$\Phi_{22}(1.0)$	0.967	0.229	1.007	0.208	0.957	0.158	1.002	0.143	0.995	0.107

註：同附表一。

驗證性因素分析：小樣本參數估計方法之模擬比較

附表三 GLS參數估計之模擬結果（雙因素模式）

樣本大小 參數	50		75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	1.019	0.113	1.017	0.010	0.999	0.073	0.998	0.071	1.000	0.048
$\lambda_{31}(1.0)$	1.020	0.109	1.003	0.085	0.988	0.078	1.004	0.061	1.002	0.044
$\lambda_{52}(1.0)$	1.009	0.111	1.023	0.098	1.007	0.072	1.002	0.066	0.999	0.053
$\lambda_{62}(1.0)$	1.013	0.115	1.023	0.097	1.003	0.069	1.001	0.060	1.006	0.057
$\delta_{11}(0.2)$	0.180	0.054	0.184	0.058	0.183	0.042	0.185	0.039	0.187	0.032
$\delta_{22}(0.2)$	0.173	0.065	0.178	0.050	0.185	0.043	0.192	0.035	0.192	0.031
$\delta_{33}(0.2)$	0.164	0.064	0.184	0.053	0.186	0.044	0.191	0.033	0.190	0.027
$\delta_{44}(0.2)$	0.172	0.060	0.187	0.054	0.184	0.047	0.190	0.035	0.191	0.032
$\delta_{55}(0.2)$	0.173	0.064	0.174	0.044	0.180	0.045	0.194	0.034	0.200	0.032
$\delta_{66}(0.2)$	0.175	0.064	0.177	0.049	0.186	0.043	0.190	0.040	0.187	0.033
$\Phi_{11}(1.0)$	0.943	0.231	0.950	0.194	0.993	0.175	0.970	0.150	0.986	0.108
$\Phi_{21}(0.3)$	0.271	0.159	0.301	0.130	0.300	0.113	0.296	0.094	0.300	0.084
$\Phi_{22}(1.0)$	0.936	0.218	0.947	0.218	0.967	0.176	0.955	0.131	0.974	0.128

註：同附表一。

附表四 WLS參數估計之模擬結果（三因素模式）

樣本大小 參數	75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	1.005	0.146	0.998	0.093	0.995	0.065	0.994	0.050
$\lambda_{31}(1.0)$	1.011	0.139	1.003	0.094	1.005	0.067	1.005	0.057
$\lambda_{52}(1.0)$	1.010	0.093	0.987	0.087	0.994	0.064	1.004	0.055
$\lambda_{62}(1.0)$	1.009	0.120	0.991	0.095	0.997	0.065	1.006	0.058
$\delta_{11}(0.2)$	0.162	0.069	0.171	0.049	0.190	0.042	0.179	0.032
$\delta_{22}(0.2)$	0.170	0.065	0.175	0.053	0.186	0.043	0.192	0.036
$\delta_{33}(0.2)$	0.170	0.067	0.176	0.050	0.185	0.040	0.184	0.035
$\delta_{44}(0.2)$	0.168	0.061	0.174	0.051	0.180	0.035	0.190	0.034
$\delta_{55}(0.2)$	0.167	0.062	0.180	0.054	0.185	0.043	0.186	0.033
$\delta_{66}(0.2)$	0.163	0.055	0.184	0.048	0.182	0.040	0.189	0.031
$\Phi_{11}(1.0)$	0.940	0.274	0.957	0.220	0.967	0.155	0.996	0.119
$\Phi_{21}(0.3)$	0.290	0.196	0.298	0.146	0.303	0.103	0.292	0.082
$\Phi_{22}(1.0)$	0.942	0.220	0.969	0.205	0.981	0.154	0.967	0.129

註1：同附表一。

註2：WLS在樣本數50時無法估計參數。

註3：表中僅列出與雙因素模式相對之參數估計部分。

附表五 ML參數估計之模擬結果（三因素模式）

參數	50		75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	1.011	0.095	0.997	0.082	0.996	0.069	0.992	0.054	0.997	0.048
$\lambda_{31}(1.0)$	1.002	0.096	1.006	0.089	0.984	0.069	0.999	0.054	1.005	0.047
$\lambda_{52}(1.0)$	0.989	0.115	0.998	0.077	0.992	0.064	1.008	0.060	0.995	0.040
$\lambda_{62}(1.0)$	1.008	0.096	0.990	0.085	1.004	0.067	1.005	0.052	0.999	0.050
$\delta_{11}(0.2)$	0.198	0.066	0.198	0.053	0.197	0.044	0.196	0.034	0.197	0.032
$\delta_{22}(0.2)$	0.180	0.055	0.195	0.053	0.191	0.039	0.197	0.039	0.204	0.035
$\delta_{33}(0.2)$	0.200	0.074	0.182	0.049	0.200	0.045	0.194	0.035	0.196	0.033
$\delta_{44}(0.2)$	0.175	0.057	0.188	0.041	0.195	0.048	0.196	0.036	0.198	0.028
$\delta_{55}(0.2)$	0.203	0.058	0.191	0.048	0.203	0.038	0.203	0.038	0.196	0.029
$\delta_{66}(0.2)$	0.188	0.067	0.208	0.046	0.197	0.042	0.195	0.041	0.196	0.030
$\Phi_{11}(1.0)$	0.957	0.214	0.991	0.190	1.008	0.169	0.996	0.133	0.995	0.122
$\Phi_{21}(0.3)$	0.283	0.152	0.287	0.131	0.306	0.123	0.296	0.085	0.297	0.079
$\Phi_{22}(1.0)$	0.960	0.238	1.025	0.196	1.011	0.169	0.987	0.120	0.998	0.130

註：同附表一及附表四註3。

附表六 GLS參數估計之模擬結果（三因素模式）

參數	50		75		100		150		200	
	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差	估計值	標準差
$\lambda_{21}(1.0)$	0.994	0.122	0.996	0.097	1.005	0.069	1.006	0.058	0.991	0.054
$\lambda_{31}(1.0)$	0.986	0.134	0.998	0.095	1.006	0.078	1.013	0.063	0.998	0.050
$\lambda_{52}(1.0)$	0.994	0.120	0.994	0.087	1.018	0.086	1.001	0.067	0.994	0.050
$\lambda_{62}(1.0)$	1.006	0.115	0.997	0.091	1.013	0.076	0.991	0.050	0.989	0.049
$\delta_{11}(0.2)$	0.154	0.064	0.163	0.050	0.175	0.038	0.187	0.033	0.184	0.034
$\delta_{22}(0.2)$	0.143	0.065	0.156	0.049	0.172	0.045	0.185	0.034	0.191	0.035
$\delta_{33}(0.2)$	0.157	0.061	0.175	0.054	0.175	0.048	0.181	0.031	0.192	0.028
$\delta_{44}(0.2)$	0.150	0.060	0.164	0.052	0.179	0.041	0.180	0.033	0.185	0.034
$\delta_{55}(0.2)$	0.152	0.059	0.166	0.049	0.169	0.044	0.182	0.033	0.186	0.033
$\delta_{66}(0.2)$	0.155	0.055	0.167	0.049	0.176	0.050	0.185	0.039	0.190	0.031
$\Phi_{11}(1.0)$	0.867	0.230	0.933	0.182	0.914	0.175	0.947	0.140	0.966	0.121
$\Phi_{21}(0.3)$	0.270	0.164	0.281	0.128	0.299	0.115	0.294	0.093	0.293	0.081
$\Phi_{22}(1.0)$	0.899	0.234	0.949	0.211	0.916	0.188	0.985	0.146	0.986	0.111

註：同附表一。

參 考 文 獻

- 王健全、陳厚銘（民83）。促進產業升級條例研究發展之投資抵減效果評析。台北：中華經濟研究院。
- 陳正昌（民80）。台灣地區教育發展、社會變遷與犯罪問題研究。政治大學教育研究所未出版碩士論文。
- 蔡坤宏（民83）。LISREL一些模擬結果之涵意與分析。教育研究，第36期，頁49-63。
- Anderson, J. C. & D. W. Gerbing (1984). "The Effect of Sampling Error on Convergence, Improper Solutions, and Goodness-of-Fit Indices for Maximum Likelihood Confirmatory Factor Analysis." Psychometrika, 49:155-173.
- Babakus, E., Ferguson, Jr C. E., & K. G. Joreskog (1987). "The Sensitivity Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions." Journal of Marketing Research, 24:222-8.
- Ceduck, R. & Browne, M.W.(1983). "Cross-Validation of Covariance Structures." Multivariate Behavioral Research, 18:147-67.
- Bollen, K. A. (1989). Structural Equations with Latent Variables. New York: John Wiley & Sons.
- Boomsma, A. (1982). On the Robustness of LISREL against Sample Size and Nonnormality. Amsterdam: Sociometric Research Foundation.
- Boomsma, A. (1985). "Nonconvergence, Improper Solutions, and Starting Values in LISREL Maximum Likelihood Estimation." Psychometrika, 50:229-42.
- Browne, M. W. (1974). "Generalized Least-Squares Estimators in the Analysis of Covariance Structures." South Africa Statistical Journal, 8:1-24.
- Browne, M. W. (1982). Covariance Structures. In D. M. Hawkins, ed., Topics in Multivariate Analysis. Cambridge: Cambridge University Press, pp. 72-141.
- Browne, M. W. (1984). Asymptotic Distribution Free Methods in Analysis of Covariance Structures. British Journal of Mathematical and Statistical Psychology, 37: 62-83.
- Hollander, M. & D. A. Wolfe (1973). Nonparametric Statistical Methods. New York: John Wiley & Sons.
- Jennrich, R. I. & S. Y. Lee (1979). "A Study of Algorithms for Covariance Structure Analysis With Specific Comparisons Using Factor Analysis." Psychometrika, 44: 99-113.
- Joreskog, K. G. & Dag Sorbom(1989). LISREL 7 User's Reference Guide. Scientific Software Inc.
- Long, J. Scott (1983). Covariance Structural Models-An Introduction to LISREL. Beverly Hills

and London: Sage Publications.

Marsh, H. W., J.R. Balla, & R.P. McDonald (1988). "Goodness-of-Fit Indexes in Confirmatory Factor Analysis: The Effect of Sample Sizes." Psychological Bulletin, 103: 391-410.

Muthen, B. O. (1988). LISCOMP: Analysis of Linear Structural Equations with a Comprehensive Measurement Model. Mooresville, IN: Scientific Software, Inc.

Randle, R. H. & D. A. Wolfe (1979). Introduction to the Theory of Nonparametric Statistics. New York: John Wiley & Sons.

Rigdon, E. E. & C. E. Ferguson, Jr.(1991). "The Performance of the Polychoric Correlation Coefficient and Selected Fitting Functions in Confirmatory Factor Analysis With Ordinal Data." Journal of Marketing Research, 28: 491-497.