政治科學論叢／第 73 期／民國 106 年 9 月／頁 81-126

探討以電訪資料及「入選機率調整法」修正網路調查偏誤的可行性*

俞振華**、涂志揚***

摘　要

雖然網路調查已在我們日常生活中隨處可見，但民眾對於網路調查的結果往往半信半疑。對於網路調查最直接的批評即爲，網路調查的樣本非來自隨機抽樣程序，爲非隨機樣本（Non-random sample）。因此，網路調查的結果往往無法用來推論母體特徵。本研究分析多波「臺灣選舉與民主化研究」（Taiwan's Election and Democratization Studies, TEDS）計畫中的網路及電話調查平行測試資料，利用同時置於兩種不同調查方式中的所謂「Webographic」問題（或稱「網路使用者特徵」變數），來辨識網路使用者與非使用者的差異處，並藉入選機率調整法（propensity score adjustment, PSA）調整網路調查結果。換言之，本研究將電訪調查結果視爲代表「正確結果」的「對照組」（參考樣本），

* DOI:10.6166/TJPS.73(81-126)。本研究進行網路調查所使用的平台，部份建置係接受行政院科技部計畫「網路民意調查的建置與應用」（計畫編號 101-2410-H-004-079-SSS）補助。同時，本文使用的資料全部係採自「2012 年至 2016 年『臺灣選舉與民主化調查』四年期研究規劃（1-4）：總統滿意度電訪及網路調查案」（TEDS2012-2016 PA 系列）（記畫編號 101-2420-H004-034-MY4）。「臺灣選舉與民主化調查」（TEDS）爲行政院科技部補助之多年期計畫，計畫主持人爲國立政治大學黃紀教授，詳細資料請參閱 TEDS 網頁：http://www.tedsnet.org。作者感謝上述機構及人員提供資料協助，惟本文之內容概由作者自行負責。

** 政治大學選舉研究中心副研究員暨政治學系副教授，E-mail: ericyu@nccu.edu.tw。
***政治大學政治學系碩士，E-mail: tu.chihyang@gmail.com。

收稿日期：105 年 12 月 19 日；通過日期：106 年 8 月 29 日

臺灣大學學術期刊資料庫

並透過統計模型及結合電訪資料來調整網路調查結果（觀察樣本）。總之，本研究旨在從多個角度驗證 PSA 結合電訪與網路調查資料的可行性，並提出此項作法的優勢與限制。

關鍵詞：網路調查、非隨機樣本、Webographic 變數（「網路使用者特徵」變數）、入選機率調整法

臺灣大學學術期刊資料庫

壹、前言

近年來臺灣民眾使用網路的比例已大幅增加，網路甚至已成爲大多數民眾日常生活當中不可或缺的一部份。[1] 隨著網路平台的崛起，網路調查是否可以取代傳統的調查方法（譬如電話調查或郵寄問卷），一躍成爲新的調查利器？事實上，過去數年來，使用網路調查資料的各項研究已有漸漸增加的趨勢（Mercer et al, 2017；Dutwin and Buskirk, 2017；Baker et al., 2010；Dever et al., 2008；Hewson et al., 2003；Schonlau et al., 2007），並有諸多利用網路調查資料撰寫學術性研究報告或論文，刊登在專業學術期刊或書籍中。[2] 不過，除了極少數網路調查平台所產出的資料外（譬如美國的 YouGov 或 Knowledge Network），學界及實務界對於多數網路調查結果仍有諸多疑慮，主要的原因在於：儘管網路普及已成趨勢，但使用者急遽地增加不代表所謂的「數位落差」（digital divide）問題已被解決（Norris, 2001），即網路的使用者仍然以特定地區（譬如經濟發展程度較高的都會區）的民眾，或是特定族群（譬如年輕、教育程度較高者）爲主要對象（俞振華，2013；游清鑫等，2007；Nie and Erbing, 2000）。也正因爲數位落差的問題持續存在，網路調查結果的樣本代表性也就一直被外界質疑，於是限制了網路調查的發展（翟本瑞，2002）。畢竟，當樣本代表性失眞時，估計值的偏誤將使得我們無法利用樣本資料推論母體特性。然而，在較不強調樣本代表性的市場調查領域，網路調查早已成爲利器，許多跨國市場調查公司紛紛在臺灣設立據點，積極透過網路進行產品行銷或針對特定族群實行消費行爲調查。[3]

短期來看，網路調查的樣本代表性問題不會因爲網路日漸普及而獲得根本上的解決。但難道因爲這層障礙就放棄發展網路調查？我們是否能發

---

1 臺灣網路資訊中心委託鄭天澤等（2015）的調查報告即顯示，過去 10 年來（自 2006 年起），12 歲以上臺灣民眾的個人上網率，已由 67.2% 增加到 83.7%，家户上網比例也達 89.2%。依此推估，目前全臺上網人口已超過 1,800 萬人。

2 以跨國調查公司 GfK 的 Knowledge Panel 爲例，截至 2012 年爲止，已有超過 400 件學術出版品（領域橫跨政治學、心理學、公共衛生等）使用該公司的網路調查資料（Gfk, 2012）。

3 數家較大規模的網路市場調查公司包括東方快線、波士特、點眾等皆有外資背景。

臺灣大學學術期刊資訊庫

展某種模式，譬如利用其他調查方式所獲得的資料及統計技術，來修正網路調查資料的樣本代表性，使得網路調查所獲得的結果仍能正確地推估母體特徵？本研究旨在發展網路調查資料的修正機制。具體來說，我們利用目前臺灣最普遍的調查模式－即電話調查所獲得的資料，作爲「對照組」（reference group）（即參考樣本），繼而透過該「對照組」並設定統計模型及採用不同模式的入選機率調整法（propensity score adjustment, PSA），來修正網路調查結果（即觀察樣本）所可能產生的偏誤。本文第貳節探討網路調查的偏誤與可行的修正機制，第參節具體說明資料來源及本研究所採行的 PSA 修正方式，第肆節爲資料分析與比對，第伍節利用本研究的資料，進一步探討網路調查誤差與樣本數的關聯性，第陸節爲討論與建議。

貳、網路調查的偏誤及修正

傳統的調查方式，不論是面對面訪問或電話調查，一個很重要的假定是：調查的樣本爲隨機樣本，即每位潛在受訪者接受調查的機率是相同的（洪永泰，2009）。依照樣本建立是否採隨機抽取的方式，學者將網路調查方式大致歸類爲兩種類型（Elliott and Valliant, 2017；Couper, 2000），包括非機率樣本（non-probability sample）及機率樣本（probability sample）。前者是指受訪者的來源不具代表性，或強調受訪者可自我選擇參與網路調查。由於這類型調查的樣本未透過隨機抽樣的程序，我們無法利用其調查結果推估母體特徵。後者則是指調查樣本基於特定型式的隨機抽樣，於是在樣本具有代表性的情況下，我們可以利用調查結果推論母體特徵。

不過，網路調查不論是否基於機率樣本，其樣本代表性仍被視爲存有嚴重樣本自我選擇機制的問題（Mercer et al., 2017；Yeager et al., 2011；Couper et al., 2007；Rivers, 2007）。Lee (2004；2006) 歸結網路調查的各項程序後即指出，網路調查從樣本招募、抽樣、到回答問卷等過程皆存在選擇機制的問題，而這些問題使得最後網路調查結果和母體產生了偏誤。具體而言，由於不是所有人都有網路，因此網路母體就已經和目標母體有所差距。在樣本招募的過程當中，又只有那些自願參與的人才會加入網路調查受訪者名單。而當我們利用受訪者名單抽樣時，代表性又離目標母體遠了些。而在最

臺灣大學學術期刊資料庫

後一階段，顯然不是抽到的樣本都會回覆問卷，只有自願回答的人才會回答。於是，對於目標母體而言，最後回覆網路問卷者的代表性又更差了。以政治大學選舉研究中心的網路調查平台「線上調查實驗室」（PollcracyLab）爲例，就算該平台的受訪者名單是透過電話調查蒐集而來，屬於機率樣本（或至少是「準機率樣本」: pseudo- probability sample），但樣本特徵仍反應出所謂的「數位落差」，即會願意加入網路調查受訪者資料庫的民眾主要還是以年紀輕、受過高等教育、及居住在都會區的族群爲主（俞振華，2013）。另外，願意加入受訪者名單的民眾和實際會接受調查的受訪者在人口特徵方面亦存有差異，即有些人雖然願意加入受訪者資料庫，但從來不填答寄給他／她的問卷。總之，網路調查結果之所以難以直接用來推論母體特徵，主要原因在於我們根本無從定義網路調查的抽樣架構（Elliot and Valliant, 2017 ; Lee, 2006），但電話調查並沒有這個問題，這也是「線上調查實驗室」的受訪者名單取自於電話調查樣本的主因。不過實務上，就算採取用電訪蒐集而來的所謂「機率樣本」作爲網路調查的受訪者名單，網路調查結果的選樣偏誤問題仍舊存在。

以上說明點出了現階段網路調查在理論與實務面所面臨的困境與挑戰。雖然傳統上，當調查資料代表性失眞時，爲了避免對母體推估偏誤，研究者往往會將調查結果加權處理。常見的加權方法包括多變項反覆加權（raking）、事後分層（post-stratification）、或是前兩種方法的延伸加權法（洪永泰，1996；陳肇男，2001）。但由於以上加權方法只是讓加權結果中某些人口特徵資料接近母體參數，而忽略了其它包括認知、態度、或行爲變數的分佈，因此不見得能完全解決調查推估的偏誤（杜素豪等，2009；杜素豪，2015）。換言之，如果要全面地衡量並修正網路調查樣本偏誤的情況，除了改善機率抽樣程序這項斧底抽薪之計外，至少要將樣本調整至和賴以權充「黃金標準」的「對照組」差不多，而這項調整機制無法只利用少數人口變數事後加權。也因此，尋找可用來修正網路調查樣本的「對照組」遂成爲是否能夠進一步突破網路調查結果偏誤的關鍵點。[4] 至於這個

[4] 廣義來說，「對照組」的核心概念，在於使用機率樣本或甚至使用普查資料來修正非機率樣本的網路調查結果。目前幾項網路調查結果的調整機制，主要包括以下四種：

臺灣大學學術期刊資料庫

對照組的條件應該爲何？Lee (2006) 及 Lee and Valliant (2009) 指出，當我們利用某一非母體的「對照組」（即參考樣本）資料來修正網路調查結果時，這個對照組的涵蓋率及回答率都要比網路調查來的高才行，譬如利用 Random-Digit-Dial (RDD) 的電話訪問或是大型面訪樣本所得的調查結果。[5] 近來諸多學者採用入選機率調整法（propensity scores adjustment, PSA）來調整網路調查的結果，即將網路調查和「對照組」（即參考樣本）資料合併後，透過估計參與網路調查可能性的參數，同時考慮人口特徵變項及其他非人口變項對網路調查結果（即觀察樣本）進行調整，發現確實較傳統加權方式更能消除網路調查的涵蓋誤差（Lee, 2004；2006；Lee and Valliant, 2009；Loosveldt and Sonck, 2008；Schonlau et al., 2004；Schonlau et al., 2007）。不過，不論是以哪種樣本做爲「對照組」來修正網路調查結果，過去相關實證研究幾乎都是西方的研究，國內在這方面的實證研究幾乎爲零，只有杜素豪等（2009）及杜素豪（2015）利用該方式以面訪資料作爲對照組來調整電訪調查結果。換言之，雖然在西方國家，利用入選機率調整法並以電訪或面訪資料作爲「對照組」來修正網路調查資料的方法行得通，但在臺灣的適用性則仍待進一步驗證。

既然「對照組」是修正網路調查資料的關鍵，那我們應該如何選擇「對照組」？過去臺灣學術性的面訪調查多半以戶籍資料爲抽樣底冊，調查執

一、採用母體資料進行多重變數的事後分層加權；二、透過與機率樣本合併及輔助變數（auxiliary variable）來修正；三、透過模型輔以母體資料來修正；及四、採用樣本配對的方式，即從近似母體的大型資料庫中抽取樣本，繼而從網路調查資料庫中抽取樣本與之配對後，調查配對後的網路調查資料庫樣本（近期有關各種修正方法的討論，請參考 Dutwin and Buskirk, 2017；Elliott and Valliant, 2017；Mercer et al., 2017；Schonlau and Couper, 2017；俞振華，2013）。總之，上述四種方法各有利弊，採用何種方法還需取決於成本及資源。譬如以樣本配對的方式而言，除非有近似母體的大型資料庫可供抽樣，不然一般研究者仍無法採用該方式來修正網路調查結果。又譬如模型輔以母體資料的修正方式，其母體資料必須涵蓋諸多模型中所使用的變數，往往也不容易執行。由於許多實務上的限制，目前採用其他根據機率抽樣的調查資料（包括電話調查或面對面調查）來修正網路調查資料的作法，仍是最可行且普遍採用的方式。

5 以本研究來說，電訪樣本的回答率及涵蓋率仍比網路調查高出許多，有關本研究所使用之各波電話調查與網路調查的成功樣本數、接觸數、及調查成功率等詳細比較及說明請參見附錄 1 的附表 1。

臺灣大學學術期刊資料庫

行之前還會發函通知受訪者，於是一般而言，學術性面訪調查的涵蓋率最高，拒訪率最低（洪永泰，2005）。西方國家的學術性調查研究，也往往將面訪視爲是最嚴謹的調查方式。因此，前述如 Lee 和杜素豪等的研究皆將面訪調查結果視爲黃金標準，做爲調整網路或電訪調查結果的依據。至於目前最常見的家用電話調查，許多人認爲，在唯手機族大量增加，家用電話使用頻率大幅下降的今日，該調查方式的涵蓋率大有問題，實不應拿來做爲對照之用。不過，先前許多利用電訪資料來修正網路調查結果的試驗皆獲得不錯的成效：譬如，網路調查機構 Harris Interactive 即是以電訪樣本來修正網路調查結果（Terhanian and Bremer, 2000）。Schonlau et al. (2007) 及 Dutwin and Buskirk (2017) 也採用類似的方式，將所謂「Webographic」（即「網路使用者特徵」）的問題，同時置於網路調查及家用電話調查當中，用以辨識網路使用者與非使用者的不同，並藉入選機率調整法調整網路調查結果。總之，雖然家用電話調查的樣本代表性逐漸受到學者質疑，但該調查模式目前仍爲臺灣各類型民意調查的主流。畢竟和面訪調查相比，電話調查執行上較爲簡便，調查成本也較低廉，且其樣本涵蓋率及回答率普遍來說仍優於網路調查（Yeager et al., 2011）。因此，本研究仍將電訪調查結果列爲對照組（參考樣本），透過電訪資料及「Webographic」變數（或稱「網路使用者特徵」變數）來修正網路調查結果（觀察樣本）。[6]

至於「Webographic 變數」的問題主要包涵哪些？簡單來說，這些問題往往和調查研究的主旨無關，譬如一些態度（譬如是否喜歡學習新事物）、事實（譬如上個月是否旅行、是否有房子）、或隱私權（譬如是否應調查公司員工有無愛滋病）等問題，來區辨網路受訪者及電訪受訪者的差異。而造成兩種不同受訪者在回答這些問題時有所差異的因素很多，譬如：網路受訪者不需面對訪員，在回答敏感或個人問題時較不會顧慮「社會期許」（social desirability）的壓力，又或者網路受訪者的生活型態及態度本來就和電訪受訪者有本質上的差異，使得我們能透過問卷題目，掌握兩者

[6] 目前臺灣各主要民調機構（包括政治大學選舉研究中心）皆開始進行手機樣本調查，背後的原因即在於補強電話調查的涵蓋率。倘若日後電話調查（其中包括家用電話及手機樣本）將涵蓋率及回答率向上提升並接近面訪水準（即黃金標準），則以電話調查資料作爲「對照組」的方法仍不失爲修正網路調查的一種方式。

臺灣大學學術期刊資料庫

在不同生活面向上的差別。俞振華（2013）利用 Schonlau et al. (2007) 的變數設計，透過平行測試來探討臺灣網路及電訪受訪者的不同處，發現兩種受訪者在自我認知及行爲事實等方面，和美國的情況類似，都有顯著的差異。[7] 另外，該測試還發現，除了那些借自於 Schonlau et al. (2007) 的變數有助於區辨臺灣的網路及電訪受訪者外，生活作息上的不同（起床、就寢的時間），其實最能反應臺灣網路受訪者及電訪受訪者的差異，也最有助於我們利用電訪資料來校正網路調查結果。[8] 不過，該研究最大的限制，是只做了一次平行測試，且利用 Webographic 變數及入選機率調整法來修正網路調查結果這項方法是否穩健（robust），仍需要更深入及全面地討論。

參、資料及修正機制

本文使用的資料是「2012 年至 2016 年『臺灣選舉與民主化調查』（TEDS）4 年期研究規劃：總統滿意度電訪及網路調查案」。此計畫以長期追蹤民眾對總統施政滿意度以及對各項重要政治議題的看法爲研究主軸，以每季爲間隔，同時進行網路與電話調查。[9] 自 2012 年 9 月起至 2016 年 3 月止，意即從 2012 年馬英九總統第二個任期上任約百日後，到 2016 年 520 政權交接前，每隔 3 個月執行一次，共計累積 15 波調查。本文以此計畫的資料來探討網路調查結果修正機制，主要基於以下兩項優勢：(1) 此計畫的網路與電話調查皆在同一時間，使用同一套問卷，且由同一機構（政治大學

7 自我認知的題目包括以下三題：(1) 請問您會不會渴望學習到新的事物？(2) 請問您會不會常常感覺到孤獨？(3) 請問您遇到事情會不會常有「賭一把」的念頭？行爲事實的題目則有以下 4 題：(1) 最近 1 個月內，請問您有沒有旅行過？(2) 最近 1 個月內，請問您有沒有做運動？(3) 最近 1 個月內，請問您有沒有閱讀完 1 本書？(4) 請問您身邊（親朋好友）有沒有人是同性戀、雙性戀或變性者？上述變數當中，只有「有沒有做運動」這一項，網路受訪者及電訪受訪者之間沒有顯著的差異。

8 生活方式的題目，主要只是問受訪者的作息時間，包括請問您平常約幾點起床及幾點就寢？

9 TEDS 網路調查的樣本，完全來自於電訪調查。換言之，每波 TEDS 電話調查都會尋問受訪者是否願意加入 TEDS 網路調查受訪者名單。有關 TEDS 受訪者名單的建構方式，請參閱附錄 1 之附表 1 的說明或 TEDS2012-2016 PA 系列之結案報告書。

臺灣大學學術期刊資料庫

選舉研究中心）執行調查，使我們在進行兩種資料的比較分析時，不會受到因執行單位不同，所帶來可能之「機構效應」的影響；(2) 此計畫長期、多次的調查，有助於我們透過多次的分析，檢視本研究所採行的網路調查修正機制是否穩健。

本文針對 5 道政治評價題，進行修正分析比較，其中包括 3 題首長及政黨主席滿意度調查，及兩題經濟評估題（具體的問卷題目及題目代稱請見表 1）。會選擇這 5 個題目的主要原因，是因爲滿意度調查係坊間常見的問卷調查題目，許多媒體民調也幾乎每一季都做，較具實用性，也有利於我們日後使用外部資料（譬如 TVBS 電視台的總統滿意度調查），來檢驗本研究修正後的網路調查資料是否穩定。經濟評估同樣也是媒體時常報導的變數，除了與首長滿意度息息相關外，也是學界在探討經濟投票時，最常使用的變數。當然，TEDS 這一系列的電訪與網路調查還有其他諸多變數，譬如政治效能感或兩岸交流評價等。但由於多數題目並非在每一波調查中都有納入，爲了增加比較的延續性，本研究就以這 5 題爲例，利用電訪資料爲「對照組」，發展網路調查結果修正機制。

表 1　比較及修正對象：五項政治評價題

項目	問　卷　題　目	代　稱
1	首先請問您對馬英九這半年來擔任總統的整體表現滿意或不滿意？（二分變數）	（總統滿意度）
2	請問您對×××擔任行政院長以來的整體表現滿意或不滿意？（二分變數）	（行政院長滿意度）
3	請問您對×××擔任民進黨主席以來的整體表現滿不滿意？（二分變數）	（民進黨主席滿意度）
4	請問您覺得臺灣現在的經濟狀況與半年前相比，是比較好、比較不好，還是差不多？（三分變數）	（回顧經濟評估）
5	請問您覺得臺灣未來半年的經濟狀況是會變好、會變不好，還是差不多？（三分變數）	（前瞻經濟評估）

資料來源：「2012 年至 2016 年『臺灣選舉與民主化調查』四年期研究規劃（1-4）：總統滿意度電訪及網路調查案」（TEDS2012-2016 PA 系列）

臺灣大學學術期刊資料庫

表 1 當中，總統滿意度、回顧經濟評估和前瞻經濟評估，等 3 道題目的文字在這 15 波的調查中皆未改變。行政院長滿意度和民進黨主席滿意度則受到政治人物的輪替而有所變動，惟民進黨主席滿意度在第 15 波的問卷文字有所不同，其變動情況請見表 2。陳冲院長滿意度共執行 2 波調查，江宜樺院長滿意度執行 8 波調查，毛治國院長滿意度執行 4 波調查，張善政院長滿意度執行 1 波調查。蘇貞昌主席執行 8 波調查，蔡英文主席 7 波調查，不過最後一次的問卷並非詢問蔡英文擔任民進黨主席的滿意度，而是詢問「其當選總統後，新政府整體表現的滿意度」。

由於本文是以電訪資料來修正網路調查資料，等於是將每波平行調查中的電話調查結果作爲參考樣本。當然，電訪調查結果和母體資料（譬如普查資料）相比，也不是完全沒有偏誤。不過，網路調查結果的偏誤情況更嚴重，之前學者主張的「數位落差」問題在這一系列的網路調查資料中依舊存在，也使得我們更有理由利用電訪資料來修正網路調查結果。爲了節省正文篇幅，在此不再贅述網路調查結果的偏誤程度及「數位落差」的問題。[10]

表 2 行政院院長滿意度與民進黨主席滿意度問卷變動情況

行政院長	任職期間	執行調查時間
陳 沖	2012 年 02 月至 2013 年 02 月	第 01～02 波調查
江宜樺	2013 年 02 月至 2014 年 12 月	第 03～10 波調查*
毛治國	2014 年 12 月至 2016 年 02 月	第 11～14 波調查
張善政	2016 年 02 月至 2016 年 03 月	第 15 波調查
民進黨主席	**任職期間**	**執行調查時間**
蘇貞昌	2012 年 05 月至 2014 年 05 月	第 01～08 波調查
蔡英文	2014 年 05 月至 2016 年 05 月	第 09～14 波調查

第 15 次調查民進黨主席滿意度問卷題目：
請問您對蔡英文今年 1 月當選總統以來的整體表現滿不滿意？

註：*第 10 次調查時間爲 2014 年 12 月 11 日至 15 日
資料來源：同表 1

[10] 有關本研究所使用之各波電話調查與網路調查的電訪受訪者、網路受訪者、及母體人口三者間的特徵比對分析與討論請參見附錄 1（附圖 1-4 及其說明）。

臺灣大學學術期刊資料庫

本研究採用 Lee (2006) 的入選機率調整法（或稱傾向分數調整法），來調整網路調查資料。首先，我們先將電話調查樣本和網路調查樣本結合，並創建一個新的二分變數：0 爲電話調查受訪者，1 爲網路調查受訪者。其次，我們將此變數設爲依變數，再以 Webographic 變數、人口變數、和政治類型變數爲自變數建立「勝算對數模型（logit model）」。Webograhic 變數爲起床與就寢時間（24 小時制）；人口變數則爲性別、年齡（分組）、教育程度（分組）、職業（6 分類）、設籍地區（分組）；政治類型變數爲總統選舉投票與否（2012、2016）、政黨認同（有無）、族群認同、統獨立場（詳細說明可參考表 3）。換言之，我們透過人口變數、政治類型變數、及 Webographic 變數三個不同的視角，來區辨電訪樣本及網路樣本的差異。其中，過去臺灣的相關研究顯示，在諸多可能被用來當作 Webographic 變數中，「起床時間」及「就寢時間」比起其他西方發展的變數（譬如註 4 中所列的），有更好的預測能力（predict power），即更能夠有效地定義臺灣電訪及網路樣本間的差異（俞振華，2013）。[11]

然而，本研究在選定下述變數後，回過頭來檢測資料，發現以下兩個問題：(1) 有的調查沒有詢問就寢時間：2012 年 9 月第 2 次調查；(2) 有的調查有詢問設籍縣市，但沒有調查設籍鄉鎮，故無法作設籍地區分類：2013 年 3 月第 3 次調查、2014 年 3 月第 7 次調查、2014 年 6 月第 8 次調查。在排除這幾波調查後，本文僅使用 15 波總統滿意度調查中的第 1、4、5、6、9、10、11、12、13、14、及 15 波的資料，共 11 次，調查時間則分別爲 2012

[11] 究竟哪些變數適合用來估計「入選機率」（propensity score）？文獻中有不少爭論，主要聚焦在該類變數是否應該與待推估的變數相關（Austin, 2011）。有的主張該類變數與推估變數是否相關不是重點，甚至認爲應該不要相關，畢竟這類變數旨在「抓出」（capture）電訪受訪者與網路調查受訪者的差異處，並藉此來調整網路調查因涵蓋不足所產生的選樣偏誤（Schonlau et al., 2007）。但也有主張寧可把所有相關或沒相關的但同時在參考樣本及共變項都放進估計「入選機率」的模型當中，畢竟，任何遺漏變項（omitted variable）都會產生偏誤（Rubin and Thomas, 1996；Lee, 2006），而模型設定錯誤（model misspecification）的影響不大（Drake, 1993）。本研究後續分析採用杜素豪（2015：124）的建議，即「選擇與待推估之調查結果息息相關且參考樣本與觀察樣本中共同具有的共變項（社會人口與／或非社會人口特徵變項）」。我們曾測試過不同的模型設定，譬如只放入人口特徵及 webographic 變數，但修正的效果並不好，於是加入與推估變項相關的共變項後（即政治類型變數），修正效果才比較好。

臺灣大學學術期刊資訊庫

年 9 月、2013 年 6 月、2013 年 9 月、2013 年 12 月、2014 年 9 月、2014 年 12 月、2015 年 3 月、2015 年 6 月、2015 年 9 月、2015 年 12 月、2016 年 3 月。

表 3　PSA 模型變數編碼方式

變　　數	編　　碼　　方　　式
依變數	
調查類型	電話調查、網路調查
Webographic 變數	
起床時間	1～24 時
就寢時間	1～24 時
人口變數	
性別	男（參照組）、女
年齡	20～29 歲（參照組）、30～39 歲、40～49 歲、50～59 歲、60 歲以上
教育程度	小學及以下（參照組）、國中或初中、高中或高職、專科、大學及以上
職業	軍公教（參照組）、白領、藍領、農林漁牧、家管、學生及其他
設籍地區	大臺北都會區*（參照組）、新北市基隆、桃竹苗、中彰投、雲嘉南、高屏澎、宜花東
政治類變數	
總統選舉投票	有投票、無投票（參照組）
政黨認同	有政黨認同、無政黨認同（參照組）
族群認同	臺灣人（參照組）、中國人、雙重認同
統獨立場	偏向統一（參照組）、偏向獨立、維持現狀

註：*包含臺北市、新北市板橋區、新北市三重區、新北市中和區、新北市永和區、新北市新莊區、新北市新店區

資料來源：同表 1

在結合兩種資料及選定變數後，PSA 的操作包括以下 5 個步驟：(1) 建立「勝算對數模型」、(2) 計算預測機率、(3) 分組、(4) 計算修正權重（adjusted weight）、(5) 分組加權。[12] 其中第三步驟的分組數目和分組方法是本文想要

[12] 有關 PSA 的操作方法，主要分爲四種，包括配對（matching）、入選機率倒數（權重）法（inverse propensity scores）、多變量調整法（covariate adjustment）、及次樣本分組（subclassification 或 stratification）（杜素豪，2015；Austin, 2011）。本文採用的方法是次樣本分組，屬於多對一的配對方法，除了操作上較容易且有效率以外，也較廣爲被調查研究學界採用。此外，該方法較不受各種條件限制，舉例來說：和傳統配對方式相比，利用次樣本分組法時「配對」的預設條件少（譬如不用假定配對樣本是否有替

臺灣大學學術期刊資料庫

討論的重點。在分組方法上，目前的文獻多使用 Lee (2006) 的作法，將合併樣本的預測機率排序（rank）後均分成 5 組（下簡稱排序分組），再用人工調整的方式避免同樣的預測機率卻在不同組。這樣的做法看似合理，卻隱藏一個風險－即如何確認組內差異最小，組間差異最大？雖然可以事後使用箱型圖（box-plot）來了解預測機率分組的結構，但採用人工「目測」調整分組除了欠缺效率外，整個程序也欠缺方法上的理論依據。本文主張，爲了確認特定的排序分組是最適當的分組方式，我們可以直接使用較具理論依據的集群分析（cluster analysis）來進行分組。因此，我們將排序分組與集群分析中的平均數法（K-means）分組作比較，藉以探知哪一個方法較爲有效，即哪一個方法能使修正後的網路調查點估計值與電訪結果差距較小。

在分組數目方面，Lee (2006) 和 Lee and Valliant (2009) 在分組時提供的 5 組建議是根據 Cochran (1968) 和 Rosenbaum and Rubin (1984) 區分 4～5 組的研究結果，不過這兩篇文獻的建議是在探討生物實驗時所得出的結論，應用在網路調查研究時仍有疑義。杜素豪（2015）對此亦提出質疑，經過 4～10 組的分組修正比較後，發現在排序分組中，五分組的確是最爲有效的分組數目，因此本文不再探討排序分組的數目，而直接使用 5 組的排序分組和不同分組數目的集群分析作比較。

具體來說，修正步驟的第一步是利用上述表 3 的變數建構「勝算對數模型」。[13] 第二步是根據第一步的模型係數，估計依變數的預測機率，愈接近

代），比較不像入選機率倒數法容易受到極端值影響，或是不像多變量調整法那樣，容易受到推估變項的模型設定（model specification）所影響。此外，透過分組法操作 PSA 時，也便於納入所使用之樣本原本的權重（即原調查權數）（杜素豪，2015）。

13 有關模型的建構，匿名審查人提出了 common support 及平衡性檢定兩項問題。關於前者，本研究各波調查的重疊率在 88.47%～97.89 之間，還算不錯。至於是否應該刪除 off support 的樣本個案？Crump et al. (2009) 即曾建議，在進行傾向分數配對時，可以刪除傾向分數在 0.1 以下及 0.9 以上、或未重疊區域的觀察值，以提升配對後，處理效果（treatment effect）估計的準確性。不過，由於該作法只使用了有較高比例 common support 的個案，雖然提升了內在效度，但同時也犧牲了外在效度，畢竟，不論是刪了處理組（treatment group）或是控制組（control group）的個案，都會影響配對時的代表性，使得估計出的處理效果之外在效度打折扣。當然，若研究者的目的主要是探討接受處理者的平均處理效果（Average treatment effect on treated, ATT）時，採用此方法

臺灣大學學術期刊資料庫

0 代表該樣本愈可能來自於電話調查，愈接近 1 則代表該樣本愈可能來自於網路調查。第三步的分組，本文採用兩種方法，(1) 排序分組，將樣本的依變數預測機率從 0～1 排序後，等分成五組（c=5）。另外，爲了避免同樣的機率卻不同組，還需人工檢查和修正分組。譬如說，假設電話樣本共有 250 案，分 5 組，則每組有 50 案。若排序第 50 案和第 51 案預測機率相同，則第 51 案計入前一分組，以此類推。(2) 集群分析，使用平均數法，將樣本分別分成 3、5、7、9 組（c=3, 5, 7, 9），藉以探討集群分析時的適當組數，並測試是否會因爲組數過多而增加誤差，使得分組修正的效果下降。第四步，計算修正權重（或稱修正權數）。根據 Lee (2006: 333-335) 的推導：當電話樣本及網路樣本分別都有權重（如以下 (I) 式中的 d_K^R 與 d_j^W），且權重分別爲抽樣機率的倒數時，透過 (I) 式我們可以得出修正權重（f_C）：即以

對於提升內在效度或許有幫助。但本文主要是透過入選機率模型取得修正依據，藉以調整網路調查結果，而不是單純探討調查方式（mode effect）對於如總統滿意度等主要推估變項究竟有多大的影響。事實上，網路調查樣本及電話調查樣本極可能具有不同的樣本代表性，當我們因爲無法適當配對而將其中某些樣本個案刪除時，等於是只探討調查方式所帶來的差異（即使用兩種調查都有可適當配對的樣本個案），而沒有觸及代表性的差異。總之，由於本研究各波的 common support 或有不同，無法找出統一刪除樣本的準則，且本文強調透過傾向分數配對後，修正網路調查結果的外在效度，因此先不考慮刪除觀察值。關於後者，作者坦承本研究使用的各波資料，並非每波選擇模型中的共變項都通過平衡性檢定，且不同分組也有差異。不過，如同 Caliendo and Kopeinig (2008) 所強調的，選擇模型的共變項選擇主要還是以理論爲主，且 Augurzky and Schmidt (2001) 透過模擬分析後指出，選擇模型中所使用的共變項，要包括與選擇及待推估變項相關的共變項（譬如本研究中的 webographic 變項及人口特徵變項）、及只與推估變項相關的共變項（譬如本研究中的政治類型變項）。換言之，在各波不同分組平衡性檢定結果存在差異的情況下，本研究在建構模型時，共變項選擇並非只以平衡性爲依歸。Imai et al. (2008) 的分析即指出，利用各種平衡性檢定做爲決定選擇模型應該包括（或刪減）哪些共變項其實是一種謬誤（fallacy），即不論是從實務或理論的角度來檢視都不恰當。譬如實務上，當我們配合 common support 的要求刪除部份樣本個案後，共變項平衡總是比較容易達成（畢竟造成「不平衡」的樣本個案已被刪除），於是我們更可能納入沒有關聯性的共變項。此外在理論層次上，我們明明只是在配對前後，比較兩個樣本（譬如本文中的觀察樣本及參考樣本）的共變項特徵，但所做的平衡性檢定其實是在檢證：兩個樣本背後的母體是否存有差異。總之，單純以平衡性檢定做爲共變項的選擇依據，目前頗有爭議，且檢定的方法很多，採用不同的方法所得出的結果不見得有一致性。

臺灣大學學術期刊資料庫

經原電話樣本權重加權後的電話樣本之各組比例，來修正經由原網路樣本權重加權後的網路樣本中各組的比例。然後再透過 (II) 式，以修正權重調整原本網路調查樣本權重，即以第 c 組修正權重（f_C）乘以原網路調查樣本權重（d_j^W），得出網路樣本每一案的 PSA 加權權重（$d_j^{W.PSA}$）。本研究的網路調查樣本雖然是由電話調查搜集而來，然而各網路樣本之原電話調查時所使用的權重（即經人口變數反覆加權後所得之權重）並未記錄在網路受訪者資料庫當中，使得我們在使用網路樣本時，無法假定各案被抽中的機率等值於其在電話樣本時被抽中的機率。換言之，本研究在無法以原電話樣本權重做爲網路樣本權重的情況下，做爲初探研究，我們分別假定電訪樣本及網路樣本的權重相同（即未考慮加權的情況）。在這樣的假定下，Lee 建議以 (III) 式計算修正權重，即分別計算第 c 組的電話樣本數（n_c^R）除以電話樣本總數（n^R），和第 c 組網路樣本數（n_c^W）除以網路樣本總數（n^W），再將上述兩個數據相除後，可得第 c 組的修正權重（f_C）。第五步，由於假定網路調查樣本權重相同（$d_j^W = 1$），於是 (IV) 式中的 $d_j^{W,PSA}$ 直接以 f_c 代入。換言之，修正標的（y_i）每一分組的樣本經修正權重加權並加總後，除以所有分組樣本案數的總和，得出修正結果（$\hat{\bar{y}}^{W.PSA}$），即爲經使用電訪資料結合 PSA 修正後的網路調查結果。

$$f_c = \frac{\sum_{k\in(S_c^R)} d_k^R \Big/ \sum_{k\in(S^R)} d_k^R}{\sum_{j\in(S_c^W)} d_j^W \Big/ \sum_{j\in(S^T)} d_j^W} \equiv \frac{\hat{N}_C^R / \hat{N}^R}{\hat{N}_C^W / \hat{N}^W} \tag{I}$$

$$d_j^{W.PSA} = f_c \cdot d_j^W = \frac{\hat{N}_c^R / \hat{N}^R}{\hat{N}_c^W / \hat{N}^W} \cdot d_j^W \tag{II}$$

$$f_c = \frac{n_c^R / n^R}{n_c^W / n^W} \tag{III}$$

$$\hat{\bar{y}}^{W.PSA} = \frac{\sum_c \sum_{j\in(S_c^W)} d_j^{W.PSA} \cdot y_j}{\sum_c \sum_{j\in(S_c^W)} d_j^{W.PSA}} \tag{IV}$$

臺灣大學學術期刊資料庫

其中 S_c^R = 電話樣本（參考樣本）第 c 組，S_c^W = 網路樣本第 c 組，d_K^R = 原電話樣本權重（採相同權重），d_j^W = 原網路樣本權重（採相同權重），n = 樣本數。

肆、資料分析結果與比對

由於本文所使用的調查資料有 11 波，PSA 的操作分組方法也有 5 種，倘若將各波的修正結果製成表格，較不容易呈現並比較 PSA 修正效果。因此本文利用各波的誤差長條圖，輔以表格總結修正前後的平均點估計誤差值及誤差區間，藉以了解並比較三個二分變數（總統（馬英九）、行政院院長、及民進黨（反對黨）主席滿意度）及兩個三分變數（回顧經濟評估及前瞻經濟評估）的網路調查修正結果。[14] 其中圖型的部份是將各波調查結果，按兩項修正模式分成五欄子圖，包括 1 種排序分組（分成 5 組）及 4 種集群分組（分成 3、5、7、9 組）（X 軸上排），並依時序呈現各波修正前及修正後的網路調查結果與電訪結果相較之誤差長條圖，即以修正前與修正後的網路調查點估計值，分別減去電話調查點估計值得之，並以百分點表示。淺色的長條圖爲修正前的網路調查誤差，深色的長條圖爲修正後的網路調查修正誤差。另外爲了協助解讀，我們還將各圖加上 +/−3 個百分點的估計誤差虛線。

圖 1 顯示，透過排序分組和各種不同的集群分組方式修正後，總統滿意度和行政院院長滿意度的網路調查結果（觀察樣本）和本研究的「對照組」，即電話調查結果（參考樣本），差距都不大，但民進黨主席滿意度的修正結果則是差強人意，不論有無修正，多半是低估滿意度，只是修正後低估程度小一點罷了。顯然，網路調查系統性地低估了民進黨主席滿意度的趨勢。

在圖 1 所列的 11 波的調查結果中，總統滿意度和行政院院長滿意度的網路調查結果在修正前後之點估計誤差值約半數都落在 +/−3 個百分點內。

[14] 除了誤差值長條圖以外，我們還將電話調查結果、修正前的網路調查結果、及修正後的網路調查結果等 3 項測量值依時序製成了點估計折線圖，請參見附錄 2。

臺灣大學學術期刊資料庫

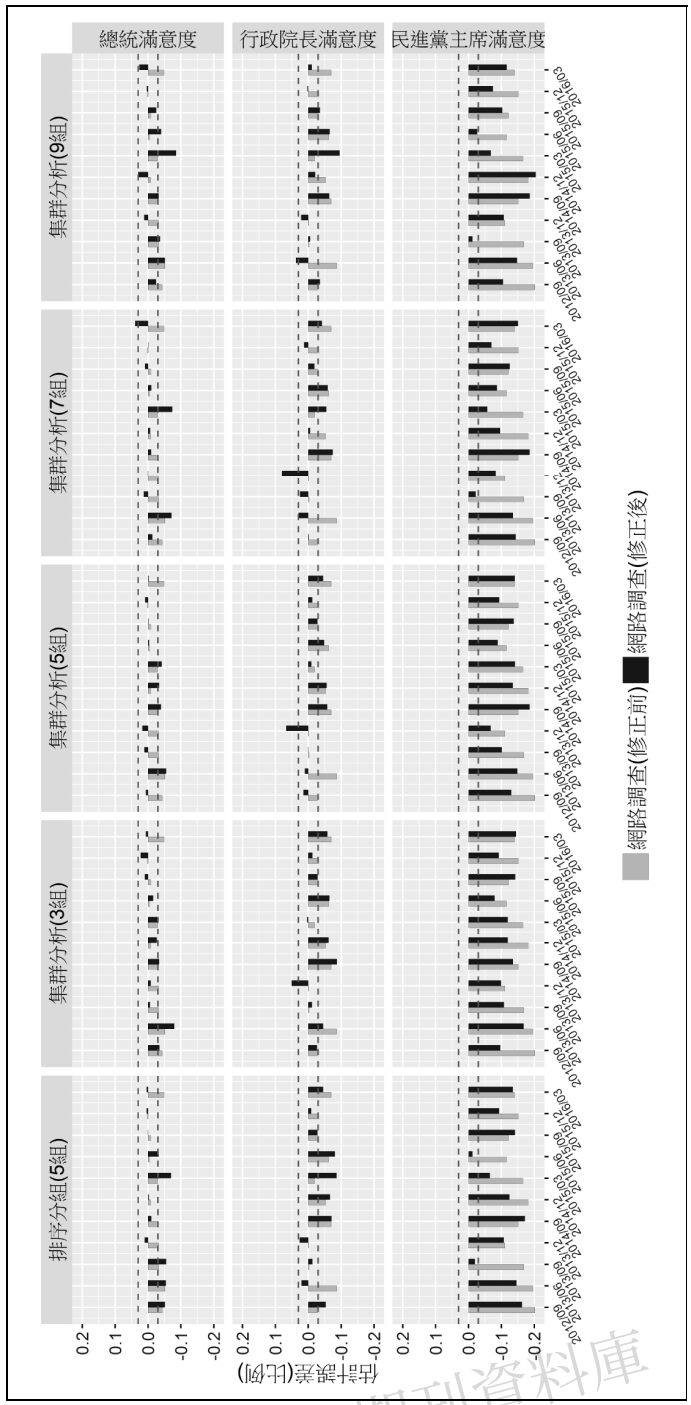
對照表 4 我們可以發現，排序分組和不同集群分析分組方式未見顯著差異，即同樣約半數修正後的結果落在 +/−3 個百分點內。不過，集群分析分組當中的 3、5、7 組似乎對於調整總統滿意度有較好的效果，調整完後的網路調查結果與電訪結果相比，約有 6 成（11 波當中的 7 波）甚至 7 成（11 波當中的 8 波）以上的點估計值誤差在 3 個百分點以內。另外，集群分析分組數目提升不見得能有效降低誤差值。譬如，分 9 組的調整結果其誤差值在 +/−3 區間的比例並沒有較其他較少分組數的調整結果來得更多。至於民進黨主席滿意度，修正後雖然如同表 4 所示，即網路調查的點估計誤差絕對值下降，但表 5 顯示，絕大多數（8 成以上）的誤差值仍超過 +/−3 個百分點的區間，顯示修正效果不佳。

表 5 則從 11 波點估計之平均誤差絕對值的角度，來比較不同修正方法的優越性。在未修正以前，總統滿意度、行政院長滿意度、及民進黨主席滿意度的網路調查估計值和電訪結果的誤差絕對值分別是 2.60、4.11 及 15.51 個百分點。透過排序分組的 PSA 模式修正後，總統及行政院院長滿意度的誤差絕對值並沒有顯著的縮小，只有民進黨主席滿意度的誤差絕對值縮小到 10.70 個百分點。而透過集群分析分組的數個 PSA 模式當中，分 5 組的方式對於縮小首長（包括總統及行政院院長）滿意度的誤差絕對值最有幫助，該方法使得兩項網路調查結果的平均誤差絕對值縮小到 1.99 及 3.17 個百分點。但 9 組的集群分析分組才使得民進黨主席滿意度的誤差絕對值縮到最小，但仍高達 10.46 個百分點。普遍來說，集群分析的分組方式較能縮小誤差，而排序分組的方式則效果有限。

至於經濟評估的問卷題型，圖 2 顯示，不論是透過排序分組或集群分組，雖然修正後的回顧經濟評估網路調查結果和電訪結果相比，誤差值都不大，但事實上，修正前的網路調查結果與電訪調查結果差距更小，意即修正後的結果不見得比修正前的結果還好。甚至，若網路調查修正前的點估計值與電話調查相近（譬如誤差值接近 0），則 PSA 修正後反而會將誤差值擴大。另外，針對前瞻經濟評估，圖 3 顯示的狀況和圖 2 類似，即各項修正方式似乎都未能將前瞻經濟評估的網路調查結果修正得更接近電話調查結果。整體來看，當測量的變數是 3 分類的類別變數時，使用 PSA 修正後的效果似乎有限。

臺灣大學學術期刊資料庫

圖 1 總統、行政院長、及民進黨主席滿意度修正誤差值比較



資料來源：「2012 年至 2016 年『臺灣選舉與民主化調查』四年期研究規劃（1-4）：總統滿意度電訪及網路調查案」（TEDS2012-2016 PA 系列）

臺灣大學學術期刊資料庫

表 4　首長及反對黨主席滿意度修正前後比較：誤差值區間 +/−3 個百分點

方法組數		總統滿意度		行政院院長滿意度		民進黨主席滿意度	
		範圍內	範圍外	範圍內	範圍外	範圍內	範圍外
未修正	-	6 （54.55）	5 （45.45）	5 （45.45）	6 （54.55）	0 （ 0.00）	11 （100.00）
排序	5	6 （54.55）	5 （45.45）	5 （45.45）	6 （54.55）	2 （18.18）	9 （ 81.82）
集群	3	8 （72.73）	3 （27.27）	5 （45.45）	6 （54.55）	0 （ 0.00）	11 （100.00）
	5	7 （63.64）	4 （36.36）	6 （54.55）	5 （45.45）	0 （ 0.00）	11 （100.00）
	7	8 （72.73）	3 （27.27）	6 （54.55）	5 （45.45）	1 （ 9.09）	10 （ 90.91）
	9	6 （54.55）	5 （45.45）	5 （45.45）	6 （54.55）	2 （18.18）	9 （ 81.82）

備註：1. 每種修正方法內含 11 波調查。
2. 括號內爲列百分比，範圍內與範圍外加總爲 100.00。
資料來源：同表 1

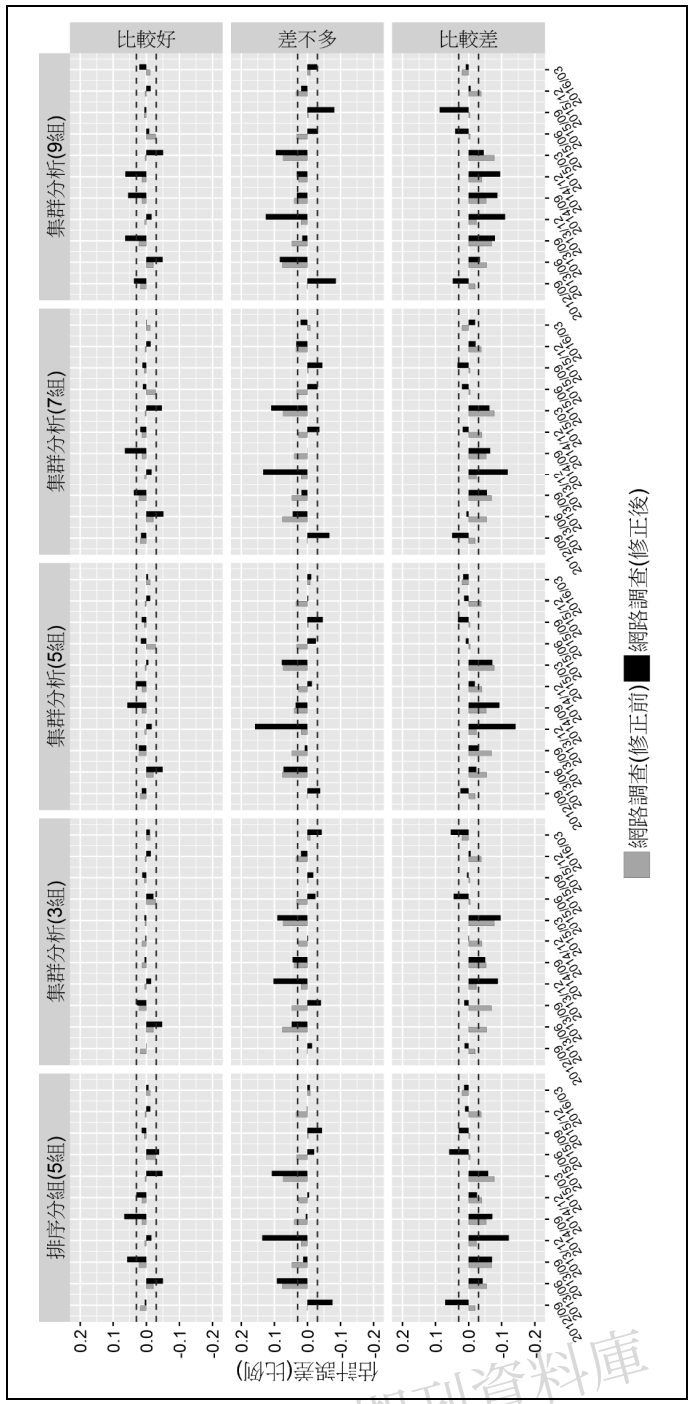
表 5　首長及反對黨主席滿意度修正前後比較：誤差絕對值（百分點）

方法組數		總統滿意度		行政院院長滿意度		民進黨主席滿意度	
		平均值	標準差	平均值	標準差	平均值	標準差
未修正	-	2.60	1.69	4.11	2.75	15.51	2.98
排序	5	2.66	2.49	4.53	2.65	10.70	5.20
集群	3	2.46	2.02	4.06	2.51	11.84	2.55
	5	1.99	1.81	3.17	2.24	12.45	3.20
	7	2.23	2.56	3.66	2.61	10.49	4.55
	9	3.32	2.05	3.57	2.74	10.46	5.69

資料來源：同表 1

臺灣大學學術期刊資料庫

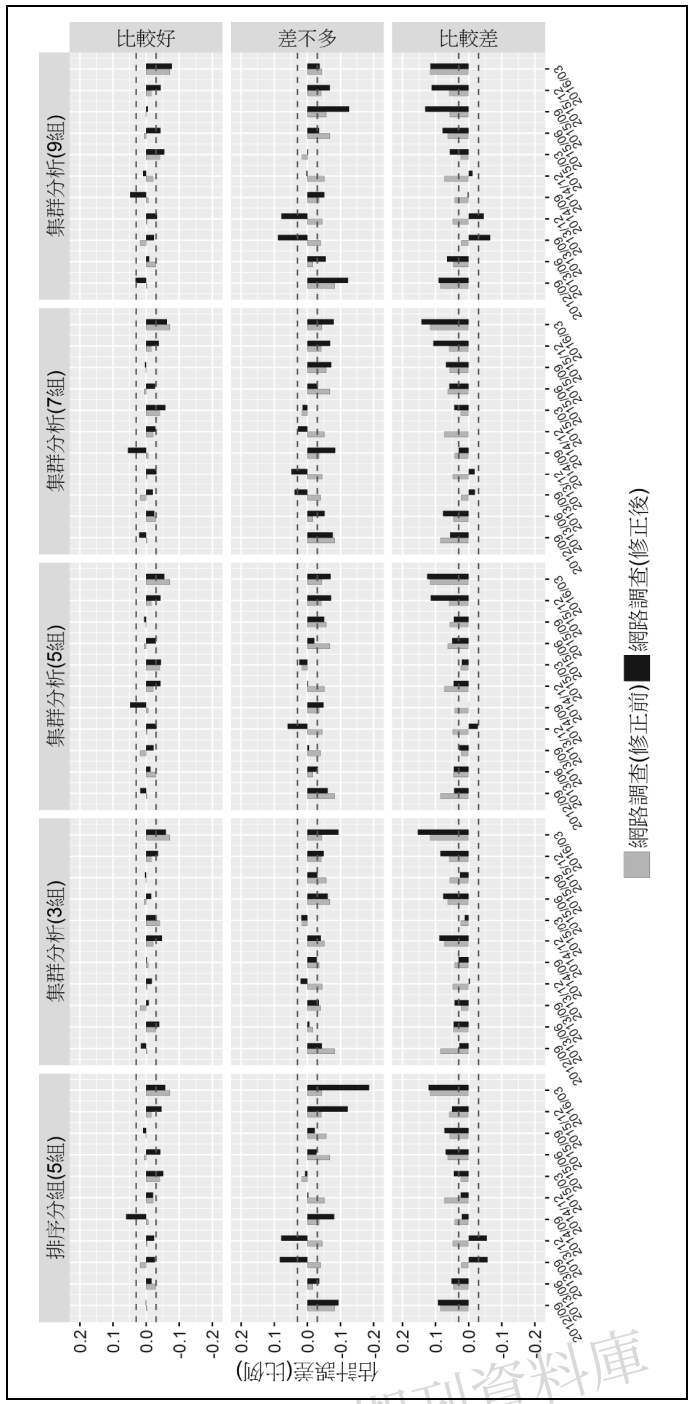
圖 2 回顧經濟評估修正誤差值比較



資料來源：同圖 1

臺灣大學學術期刊資料庫

圖 3　前瞻經濟評估修正誤差值比較



資料來源：同圖 1

臺灣大學學術期刊資料庫

表 6 我們透過誤差值區間，來評估哪一種修正方法較爲優越。若和未修正的結果比較，排序分組的修正結果與集群分析分組的結果似乎時好時壞，缺乏一致性。譬如：針對回顧經濟評估，11 波未修正前網路調查結果「比較好」的點估計值和電話調查結果的差距都在 +/−3 個百分點，但排序分組的修正結果則只有 6 波的點估計值在 +/−3 個百分點的誤差值區間內。不過，針對前瞻經濟評估，修正前只有 2 波未修正前網路調查結果「差不多」的點估計值和電話調查結果的差距在 +/−3 個百分點以內，但經過排序修正後，則有 4 波的點估計值在誤差區間內。不同的集群分析分組結果也同樣呈現類似不一致的情形，但相對來說，分 9 組的集群分析分組方式的修正效果最差。主要原因是：當某一選項中的某一分組內的網路觀察樣本數目趨近（或等於）0 時，該組一個樣本可能被加權至 10 倍以上，且其他選項的分組權重亦會被膨脹。換言之，當分組數過多而組內樣本數太少時，過多的權重反而導致修正結果不穩定。因此針對要「如何」選擇分組數目，日後或許應透過模擬研究，才能獲致更完善的準則。

如果我們只比較修正前後的誤差絕對值，則更突顯修正後的結果沒有比修正前還要好。透過表 7 我們發現，修正前的網路調查結果，不論是回顧還是前瞻經濟評估，其各選項的誤差絕對值都在 3.5 個百分點以內。但修正後的結果往往更糟，在幾個修正方法之間，集群分析 7 分組的結果相對好些，至少各個選項的誤差絕對值也都在 3.5 個百分點以下。總之，PSA 的方法在修正多分類別變數時，很可能因爲選項增加而使得各分組內樣本數減少，進而造成點估計值不穩定。換言之，我們需要多少的網路調查樣本數，才會使得修正後的點估計值較穩定？

臺灣大學學術期刊資料庫

表 6　回顧／前瞻經濟評估修正前後比較：誤差值區間 +/–3 個百分點

方法	組數	選項	回顧經濟評估		前瞻經濟評估	
			範圍內	範圍外	範圍內	範圍外
未修正	-	比較好	11 (100.00)	0 (0.00)	9 (81.82)	2 (18.18)
		差不多	5 (45.45)	6 (54.55)	2 (18.18)	9 (81.82)
		比較差	5 (45.45)	6 (54.55)	2 (18.18)	9 (81.82)
排序	5	比較好	6 (54.55)	5 (45.45)	6 (54.55)	5 (45.45)
		差不多	6 (54.55)	5 (45.45)	4 (36.36)	7 (63.64)
		比較差	4 (36.36)	7 (63.64)	2 (18.18)	9 (81.82)
集群	3	比較好	10 (90.91)	1 (9.09)	6 (54.55)	5 (45.45)
		差不多	5 (45.45)	6 (54.55)	4 (36.36)	7 (63.64)
		比較差	6 (54.55)	5 (45.45)	4 (36.36)	7 (63.64)
	5	比較好	8 (72.73)	3 (27.27)	5 (45.45)	6 (54.55)
		差不多	5 (45.45)	6 (54.55)	4 (36.36)	7 (63.64)
		比較差	6 (54.55)	5 (45.45)	4 (36.36)	7 (63.64)
	7	比較好	7 (63.64)	4 (36.36)	6 (54.55)	5 (45.45)
		差不多	3 (27.27)	8 (72.73)	2 (18.18)	9 (81.82)
		比較差	5 (45.45)	6 (54.55)	4 (36.36)	7 (63.64)
	9	比較好	5 (45.45)	6 (54.55)	4 (36.36)	7 (63.64)
		差不多	3 (27.27)	8 (72.73)	2 (18.18)	9 (81.82)
		比較差	2 (18.18)	9 (81.82)	2 (18.18)	9 (81.82)

備註：1. 每種修正方法內含 11 波調查。
2. 括號內爲列百分比，範圍內與範圍外加總爲 100.00。
資料來源：同表 1

臺灣大學學術期刊資料庫

表 7 回顧／前瞻經濟評估修正前後比較：誤差絕對值（百分點）

方法	組數	選項	回顧經濟評估		前瞻經濟評估	
			平均	標準差	平均	標準差
未修正	-	比較好	1.35	0.79	2.22	1.65
		差不多	3.49	2.15	1.96	2.05
		比較差	3.22	1.52	3.48	2.11
排序	5	比較好	4.05	3.01	4.89	3.83
		差不多	4.62	4.65	3.88	2.27
		比較差	5.41	2.25	6.73	5.29
集群	3	比較好	3.68	2.37	4.32	3.99
		差不多	5.84	3.38	5.75	2.58
		比較差	4.96	3.60	7.04	3.90
	5	比較好	3.26	2.44	4.44	4.32
		差不多	5.74	3.59	4.54	1.86
		比較差	4.06	2.39	6.07	3.98
	7	比較好	1.47	1.29	2.61	1.97
		差不多	3.13	2.12	2.51	1.81
		比較差	3.36	1.74	3.25	1.93
	9	比較好	3.39	3.36	4.27	3.07
		差不多	5.22	3.09	5.38	4.17
		比較差	5.63	3.99	6.00	2.75

資料來源：同表 1

臺灣大學學術期刊資料庫

伍、誤差與樣本數的關聯

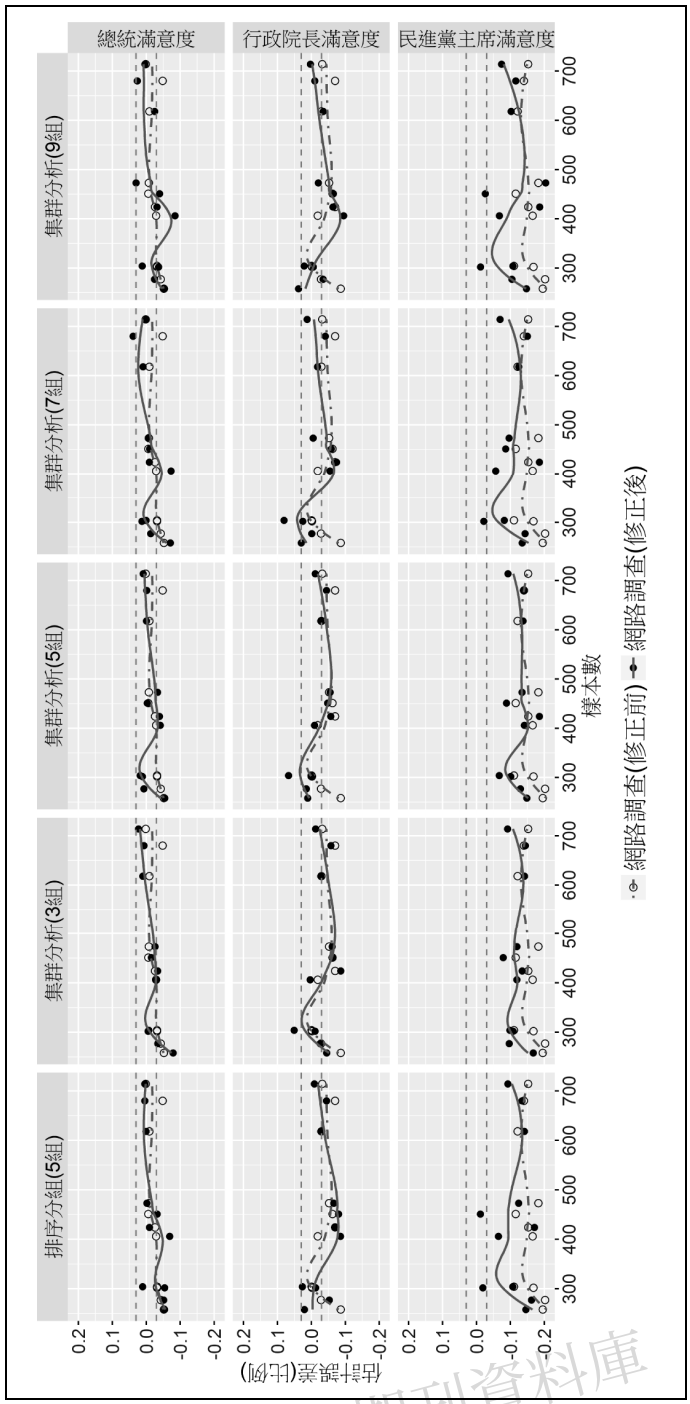
統計分析永遠不會嫌樣本數太多。我們可以想像，如果網路調查的樣本數增加，則透過 PSA 分組時，各組的樣本數自然也會增多，估計值也會相對穩定。本文分析的 11 波網路調查的樣本數並非一致，少則 258 案，多則爲 714 案（平均樣本數爲 446 案）。由於各波樣本數的差異，我們可以進一步分析：究竟網路調查的樣本數多寡是否影響修正前後的結果？是否樣本數增加能夠使得修正後的點估計值誤差縮小？圖 4 至圖 6 依序顯示透過不同修正方法，修正前及修正後所得出的首長及反對黨主席滿意度（包括總統滿意度、行政院院長滿意度、及民進黨主席滿意度）、回顧經濟評估、及前瞻經濟評估的網路調查點估計誤差與樣本數的散狀圖，輔以無母數統計中的 LOWESS 法作趨勢線，以及估計誤差值爲 +/−3 的水平線。其中曲折的虛線爲修正前網路調查點估計誤差與樣本數的趨勢線，另一個曲折的實線則爲修正後網路調查點估計誤差與樣本數的趨勢線。

圖 4 顯示，隨著網路調查樣本數增加，總統滿意度及行政院院長滿意度的修正誤差似乎有往 +/−3 個百分點區間中點收斂的趨勢（漸趨近 0），顯示樣本數增加或許有助於縮小誤差值，符合預期。不過，民進黨主席滿意度的修正誤差值顯然就未受到樣本數增加的影響，即不論樣本數多寡，網路調查的點估計值仍系統性地低於對照組的基準，並未朝 +/−3 個百分點的區間收斂。

針對回顧經濟評估的三個選項，圖 5 顯示，當網路調查樣本數增加時，只有「比較好」的修正誤差值在 +/−3 的區間內震盪，且有隨樣本數增加而收斂的趨勢，其他兩個選項的誤差值震盪幅度都很大，也看不太出收斂的趨勢。至於圖 6 的前瞻經濟評估，我們發現調查樣本數較多的三波調查，不論是哪個選項，其修正後的誤差值都反而更偏離 +/−3 的區間。總之，從這三個圖來看，本研究無法確認透過 PSA 修正後的網路調查結果，其誤差值一定會隨著樣本數增加而縮小。

臺灣大學學術期刊資料庫

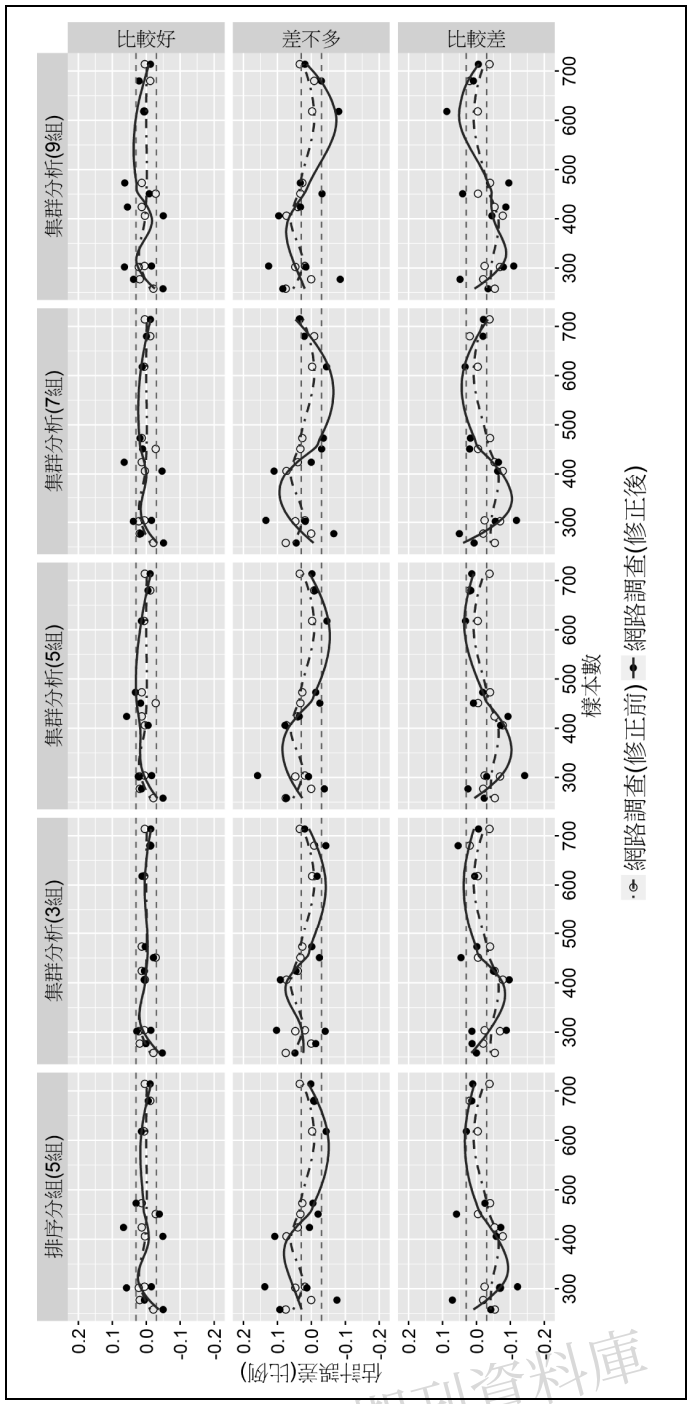
圖 4 總統、行政院長及民進黨主席滿意度修正前後與樣本數散佈圖



資料來源：同圖 1

臺灣大學學術期刊資料庫

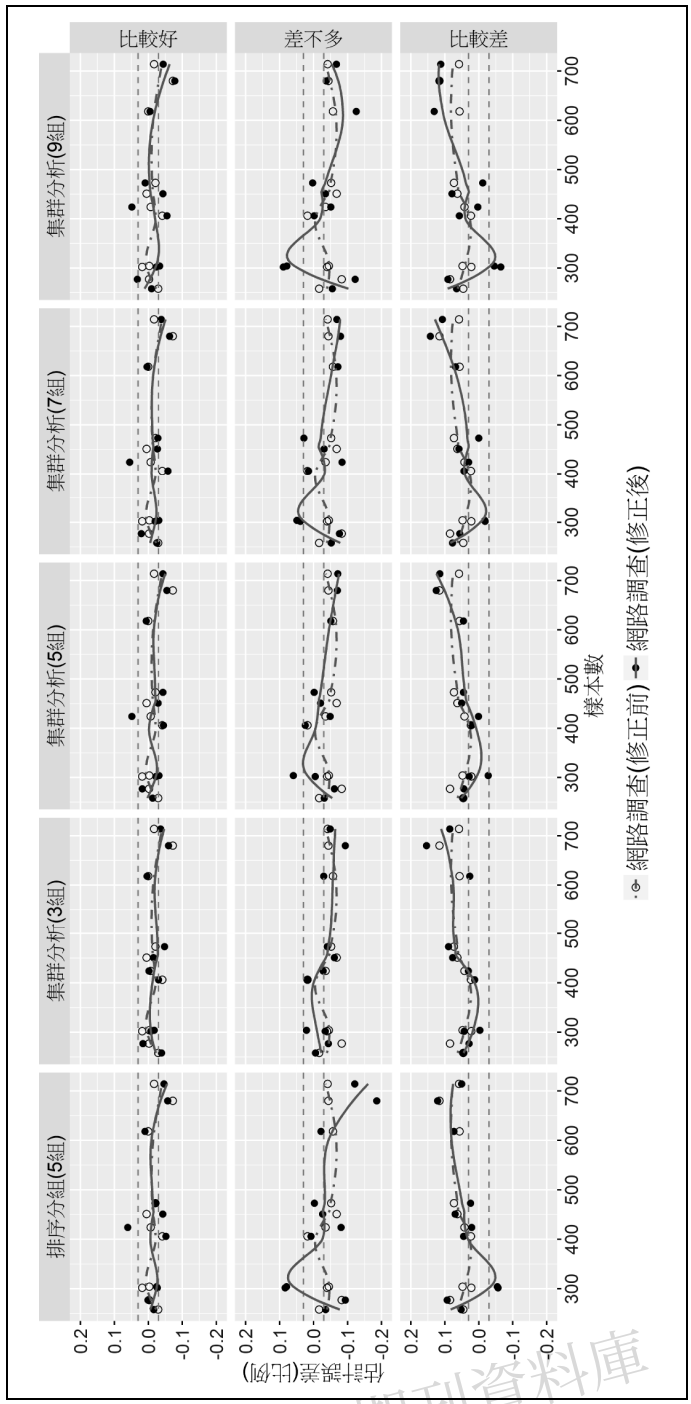
圖 5　回顧經濟評估修正前後與樣本數散佈圖



資料來源：同圖 1

臺灣大學學術期刊資料庫

圖 6 前瞻經濟評估修正前後與樣本數散佈圖



資料來源：同圖 1

臺灣大學學術期刊資料庫

陸、討論與建議

總結來看，針對坊間時常調查的首長（包括總統及行政院院長）滿意度和民進黨（反對黨）主席滿意度，本研究發現：在單純使用二分變數（滿意／不滿意）的情況下，經 PSA 修正後的首長滿意度網路調查結果，比修正前更接近電訪結果。不過，同樣的方法應用在民進黨（反對黨）主席滿意度時，修正結果就大打折扣－即修正後的誤差值雖比修正前小一些－但還是存在系統性地偏誤，即網路調查結果系統性地低估民進黨主席的滿意度。至於是什麼原因造成這樣的結果？我們沒有很明確的答案，可能的猜測是「題序效應」（question order effect）所產生的結果。[15] 不過，確切的因素還留待日後更多的測試及比較分析。

此外，針對學界或坊間調查都關心的整體經濟狀況評估問題，不論是「回顧經濟評估」或「前瞻經濟評估」，透過各種 PSA 方法修正後的結果都差強人意。具體來說，PSA 對於這兩個三分類選項題的修正效果並不穩定，時好時壞，又或對於特定選項（譬如「比較好」）修正效果較好，其他選項就較差。至於爲何 3 個選項的經濟評估題會出現較大的誤差？根據過往的

[15] TEDS 電訪及網路調查的問卷題目多半與總統及政府施政滿意度有關，除了總統整體滿意度外，還有許多總統分項施政滿意度的問題，包括滿意或不滿意總統在處理兩岸關係、外交方面、國防方面、民生問題、或促進經濟發展等諸多議題面向。另外，問卷也會接續詢問行政院長的施政滿意度及對執政黨主席的整體評價（當馬英九不兼任黨主席時）。於是，這一系列的題組通常要到最後一題，才會詢問受訪者對於民進黨主席的整體表現滿意不滿意。由於這些問題的題序是固定的，並沒有隨機出現，於是在自填的網路調查中，某些受訪者可能會根據初始的填答印象，一路針對馬政府的整體及個別項施政「點擊」不滿意，而沒有注意最後一題其實是在問反對黨主席的表現，和馬政府施政並沒有關係。換言之，民進黨主席滿意度在網路調查中被系統性地低估可能是肇因於所謂的「問卷題序效應」（question order effect）（Dillman, 2000）。Dillman et al. (2009: 160, 312) 比較自填問卷調查與依靠訪員完成的調查結果後發現，自填問卷較易出現「題序效應」，即自填問卷時，受訪者對於同類型題目較可能根據之前答題的印象一路填答下去，而不再仔細思考。當然，本研究網路調查的民進黨主席滿意度之所以被系統性地低估是否眞的是因爲「題序效應」所造成的，還需要更嚴謹的研究才能確定。不過，由於本研究使用的所有問卷，題序都是固定的，在已經無法操作題序做進一步比較分析的情況下，網路調查中「題序效應」較嚴重也只能是本文所提出的一項假設，留待日後學者進一步研究。

臺灣大學學術期刊資料庫

研究，可能的因素包括以下兩點：第一、回答選項過多；第二、提供中間選項。針對第一點，Holbrook et a1. (2006) 即認爲，過多的選項會干擾受訪者，不易區辨選項間的差異，進而增加受訪者答題時的困難。但以經濟評估題的 3 個選項來說，「變好」、「差不多」、及「變不好」三者間的意義有所區隔，造成受訪者辨識困難的程度應該有限。而針對第二點，提供「中間選項」往往使得受訪者將中間選項做爲「項目無反應」的選擇，或是利用中間選項敷衍回答（廖培珊，2010；Moors, 2007；Schuman and Presser, 1981）。因此，經濟評估題估計誤差較大的主要原因或許不是因選項數量上造成受訪者辨識的困擾，而是提供中間選項造成受訪者可以迴避答題。又，網路受訪者較電話受訪者較沒有機會選擇「無反應」（譬如不知道、看情形、無意見、拒答等）的選項。許多網路調查爲了避免受訪者隨便亂答或跳答，除了不提供無反應的選項以外，還往往強迫受訪者依題序填答，才可跳至下一題（或下一頁）。這樣的機制使得受訪者有可能爲了快速跳頁以填完問卷，乾脆隨便填答。於是，在沒有「無反應」選項的情況下，網路調查結果反而可能因強迫受訪者每一題皆要填答，而存在較大的估計誤差（Fricker et al., 2005）。[16]

以上分析顯示，雖然網路調查結果會因樣本自我選擇機制產生偏誤，但利用 Webographic 變數及傾向分數結合電話調查與網路調查結果後，某種程度的確可消除網路調查的偏誤。雖然如前註 4 所述，學界及調查實務界已發展出多種修正網路調查資料偏誤的方式，但事實上，學者仍無法評斷出「最好」的方法爲何，即沒有一種方式能夠全面地消除網路調查資料的偏誤（Dutwin and Buskirk, 2017）。以目前臺灣調查的實務來看，利用電話調查資料來修正網路調查資料仍是最可行的方式，畢竟臺灣缺乏個體層次的母體或是近似母體的大型資料庫，可供樣本配對或進行細緻的事後分層加權，於是只能以其他機率樣本所獲得的調查資料作爲「對照組」，藉以修正非機率樣本如網路調查所造成的偏誤。

不過，或許有人會問，如果電話調查結果是「對照組」，且都採用樣本

16 TEDS 網路調查除了開放題或自填題外，凡是選項題（包括總統滿意度、行政院長滿意度、反對黨主席滿意度、及經濟評估題）皆不提供「無反應」選項，並要求受訪者每題都要填答，不然無法跳頁。

臺灣大學學術期刊資料庫

代表性較高的電話調查探求民意了，那爲何還需要網路調查的資料？我們認爲，目前使用網路調查的主要功能還是在於節省成本及爭取時效。畢竟電話調查的金錢與時間成本還是高於網路調查，特別是爲了提高電訪涵蓋率而加上手機調查後，電話調查的成本又增加了許多。倘若日後我們只需要一半甚至更少的電話調查樣本作爲對照組（譬如，電訪樣本及網路調查樣本各 500 案），並利用網路調查及 PSA 就可以得出差不多的結果，則我們實在不需要像過去一樣，耗時耗工地進行大樣本的電話調查（譬如非得要完成 1,067 案不可）。

不過，透過 Webographic 變數結合網路與電話調查資料仍有其他成本需要計算在內：即爲了區辨網路及電話受訪者的不同，往往需要置入和調查主題完全不相關的題組。以一般電話調查問卷不超過 30 題爲基準，若置入大量的 Webographic 問題等於壓縮其它直接和主題相關的問卷題組空間，這還不包括基本人口特徵問題，於是使用 Webographic 途徑同樣耗費了相當的成本。本研究只使用人口變數加上作息時間（兩題 Webographic 生活方式題），就能夠很有效地區辨網路與電話調查受訪者，算是頗有效率的作法。

最後，針對結合電訪資料並透過 PSA 修正網路調查結果的機制，本研究除了比較不同模式及提出初步成果外，並認爲有以下 4 個研究方向，值得日後有興趣的研究者進一步探討：

第一、是否有更好的 Webographic 題目來協助我們更有效地區辨網路及電訪受訪者？該變數應該要和網路使用者的特徵及行爲相關，但和研究所欲探討的依變數（譬如首長滿意度、經濟評估等）無關。本文使用的兩個有關生活習慣的變數，表現僅差強人意，日後還有待更多的測試，以找出更好、更適合臺灣情況的 Webographic 變數。

第二、入選機率調整法還有其他調整方法，譬如配對或入選機率值倒數法等。但本研究僅使用較容易操作的次樣本分組法。其他方法是否能夠獲得更好的調整結果，或在何種情況下較適用哪一種方法（譬如參考或觀察樣本數增加），顯然還有很多值得討論的空間。另外，本研究並未將原電話調查及原網路調查的權重納入考量，而採取未加權的樣本並假定網路調查樣本權重相同的情況下，計算網路樣本之 PSA 加權權重。根據 Valliant and Dever (2011) 的模擬分析，納入網路調查原樣本權重的 PSA 修正結果會比假

臺灣大學學術期刊資料庫

定網路調查權重相同時來得好。未來若有較大的網路調查樣本，或許可採用譬如 Lee (2006) 的作法：即使用超過 1,600 案甚至 8,000 案以上的網路調查樣本，並將這樣的網路調查樣本視爲「擬母體」進行抽樣，且透過人口變數分層得出各層之抽樣機率，做爲原網路調查樣本權重的基準。

第三、本研究雖然同時考量了排序分組及集群分析分組，不過，兩者的表現並沒有系統性的差別。我們大概只能確定，在分組內樣本數不足的情況之下，過多的組數不見得會使修正效果更好。但倘若樣本數增加，則在進行分組時，自然較能夠增加組數，畢竟「無法配對」（即同組內缺少參考或觀察樣本）的可能性降低。而過去的研究顯示，增加組數將有利於減少偏誤，儘管增加組數的邊際效用會隨組數增加而遞減（Huppler and Louis, 2002）。總之，增加樣本數肯定更有利於應用次樣本分組法來修正觀察樣本的偏誤，但要增加到多少，則有待進一步探討與分析。[17]

第四、本研究只聚焦在點估計值的比較，未來的研究還應增加不確定性的測量（譬如透過模擬），藉以協助我們評估估計值的準確性。

參考書目

一、中文部分

杜素豪，2015，〈比較入選機率分組與其他加權方法對電話調查樣本的調整：上網率的推估〉，《臺灣社會學刊》，56: 115-150。

杜素豪、羅婉云、洪永泰，2009，〈以入選機率調整法修正調查推估偏差的成效評估〉，《政治科學論叢》，41: 151-176。

俞振華，2013，〈網路民意調查的理論與實務〉，陳陸輝（編），《民意調查新論》，臺北：五南，頁 89-110。

洪永泰，1996，〈抽樣調查中樣本代表性問題〉，《調查研究》，1: 7-37。

[17] 本研究的網路調查樣本是採用電話調查所搜集而來的機率（或準機率）樣本，如果要將本研究所使用的方法推廣到一般非機率樣本時，最可能面臨到的問題，就是在使用 PSA 分組後，有些組別可能因爲「數位落差」的關係，根本沒網路調查樣本，或網路調查樣本數非常少。但如果非機率網路調查樣本數較大時（例：超過 2,000 案），或許較能克服這樣的問題。譬如，較大的網路調查樣本總是會有一些老年、教育程度低、且居住在鄉村的受訪者，這樣才可能減少網路樣本和電訪樣本配對時，網路調查無樣本可配對的可能性。

臺灣大學學術期刊資料庫

洪永泰，2005，〈台灣地區抽樣調查各種母體定義、抽樣底冊、和涵蓋率的比較〉，《調查研究－方法與應用》，18: 9-44。

洪永泰，2009，〈抽樣〉，游清鑫（編），《民意調查新論》，臺北：五南，頁 111-156。

陳肇男，2001，〈大型電訪之省思〉，《調查研究》，10: 121-138。

游清鑫、黃紀、洪永泰、蔡佳泓，2007，《建置實驗性網路民意調查系統之規劃》（計畫編號：RDEC-095-06），臺北，行政院研究發展考核委員會。

鄭天澤、楊亨利、陳麗霞，2015，《2015 年台灣寬頻網路使用調查報告》，臺北：財團法人台灣網路資訊中心。

翟本瑞，2002，〈台灣網路市場調查公司的終結〉，《網路社會通訊期刊》第 27 期網頁，http://www.nhu.edu.tw/~society/e-j/27/social/27-17.htm，2015/12/15。

廖培珊，2010，〈態度量表之選項標示語：調查資料之潛藏類別分析〉，《調查研究－方法與應用》，24: 91-134。

二、英文部分

Augurzky, B. and C. M. Schmidt. 2001. *The Propensity Score: A Means to An End* (*Discussion paper series No.271*). Bonn: Institute of Labor Economics (IZA).

Austin, P. C. 2011. “An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies.” *Multivariate Behavioral Research* 46(3): 399-424.

Baker, R., S. J. Blumberg, J. M. Brick, M. P. Couper, M. Courtright, J. M. Dennis, D. Dillman, M. R. Frankel, P. Garland, R. M. Groves, C.Kennedy, J. Krosnick, P.J. Lavrakas, S. Lee, M. Link, L. Piekarski, K. Rao, R. K. Thomas, and D. Zahs. 2010. “Research Synthesis: AAPOR Report on Online Panels.” *Public Opinion Quarterly* 74(4): 711-781.

Caliendo, M. and S. Kopeinig. 2008. “Some Practical Guidance for the Implementation of Propensity Score Matching.” *Journal of Economic Surveys* 22(1): 31-72.

Cochran, W. G. 1968. “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies.” *Biometrics* 24(2): 295-313.

Couper, M. P. 2000. “Web Surveys: A Review of Issues and Approaches.” *Public Opinion Quarterly* 64(4): 464-494.

Couper, M. P., A. Kapetyn, M. Schonlau, and J. Winter. 2007. “Noncoverage and Nonresponse in an Internet Survey.” *Social Science Research* 36(1): 131-148.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik. 2009. “Dealing with Limited Overlap in Estimation of Average Treatment Effects.” *Biometrika* 96(1): 187-199.

臺灣大學學術期刊資料庫

Dever, J., A. Rafferty, and R. Valliant. 2008. "Internet Surveys: Can Statistical Adjustments Eliminate Coverage Bias?" *Survey Research Methods* 2: 47-60.

Dillman, D. A. 2000. *Mail and Internet Surveys: The Tailored Design Method.* 2nd ed. New York: Wiley.

Dillman, D. A., J. D. Smyth, and L. M. Christian. 2009. *Internet, Mail, and Mixed-Mode*. Hobiken: John Wisley & Sons.

Drake, C. 1993. "Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect." *Biometrics* 49(4): 1231-1236.

Dutwin, D. and T. D. Buskirk. 2017. "Apples to Oranges or Gala Versus Golden Delicious? Comparing Data Quality of Nonprobability Internet Samples to Low Response Rate Probability Samples." *Public Opinion Quarterly* 81(s1): 213-249.

Elliot, M. R. and R. Valliant. 2017. "Inference for Nonprobability Samples." *Statistical Science* 32(2): 249-264.

Fricker, S., M. Galesic, R. Tourangeau, and T. Yan. 2005. "An Experimental Comparison of Web and Telephone Surveys." *Public Opinion Quarterly* 69(3): 370-392.

GfK, 2012. "GfK Bibliography: Articles and Presentations based on GfK's Collected Panel—Data Analysis, or Methodology." in http://www.knowledgenetworks.com/ganp/docs/KN-Bibliography.pdf. Last update 10 April 2016.

Hewson, C., P. Yule, D. Laurent, and C. Vogel. 2003. *Internet Research Methods: A Practical Guide for the Social and Behavioral Sciences*. London: Sage.

Holbrook, A. L., Y. IK Cho, and T. Johnson. 2006. "The Impact of Question and Respondent Characteristics on Comprehension and Mapping Difficulties." *Public Opinion Quarterly* 70(4): 565-595.

Huppler H. K. and T. A. Louis. 2002. "Propensity Score Modeling Strategies for the Causal Analysis of Observational Data." *Biostatistics* 3: 179-193.

Imai, K., G. King, and E. Stuart. 2008. "Misunderstandings between Experimentalists and Observationalists about Causal Inference." *Journal of the Royal Statistical Society Series A* 171(2): 481-502.

Lee, S. and R. Valliant. 2009. "Estimation for Volunteer Panel Web Survey Using Propensity Score Adjustment and Calibration Adjustment." *Sociological Methods & Research* 37(3): 319-343.

Lee, S. 2004. *Statistical Estimation Methods in Volunteer Panel Web Surveys*. Doctoral dissertation, Joint Program in Survey Methodology, University of Maryland, College Park, MD.

Lee, S. 2006. "Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys." *Journal of Official Statistics* 22(2): 329-349.

Loosveldt, G. and N. Sonck. 2008. "An Evaluation of the Weighting Procedures for an Online Access Panel Survey." *Survey Research Methods* 2(2): 93-105.

Mercer, A. W., F. Kreuter, S. Keeter, and E. A. Staurt. 2017. "Theory and Practice in Nonprobability Surveys: Parallels between Casual Inference and Survey

臺灣大學學術期刊資料庫

Inference." *Public Opinion Quarterly* 81(s1): 250-279.
Moors, G. 2007. "Exploring the Effect of a Middle Response Category on Response Style in Attitude Measurement." Quality & Quantity 42(6): 779-794.
Nie, N. and L. Erbring. 2000. *Internet and Society: A Preliminary Report*. Stanford: Stanford Institute for the Quantitative Study of Society.
Norris, P. 2001. *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. Cambridge, MA: Cambridge University Press.
Rivers, D. 2007. "Sampling for Web Surveys." Paper prepared at the 2007 Joint Statistical Meeting, August 1, Salt Lake City, UT.
Rosenbaum, P. R. and D. B. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *Journal of the American Statistical Association* 79(387): 516-524.
Rubin, D. B. and N. Thomas. 1996. "Matching Using Estimated Propensity Score: Relating Theory to Practice." *Biometrics* 52(1): 254-268.
Schonlau, M. and M. P. Couper. 2017. "Options for Conducting Web Surveys." *Statistical Science* 32(2): 279-292.
Schonlau, M., A. van Soest, and A. Kapteyn. 2007. "Are 'Webographic' or Attitudinal Questions Useful for Adjusting Estimates from Web Surveys Using Propensity Scoring?" *Survey Research Methods* 1(3): 155-163.
Schonlau, M., K. Zapert, L. P. Simon, K. H. Sansted, S. M. Marcus, J. Adams, H. Kan, R. Turner, and S. H. Berry. 2004. "A Comparison between Responses from a Propensity-Weighted Web Survey and Identical RDD Survey." *Social Science Computer Review* 22: 128-138.
Schuman, H. and S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. New York: Academic Press.
Terhanian, G. and J. Bremer. 2000. *Confronting the Selection-bias and Learning Effects Problems Associated with Internet Research*. Rochester, NY: Harris Interactive *White Paper*.
Valliant, R. and J. A. Dever. 2011. "Estimating Propensity Adjustments for Volunteer Web Surveys." *Sociological Methods and Research* 40: 105-137.
Yeager, D. S., J. A. Krosnick, L. Chang, H. S. Javitz, M. S. Levendusky, A. Simpser, and R. Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability Samples." *Public Opinion Quarterly* 75(4): 709-747.

臺灣大學學術期刊資料庫

附錄 1
各波調查結果基本資訊及網路受訪者、電話受訪者、與人口統計資料比較（人口特徵）

附表 1 各波調查執行結果基本資訊

調查時間	電話調查			網路調查		
	成功樣本	有效接觸	成功率	成功樣本	邀請函	有效填答率
2012 年 09 月	1,084	3,525	30.75%	332	7,790	4.26%
2012 年 12 月	1,140	3,953	28.84%	139	4,424	3.14%
2013 年 03 月	1,101	4,248	25.92%	359	4,727	7.59%
2013 年 06 月	1,072	3,439	31.17%	279	4,951	5.64%
2013 年 09 月	1,075	4,149	25.91%	320	5,277	6.06%
2013 年 12 月	1,081	4,259	25.38%	318	5,503	5.78%
2014 年 03 月	1,101	4,130	26.66%	389	6,623	5.87%
2014 年 06 月	1,076	3,857	27.90%	475	6,900	6.88%
2014 年 09 月	1,086	3,963	27.40%	467	7,160	6.52%
2014 年 12 月	1,086	3,541	30.67%	509	8,207	6.20%
2015 年 03 月	1,083	3,594	30.13%	447	8,205	5.45%
2015 年 06 月	1,076	4,298	25.03%	495	8,711	5.68%
2015 年 09 月	1,076	4,222	25.49%	673	10,232	6.58%
2015 年 12 月	1,100	4,319	25.47%	783	10,602	7.39%
2016 年 03 月	1,099	4,638	23.70%	732	10,824	6.76%

註：012 年 12 月、2013 年 03 月、2014 年 03 月、2014 年 06 月等 4 波調查（灰底）未納入分析。
資料來源：「2012 年至 2016 年『臺灣選舉與民主化調查』四年期研究規劃（1-4）：總統滿意度電訪及網路調查案」（TEDS2012-2016 PA 系列）

調查樣本說明：（請參閱 TEDS2012-2016-PA 系列各年度結案報告）

一、電話調查的樣本主要有兩種，一種是以調查當年度的「中華電信住宅部電話號碼簿」爲母體清冊，依據各縣市電話簿所刊電話數佔臺灣地區所刊電話總數比例，決定各縣市抽出之電話數比例，以等距抽樣法抽出各縣市電話樣本後，爲求涵蓋的完整性，再以隨機亂數修正電話號碼的最後二碼或四碼，以求接觸到未登錄電話的住宅戶。另一部份的電話樣本則是來自政治大學選舉研究中心所

臺灣大學學術期刊資料庫

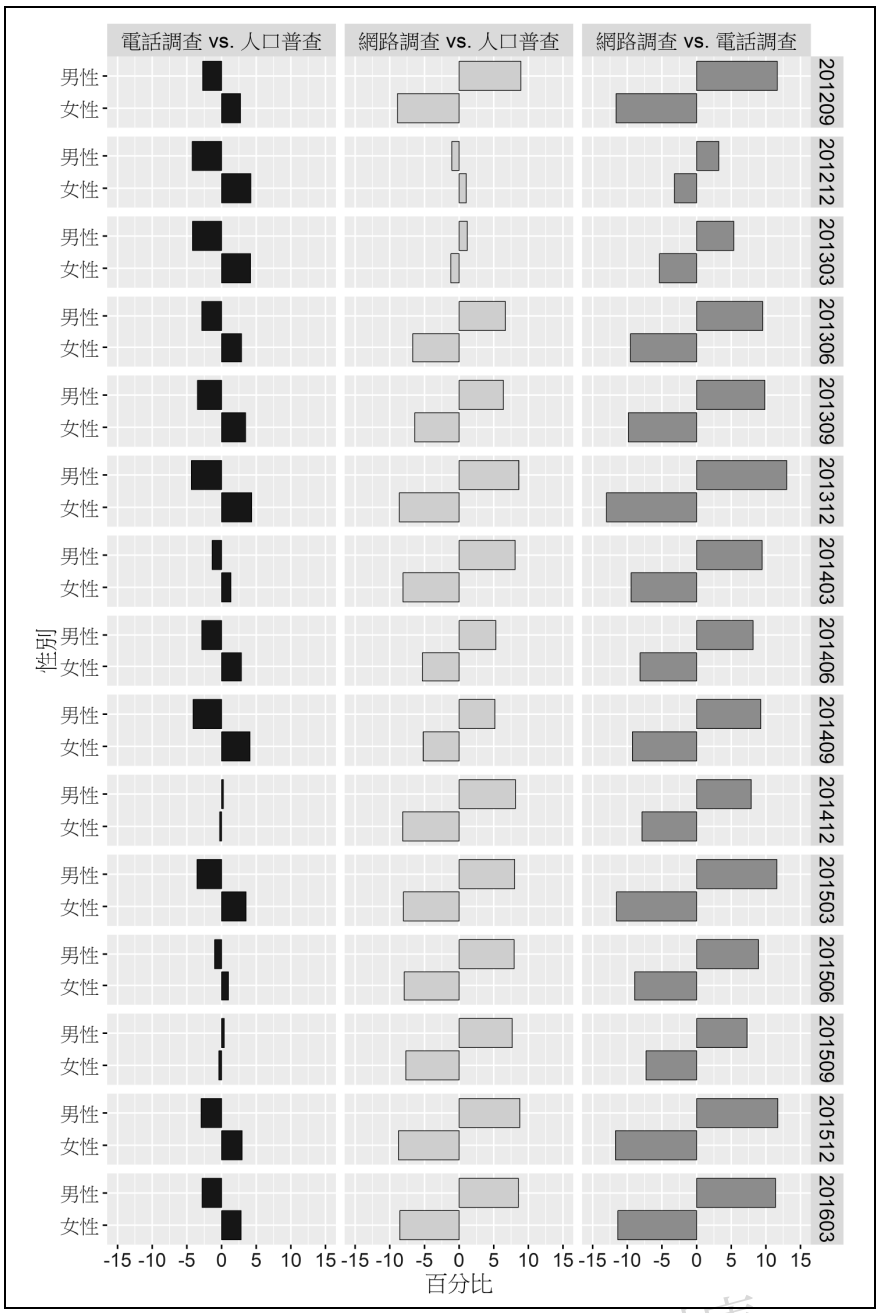
累積的電訪資料庫，先隨機抽出局碼後，再以隨機亂數產生電話號碼的最後四碼來製作電話樣本。在接通電話訪問開始之前，訪員按照（洪氏）戶中抽樣的原則（洪永泰，1996），抽出該戶中應受訪的對象再進行訪問。

二、在網路調查部份，第一次樣本主要是來自政治大學選舉研究中心的「線上調查實驗室」的受訪者名單，其餘各次樣本則來自 TEDS-PA 短卷電話訪問所蒐集之電子信箱資料組成受訪者名單。

從上述說明來看，本研究的電話調查抽樣方式已盡可能擴大樣本涵蓋率，除了「唯手機族」外，理論上只要有家用電話者，就有相同的機會被抽到。根據許勝懋（2015：45-46）分析 TEDS 2012 及 2013 年面訪調查資料後指出，當時臺灣地區 20 歲以上民眾的「唯手機族」比例大約是 6.2%，若再「加上沒家用電話也沒有手機」的 0.4%，家用電話的涵蓋率應可達到 93.4%。當然，這是在採取採取戶中選樣、多時段接觸、多次接觸、多語言接觸下所可能達到的理想狀態（該比例很接近家庭收支調查住宅電話普及率的 94.8%）。而稍早洪永泰等（2014）分析 2009 年國民健康調查資料後即發現，「唯手機族」的比例大約佔臺灣地區 18 歲以上民眾的 9.3%。換言之，「唯手機族」在本研究首先幾波的調查當中，對於電訪樣本涵蓋率的影響並不會很大。然而的確，過去幾年來，「唯手機族」比例肯定大幅成長，於是我們看到近來各主要民調機構也開始將「唯手機族」納入電話調查當中。但就算唯手機族成長到 25%，本研究中電話調查的涵蓋率還是不會比網路調查還差。畢竟，本研究的網路調查受訪者名單皆來自於家用電話調查樣本，所以該名單可被視爲是家用電訪樣本的「次級樣本」，而每個家用電訪成功樣本願意提供其電子郵件信箱參與網路調查的比率約 1/5～1/4，即若電訪樣本成功 1,000 個，所甄補到的網路調查受訪者也不過 200 個到 250 個（俞振華，2013），而其中眞正願意接受網路調查的比例也不過介於 3.14%～7.59%（網路調查成功率）。換言之，網路調查受訪者名單只是電話調查的「部份樣本」，而眞正受訪的又是部份樣本當中的少數。於是在這樣的結構下，本研究中的電訪樣本涵蓋率還是會比網路調查樣本好些，這也是本研究以電話調查樣本當作「對照組」的主要原因—即或許本研究的電訪樣本涵蓋率沒有一般大型面訪樣本的涵蓋率好，但肯定還是比網路調查的涵蓋率要好。

臺灣大學學術期刊資料庫

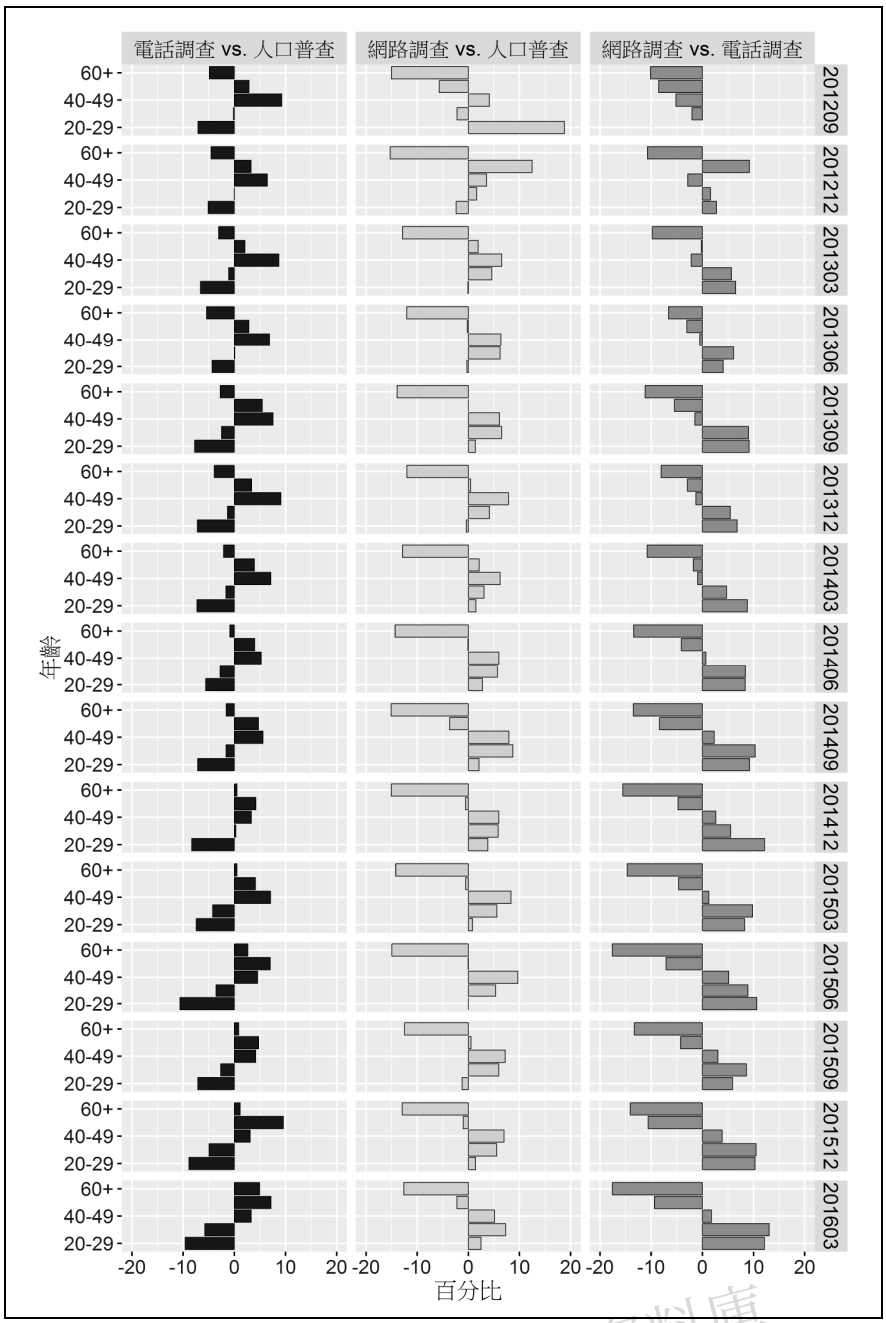
附圖 1 網路、電話調查與人口普查交叉比較「性別」比例差異



資料來源：「2012 年至 2016 年『臺灣選舉與民主化調查』四年期研究規劃（1-4）：總統滿意度電訪及網路調查案」（TEDS2012-2016 PA 系列）

臺灣大學學術期刊資料庫

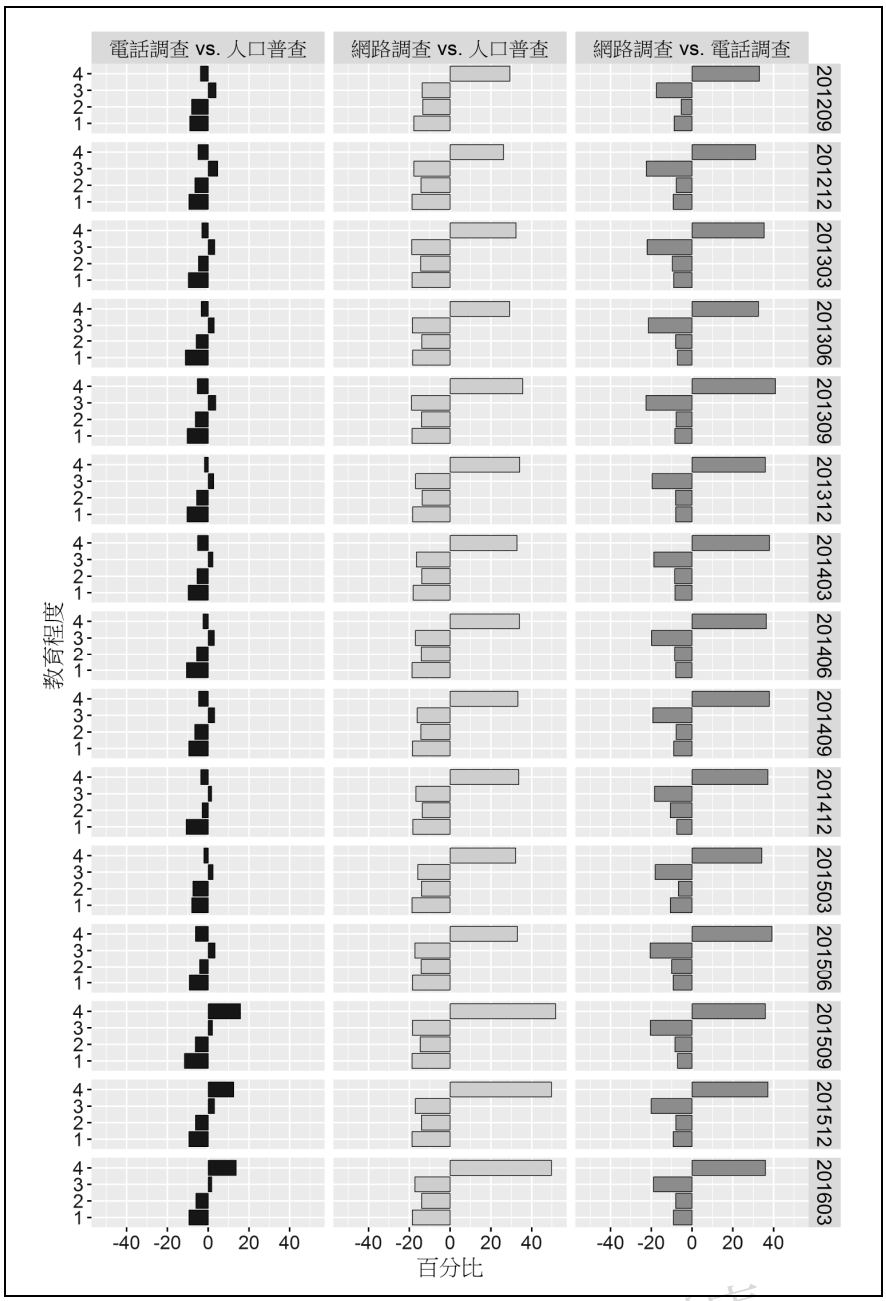
附圖 2　網路、電話調查與人口普查交叉比較「年齡」比例差異



資料來源：同附圖 1

臺灣大學學術期刊資料庫

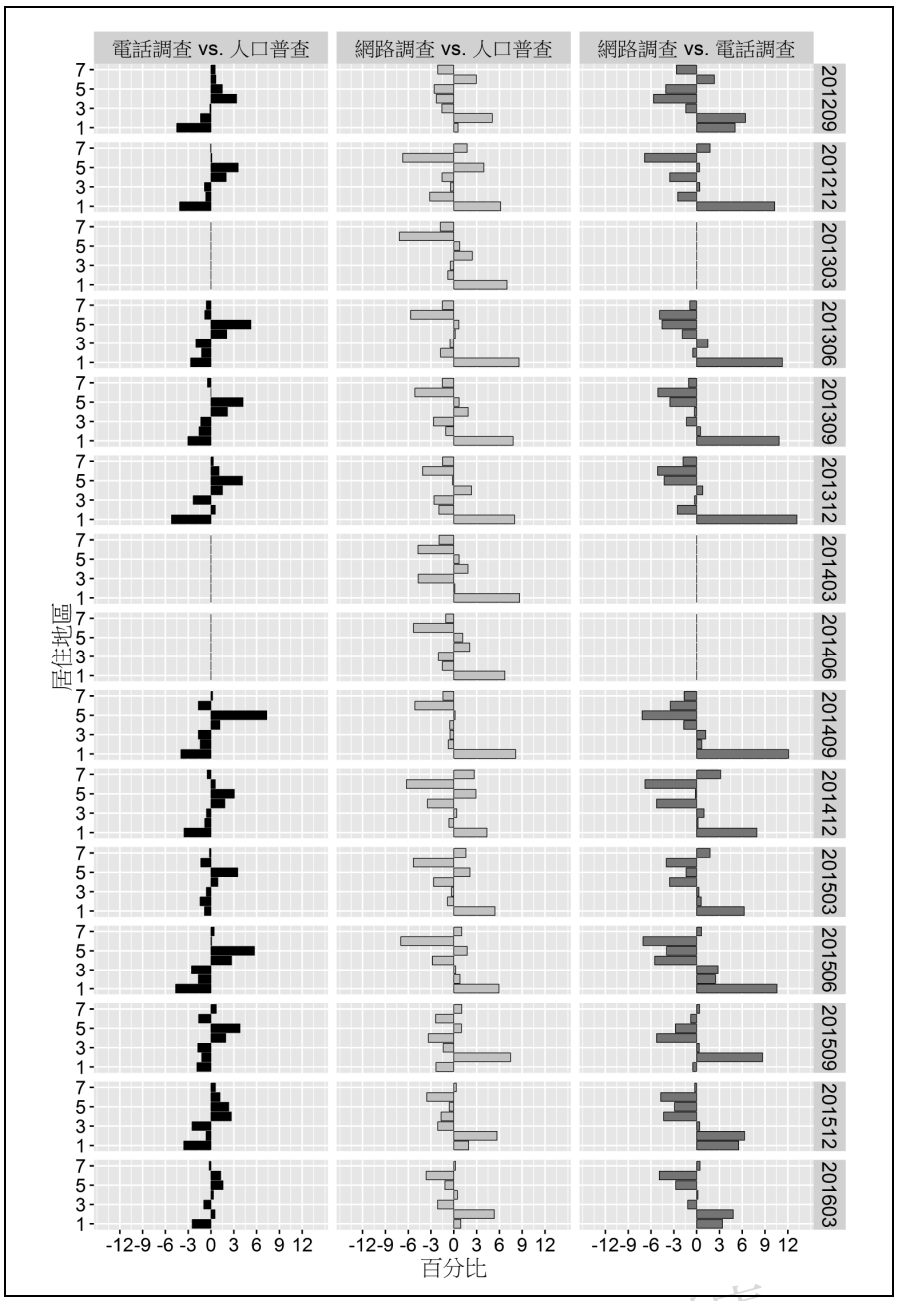
附圖 3 網路、電話調查與人口普查交叉比較「教育程度」比例差異



註：1 小學及以下，2 國中或初中，3 高中或高職，4 專科、大學及以上。
資料來源：同附圖 1

臺灣大學學術期刊資料庫

附圖 4　網路、電話調查與人口普查交叉比較「居住地區」比例差異



註：1 大臺北地區，2 新北基隆，3 桃竹苗，4 中彰投，5 雲嘉南，6 高屏澎，7 宜花東。
資料來源：同附圖 1

臺灣大學學術期刊資料庫

說明：

附圖 1（性別）

- 普遍來說，電訪受訪者的女性較多一些，但和普查人口資料相比，差異有限。
- 網路受訪者的性別差異和電訪正好相反，男性受訪者明顯較多，女性受訪者則較少。與人口資料相比，差異大概在 5～10 個百分點。

附圖 2（年齡）

- 電訪受訪者年輕人較少，特別是 20～29 歲的族群，和普查資料相比，有時甚至差到 10 個百分點以上。同時，電訪受訪者的中、老年人較多，特別是 40 和 50 多歲的族群。
- 網路調查受訪者的年齡分佈則正好與電訪相反，年輕的族群較多，特別是 30 及 40 歲的中年族群。而 60 歲以上的老年族群的人數明顯較少，多數時候都比母體少了 10 個百分點以上。

附圖 3（教育程度）

- 電話調查樣本的教育程度分佈，主要是中高教育程度（高中或大學以上）較多，低教育程度較少，但和普查資料相比，差距至多 15 個百分點左右。
- 網路調查樣本則一面倒是大學以上學歷者居多，和母體資料相比，有時偏誤甚至超過 40 個百分點。

附圖 4（居住地區）

- 電訪樣本在大都會區（譬如大臺北都會區）的比例通常較母體資料少，在中、南部農業縣（譬如中彰投及雲嘉南）的比例則較高一些。不過，這種地區性的偏誤有限，往往不到 5 個百分點。
- 網路受訪者的偏誤方向則相反，都會區的受訪者比例明顯高很多（譬如大臺北都會區及包涵高雄市的高屏澎地區），農業地區的比例則較低。平均來說，網路調查都會區的樣本偏差會達到 5～10 個百分點。

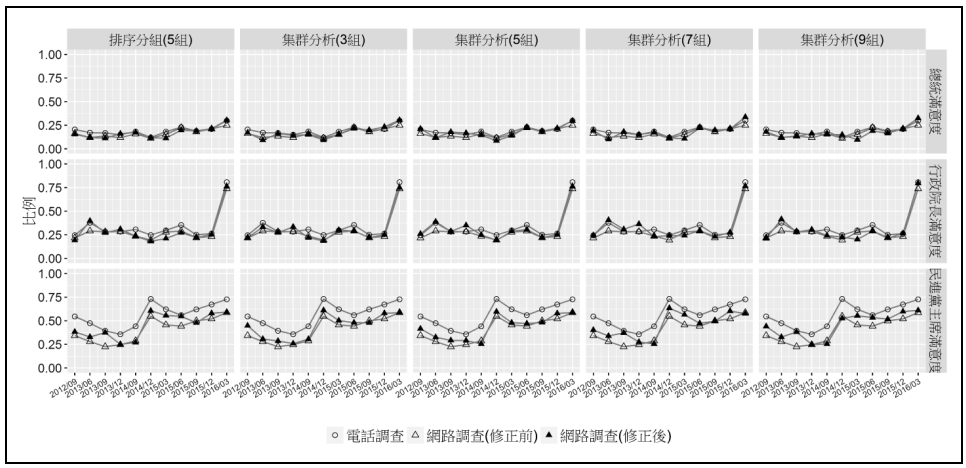
總之，和母體特徵（人口統計數據）相比，電話調查的樣本較偏向女性、中老年、中高教育水準、鄉村地區。由於偏誤幅度有限，研究者多半採事後加權處理。以求樣本符合母體特徵。網路調查的樣本特徵則反應出所謂的「數位落差」，即會接受網路調查的民眾主要還是以年紀輕、受過高等教育、及居住在都會區的族群爲主。和電話調查相比，「數位落差」所造成的網路樣本偏誤情況嚴重很多，若採用事後加權處理，則權重值勢必很極端（譬如網路調查的老年人太少，一個老年人的意見可能需要乘上 5 倍才會使年齡符合母體特徵），可能使得結果產生更嚴重的偏誤。

臺灣大學學術期刊資料庫

附錄 2
電話調查結果、修正前網路調查結果、及修正後網路調查結果折線圖

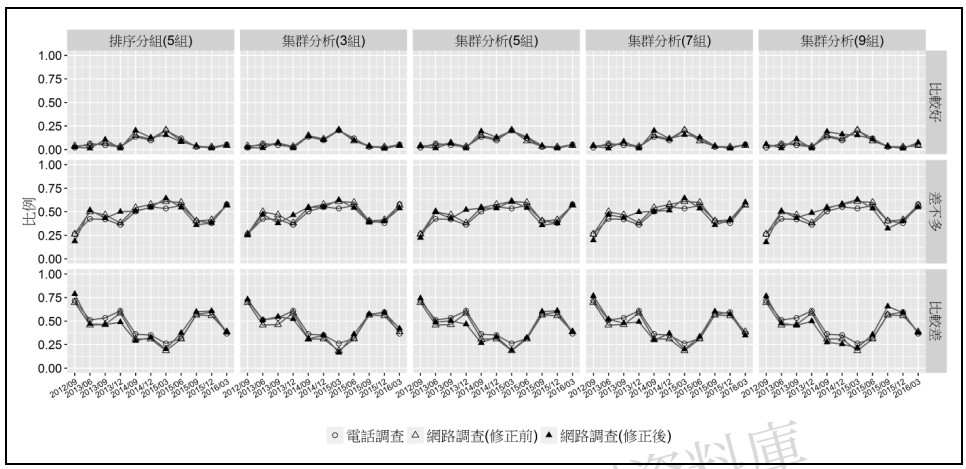
說明：以下三張附圖中，電話調查結果（以空心圓點標示）、修正前的網路調查結果（以空心三角點標示）、及修正後的網路調查結果（以實心三角點標示）

附圖 5　總統、行政院長及民進黨主席滿意度之修正前後與修正方法比較



資料來源：同附圖 1

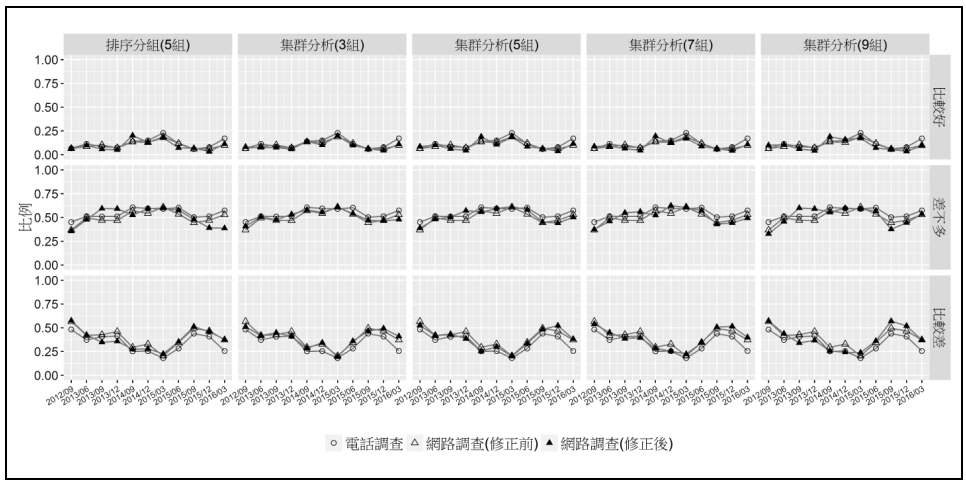
附圖 6　回顧經濟評估修正前後與修正方法比較



資料來源：同附圖 1

臺灣大學學術期刊資料庫

附圖 7　前瞻經濟評估之修正前後與修正方法比較



資料來源：同附圖 1

參考書目

洪永泰，1996，《戶中選樣之研究》，臺北：時英。

洪永泰、洪百薰、林宇璇、呂孟穎、許勝懋、吳淑惠、卓仲彥、徐書儀，2014，〈手機使用對臺灣地區電話調查涵蓋率之影響評估〉，《調查研究－方法與應用》，31: 7-30。

許勝懋，2015，〈「唯手機族」對未來電話調查的挑戰與啓發〉，《調查研究－方法與應用》，34: 33-65。

俞振華，2013，〈網路民意調查的理論與實務〉，陳陸輝（編），《民意調查新論》，臺北：五南，頁 89-110。

臺灣大學學術期刊資料庫

Exploring the Use of Telephone Surveys and Propensity Score Adjustments to Correct Web Survey Biases*

*Eric Chen-hua Yu*** & *Chih-yang Tu****

Abstract

While web surveys are observed quite frequently in our day-to-day lives, scholars rarely take their results seriously. This is mainly due to the fact that the results of web surveys are the product of non-random sampling and cannot be used to draw inferences about the target populations. This paper utilizes numerous parallel waves of telephone and web survey data, from Taiwan's Election and Democratization Studies (TEDS), to explore whether or not the possible biases produced by web surveys can be corrected with a "reference group" dataset drawn from telephone surveys. Specifically, it uses the so-called "Webographic variables" and combines the two types of data to produce proper estimates for web surveys via the propensity score adjustment (PSA) method. Overall, this analysis examines the effectiveness of using data from telephone surveys and the PSA method when dealing with possible biases of web survey results. Additionally, this paper highlights some key pros and cons of using such methods and suggests various venues for future research.

Keywords: Web Survey, Non-random Samples, Webographic Variables, Propensity Score Adjustment

---

* DOI:10.6166/TJPS.73(81-126)

** Associate Research Fellow at Election Study Center and Associate Professor at Department of Political Science, National Chengchi University, E-mail: ericyu@nccu.edu.tw.

*** Master of Arts, Department of Political Science, National Chengchi University, E-mail: tu.chihyang@gmail.com.

臺灣大學學術期刊資料庫

126 探討以電訪資料及「入選機率調整法」修正網路調查偏誤的可行性 俞振華、涂志揚

臺灣大學學術期刊資料庫