

科技部補助專題研究計畫成果報告 期末報告

再探全國性民調推估地方民意的可行性:應用改良式多層次貝氏
定理估計模型及事後分層加權預測立法委員選舉結果

計畫類別：個別型計畫
計畫編號：MOST 104-2410-H-004-090-
執行期間：104年08月01日至105年10月31日
執行單位：國立政治大學選舉研究中心

計畫主持人：俞振華
共同主持人：蔡佳泓
計畫參與人員：碩士班研究生-兼任助理人員：潘心儀

報告附件：出席國際學術會議心得報告

中華民國 106 年 01 月 18 日

中文摘要：本研究利用2016年大選前的民意調查資料，並採用多層次貝氏定理估計模型搭配分層加權的方式（multilevel regression and post-stratification: MRP），預測73個區域立委選舉結果。具體來說，本文所採用的預測模式包含三個步驟：首先，透過基本人口特徵變數（性別、年齡、及教育程度）輔以選區層級的特徵，估計不同類型選民分別支持國民黨立委參選人及民進黨立委參選人的機率。其次，我們使用內政部2015年全國人口調查資料，求得每一個選區當中，不同類型選民的聯合機率分佈。最後，將各個選區內不同類型選民當中，支持國民黨立委參選人（及民進黨立委參選人）的成年人口數加總（每個選區皆含50種類型），並分別除以各選區的總成年人口數，以推估每一選區當中，國民黨立委參選人及民進黨立委參選人的得票率。在選區樣本數有限（平均約55個）的情況下，本研究仍能透過多層次統計模型及人口調查資料輔助，得出各選區政黨候選人得票率預測值與實際得票率之間的平均誤差值之絕對值僅約5個百分點。此外，本研究成功預測61個立委選區的選舉輸贏，與「未來事件交易所」的選舉預測結果相比較，僅落後一個選區。

中文關鍵詞：多層次貝氏定理估計模型；事後分層加權；選舉預測；全國民調；立委選舉

英文摘要：This study uses pre-election national survey data and a method combining the Bayesian multilevel modeling approach with the population information for post-stratification (i.e., multilevel regression and post-stratification: MRP) to predict Legislative Yuan elections in the 73 single-member districts. Specifically, our method is consisted of three steps: first, we construct a multilevel logistic regression model to estimate the vote choice variables for the Kuomintang (KMT) and Democratic Progressive Party (DPP) candidates, respectively, given demographics and districts of residence. Second, we post-stratify on all the variables in the model by using the joint population distribution of the demographic variables within each district. Third, we then combine the above two steps and estimate the mean of support for the KMT and DPP candidates in the district level. Given that each district only has about 55 samples on average, this study shows that MRP method can be regarded as an effective tool for election prediction, as the average absolute measurement error between the estimates and actual vote shares is just about 5 percentage points. In a comparison with the pre-election district-level predictions issued by the prediction market “xFuture”, our estimates are almost as good as the results of “xFuture”.

英文關鍵詞：Bayesian multilevel modeling; post-stratification; election prediction; national survey; Legislative Yuan elections

「再探全國性民調推估地方民意的可行性：
應用改良式多層次貝氏定理估計模型及事後分層加權預測立法委員選舉結果」
研究成果報告（精簡版）

中文摘要

摘要

本研究利用 2016 年大選前的民意調查資料，並採用多層次貝氏定理估計模型搭配分層加權的方式（multilevel regression and post-stratification：MRP），預測 73 個區域立委選舉結果。具體來說，本文所採用的預測模式包含三個步驟：首先，透過基本人口特徵變數（性別、年齡、及教育程度）輔以選區層級的特徵，估計不同類型選民分別支持國民黨立委參選人及民進黨立委參選人的機率。其次，我們使用內政部 2015 年全國人口調查資料，求得每一個選區當中，不同類型選民的聯合機率分佈。最後，將各個選區內不同類型選民當中，支持國民黨立委參選人（及民進黨立委參選人）的成年人口數加總（每個選區皆含 50 種類型），並分別除以各選區的總成年人口數，以推估每一選區當中，國民黨立委參選人及民進黨立委參選人的得票率。在選區樣本數有限（平均約 55 個）的情況下，本研究仍能透過多層次統計模型及人口調查資料輔助，得出各選區政黨候選人得票率預測值與實際得票率之間的平均誤差值之絕對值僅約 5 個百分點。此外，本研究成功預測 61 個立委選區的選舉輸贏，與「未來事件交易所」的選舉預測結果相比較，僅落後一個選區。

關鍵詞：多層次貝氏定理估計模型；事後分層加權；選舉預測；全國民調；立委選舉

Abstract

This study uses pre-election national survey data and a method combining the Bayesian multilevel modeling approach with the population information for post-stratification (i.e., multilevel regression and post-stratification: MRP) to predict Legislative Yuan elections in the 73 single-member districts. Specifically, our method is consisted of three steps: first, we construct a multilevel logistic regression model to estimate the vote choice variables for the Kuomintang (KMT) and Democratic Progressive Party (DPP) candidates, respectively, given demographics and districts of residence. Second, we post-stratify on all the variables in the model by using the joint population distribution of the demographic variables within each district. Third, we then combine the above two steps and estimate the mean of support for the KMT and DPP candidates in the district level. Given that each district only has about 55 samples on average, this study shows that MRP method can be regarded as an effective tool for election prediction, as the average absolute measurement error between the estimates and actual vote shares is just about 5 percentage points. In a comparison with the pre-election district-level predictions issued by the prediction market “xFuture”, our estimates are almost as good as the results of “xFuture”.

Key Words: Bayesian multilevel modeling; post-stratification; election prediction; national survey; Legislative Yuan elections

一、前言

選舉研究當中，最引人入勝的莫過於選舉預測 (election prediction)。不論大小選舉，選前各界最關注的焦點都是誰會輸，誰會贏。台灣自從 1990 年代民主化後，幾乎年年都有大規模選舉。選舉的不確定性也吸引不少專家學者，利用總體或是個體的資料，建立各式各樣的選舉預測模型。這些預測模型除了檢證相關選舉理論外，對於競選操作的實務也有不少貢獻。

台灣過去的選舉預測模型多半利用選前的民意調查資料，藉由理論建立投票選擇模型，並透過複雜度不一的計算方式來拆解選民的投票意向。這種以個體民調資料為核心的選舉預測模式相當普遍，全球各地都有相關的研究，且大體來說，準確度都還不差 (Buchanan 1986)。然而，既然是預測，自然就會出現失準的現象，譬如 2014 年台灣的九合一地方選舉，就有許多地區的開票結果 (譬如新北市、桃園市、和新竹市) 和選前媒體及專家利用民調資料的預測有很大的出入 (李皇萱 2014)。

誠如劉義周 (2009) 所言，利用民調資料的選舉預測模式需具備三項要素才能準確：首先，需要高品質的調查資料；其次，需要有效的理論做為解釋資料的依據；最後，需要用正確的方法來分析資料。換言之，如果要改良過去的選舉預測模型，應藉由改進上述三項要素入手，即包括資料面、理論面、及方法面。本研究旨在以俞振華、蔡佳泓 (2006) 所發展的選舉預測模型為基礎，同時強化模型的資料面、理論面、及方法面，並將該模型應用在 2016 年立委選舉預測上。

過去俞振華、蔡佳泓以 Park et al. (2004) 的模式為基礎，所提出的選舉預測模式是以多層次貝氏定理估計模型搭配分層加權的方式，利用 2004 年總統大選前的全國民調資料，來推估各縣市 (即扣除金門及馬祖兩外島之 23 縣市) 的陳水扁得票率。雖然該研究使用的是全國性調查資料，各縣市的樣本數有限，但透過統計模型及普查資料輔助，使得各縣市預測值與實際得票率之間的平均絕對誤差值僅約 1.5 個百分點。具體來說，該模式包含兩個部份：首先，將全國人口分成各個階層，並利用少數的全國性個體樣本輔以區域總體特徵，估計出每一種人口階層的投票的意向；其次，配合人口普查資料得知各區域 (或次層級) 中人口階層的分佈，預測各區域或次層級的行為與態度，例如全國每一縣市的政黨得票率。該模式的核心優勢可歸納為以下三點：

- 第一、 有效地利用有限的樣本數。該研究能以不到 3500 筆全國樣本推估 25 個縣市的投票意向，主要是因為該模式結合個體層次的民意調查資料及總體層次的分區特徵 (普查資料)，才得以克服分區樣本數不足的限制。
- 第二、 可納入更多的資訊。多層次的貝氏模型可以協助研究者納入更多的資訊來預測選情，特別是之前的資訊。譬如在區域層級的模型中，納入上一次選舉時該區域的政黨得票率，自然有助於預測該區域此次的選舉結果，也有助於驗證區域層級的變數對於個體投票行為的影響。又或在貝氏模擬時，利用過去的民調資料及統計模型係數，設定此次模型係數的先驗 (事前) 分佈 (prior distribution)。於是，透過結合前次資料所形成的先驗分佈及此次調查資料，勢必也會增加推估此次選舉的效率 (Tsai 2005)。
- 第三、 將全國的民意納入分區的考量中。該模式首先利用人文區位變數與全國的實際觀察值構成統計迴歸模型，再據以配合普查資料求得縣市民意的估計值。由於模型的迴歸係數是根據全國

樣本，而非分區產生迴歸係數，因此該途徑等於考量了選舉時全國性的「投票風向」(national tie)，並借用了其它縣市的資訊來輔助求取某一縣市的估計值。這樣的分析架構其實頗符合選舉時的情境，即每個分區的選舉雖然各有其候選人或地域性的因素，但全國性的因素仍會有某種程度的影響力。

總之，該選舉預測模式能夠使我們更有效地利用有限的樣本數、結合不同層級或時期的資料、及同時考量全國及分區變數對於投票行為的影響。而上述三項特徵，都有助於我們利用有限的全國樣本推估各個分區的民意。過去數年來，雖然已有利用該模式推估區域或縣市層級的民意(蔡佳泓 2008; 2014)，但從未應用到立委選區層級。然台灣立委選舉有 73 個單一選區，倘若我們要以傳統分區民調方式預測 73 個選區的選舉結果，根據機率抽樣的理論，樣本數開根號後與抽樣誤差成反比，於是在特定誤差範圍的要求下(譬如 ± 3 個百分點)，我們可能需要約近 7 萬 8 千筆樣本($1067 \times 73 = 77891$)。就算誤差範圍要求僅為 ± 5 個百分點，也需要超過 2 萬 8 千筆樣本($385 \times 73 = 28105$)。這樣龐大的樣本數當然不是不可能達成，只是不論採用面對面訪問或是電話調查，肯定所費不貲。假設我們全國隨機抽取了 3000 個樣本，推論全國性選舉譬如總統大選，在 95% 的信心水準下，抽樣誤差約只是 ± 1.8 個百分點，可謂相當精準。但如果將這 3000 個樣本分配到 73 個單一選區，則平均每個選區大概只有約 40 個樣本，於是在同樣的信心水準下，抽樣誤差將超過 ± 15 個百分點！在抽樣誤差這麼大的情況下，我們又將如何相信依民調數字所得出的預測結果？本研究的目的即是如何利用有限的民調樣本，應用俞振華與蔡佳泓過去所發展的選舉預測模型，預測區域立委的選舉結果。如果這樣的模式可行性高，則日後以全國民調資料預測分區選舉結果的成本將大幅減少。

二、研究目的

本研究利用全國性的民調資料與其他層級資料相結合，建立多層次分析模型，並透過貝氏統計方法估計，藉以評估次層級的民意。具體來說，我們首先將全國人口分成各個階層，並利用少數的全國性個體樣本輔以區域總體特徵，估計出每一種人口階層的投票的意向。其次，配合人口普查資料得知各區域(或次層級)中人口階層的分佈，預測各區域或次層級的行為與態度，例如全國每一縣市或立委選區的政黨得票率。這種利用個體樣本建立統計模型並輔以母體資料進行事後分層加權的模式常被用來求取小區域(少樣本)的估計值(Small Area Estimation, Gelman and Little 1997)。

所謂 Small Area Estimation 的 Small 並非單指小的地理區域，而是泛指較少的樣本數。由於少量的樣本數將導致點估計值的變異數或標準差過大而降低了該估計值的效度，學者於是發展各種方法來解決小樣本所帶來的問題。這類型研究的發展主要可分為兩支：第一、直接估計法(direct estimator)；第二、間接估計法(indirect estimator)。前者是利用各種加權的方式直接處理樣本觀察值，當然小樣本的問題在這裡依舊存在；後者則是利用統計模型配合輔助的變數(auxiliary variables)間接求取樣本估計值。換言之，儘管樣本數不足，但我們可以透過統計模型和輔助的資訊來提高估計值的準度與效度。這種間接估計的方法又稱為 Model-based Estimator，是近來相關統計學者研究發展的重心，應用的層面相當廣泛，尤其在各類型專注於區域差異的民意調查與政策研究上(Rao 2003)。

除了旨在利用全國性樣本推估分區層級的民意以外，本研究另一個目的就是精進台灣選舉預測的方法。有關選舉預測方面的文獻，若是以資料的類型區分，主要可分為兩種：以總體資料預測及以個體資料預測。前者是以歸納法的角度出發，運用總體資料，建立線性或非線性模型以預測選舉結果(Kramer 1971; Tufte 1978; Rosenstone 1983; Abramowitz 1988; Lewis-Beck 1992; Tsai 2000)。這一類的

選舉預測視每一次選舉結果為一個分析單位，利用長期的資料並假設線性或非線性模型，探討諸多變數包括經濟因素（如選前第三季的 GDP）、現任者評價（如現任總統的支持度）、或選前的特殊事件（如是否選舉年有戰爭）等和政黨得票率之間的關聯性。這一類的選舉預測（election forecasting）強調變數之間關聯性的強弱，或是模型與資料之間的適合度（model fit），但往往缺乏理論依據及無法解釋因果關係，畢竟使用總體資料去推論個體投票行為將無法避免區位謬誤。此外，不是每個國家都像美國或西方民主成熟國家一樣，已累積夠多數目的選舉結果可供分析。以台灣來說，總統大選也不過舉行了 5 次，立委選舉也只有 7 次，樣本數太少使得這類型分析很難執行。過去只有學者採用這樣的模式來預測縣市長選舉，畢竟每次縣市長選舉會有超過 20 個樣本，還可以透過時間序列橫斷面分析模型（time-series cross-sectional data analysis），來探討縣市層級總體變數與縣市長得票率之間的關係（黃智聰、程小綾 2005）。總之，這種利用總體資料的選舉預測模式，現階段仍不太適用於台灣，主要的困境在於資料點（data point）不夠多。少數的例外是使用選前數波調查結果做為資料點，利用時間序列模型來預測選舉結果（徐永明 2005）。另外，還有一些學者採用總體投票資料進行各個行政單位的分析，小至投開票所，大至縣市，但一般則是以村里為分析單位。於是，根據區域單位的投票資料及人口區位變數，學者可以進行時序性或者是定時性的分析（雷飛龍、陳義彥、丁庭宇 1985；盛杏媛 1998；鮑彤 1999）。惟這類型的分析多數是用來解釋選舉結果，往往無法在選前提供精確的數字評估（洪永泰 1994）。

因此，在台灣所謂的選舉預測，多半是指利用個體資料，特別是指利用民意調查資料，來預測選舉結果。這種以個體資料預測的模式不像總體模式一般，是透過總體層級各變數間的關聯性來預測選舉結果（譬如經濟成長率與得票率）。相對地，個體資料的選舉預測模式是從選民行為的理論出發，利用調查來探究民眾的投票意向，再將個體的投票意向「加總」起來，以判斷選舉誰輸誰贏。換言之，這類型的分析模式需要解決兩個問題：第一、如何精準地透過調查方式探知個人的投票意向？如果直觀上得不到答案，則要如何旁敲側擊？第二、要如何將個體的態度「加總」？畢竟我們感興趣的不是某一人是否會投給某位候選人，而是候選人會得到多少選票。

國內過去的選舉預測文獻多半聚焦在處理第一個問題，特別是過去台灣特殊的威權政治環境，許多選民在選前不願意表態。於是，探知個人真實的投票意向或拆解未表態的選民成為選舉預測的關鍵。洪永泰(1994)即使用總體的人口區位資料，將全國劃歸不同的政治版圖。當調查的受訪者不願表態時，則採用政治版圖的資訊來預測那些未表態選民可能的投票意向。洪氏的分析途徑可謂首開台灣結合個體與總體資料預測選舉的濫觴，準確度歷久不衰，並有許多延伸模型，譬如將該模式應用在三方選舉的策略投票上（莊文忠 2000）。總之，即使在二十年後的今日，洪永泰所發展的選舉預測模式仍然被選舉實務界廣泛採用。

除了利用總體資料拆解未表態以外，還有諸多研究是單純利用民調資料來探求選民真正的投票意向，包括陳義彥（1994）利用調查資料進行集群分析，以設定不同投票意向的選民集群；劉義周（1996）應用 Kelly and Mirer（1974）的「簡單投票」（simple act）決策模式，設想選民會分階段透過政黨、候選人、及政見等多重因素來思考要把票投給誰，並將受訪者依序歸類後預測選舉結果。劉文卿（1995）同樣假定個人的投票行為會被諸多因子所影響，並引用遺傳演算法的觀念，將各個變數轉化為基因，建立基因模型。該預測模型預設選民會投給和他/她基因看起來最相似的候選人，於是我們可以據此歸類選民的投票意向並預測選舉結果。劉念夏（1996）亦使用多項變數，建立區隔模型將選民區辨為 12

群，並改良劉義周的選民決策模式，將預測模式的焦點放在選民對候選人的評價上，其次才是政黨認同。梁世武（1994）則捨棄多重變數，建立候選人形象投票預測模型，將候選人特質預設為選民投票時參考的最主要依據。李錦河、溫敏杰（1998）及李錦河、溫敏杰、陳盈太（2010）則將候選人特質轉化成產品屬性及品牌知名度，利用行銷學的概念建立選民需求及認知指標來預測選舉。盛治仁（2000）的預測模型同樣強調候選人因素，他使用「情感溫度計」測量民眾的候選人評價，從而分析未表態的選民。盛杏媛（1998）則認為，過去的個體投票預測皆忽略了選民投票的「不確定性」，認為唯有透過機率模型，計算每一個觀察值可能投給特定候選人的機率，才能完整地詮釋選舉預測的結果。盛杏媛、周應龍（2008）則直接透過統計模型來拆解未表態，即採用選樣偏誤模型來檢視未表態者與表態者間的差異，並將未表態者的資訊納入投票預測模型中，也確實提升了選舉預測的精準度。總之，上述單純依賴民調資料所建立的各項選舉預測模型，主要是透過諸多理論視角及不同的統計方法來分析選民真實的投票意向，可視為是台灣選舉預測模式的主流。

不過，上述強調個體投票行為的文獻往往不太在意將個體的態度「加總」的過程，並假定民意調查的樣本大抵符合母體特徵。的確，倘若民意調查的抽樣程序完全符合簡單隨機抽樣原理，即母體中每個人被抽到的機會是一樣的，則我們將民調樣本中支持某候選人的次數加總起來除以總受訪人數，就可以推論某候選人的支持度。但如果樣本並不完全符合母體特徵，譬如年輕人太少時，則我們在加總時就需要針對年齡層加權。當然，加權又有很多不同的作法及考量，譬如是否考量計算權重時所使用之變數間的交互關係，就可產生不同的加總結果（黃紀、張佑宗 2003）。又倘若抽樣程序有偏誤，比方說採用家用電話簿抽樣永遠抽不到唯手機族，且唯手機族的投票意向和其他人有顯著的不同時，則無論我們採用何種方式將電話調查樣本次數加總起來，其預測結果肯定和實際結果有所差距。總之，正因為從個體資料轉換成總體結果的預測模式之間存在一個加總的過程，於是該如何加總，便成為個體資料預測模型的關鍵。而俞振華、蔡佳泓（2006）以統計模型輔以普查資料的預測模式是相關研究當中，少數關注加總過程的預測模式。

總結上述選舉預測的文獻，本研究主張：未來的選舉預測模式應納入更廣泛的資料，尤其應該善用總體資料來補強個體資料的不足。此外，針對如何將個體資料「加總」起來，成為選舉結果預測值？或許採用更細緻的加權方式是一個方向。當然，目前電話調查可能造成樣本偏誤，使得選舉預測失準，值得所有透過電訪民調資料從事選舉預測者進一步關注。雖然本研究不可避免會觸及該議題，但調查模式所造成的差異畢竟不是本文的核心，在此不深入探討。¹

三、研究方法與執行成果

本研究旨在利用立委選前全國性的調查資料，透過多層次貝氏定理模型並搭配人口調查資料進行分層加權（即 MRP 模式），以預測各個選區的政黨得票分佈。根據調查資料，我們可以將某個選區的選民依照性別、年齡、教育程度等人口變數細分為許多層，然後再以這些資訊預測民調資料當中，個別選民投票支持某一政黨（或候選人）的可能性，以估計出各個分層的影響係數。而在估計過程中，最基本的作法是設定各個分層的估計係數服從常態分布，變異數則是服從反伽瑪（inverse-gamma）的事前機率分布且相互獨立。當然，這兩項事前資訊可以更進一步改善，譬如用過去的資料來設定各估

¹ 相關討論可見許勝懋（2015）。該研究利用 TEDS2013 面訪資料發現，台灣的唯手機族約佔總人口的 6% 左右，對於電話民調結果的影響有限。

計係數的事前機率分配，使得貝氏模型可以更精確地產生估計係數的事後機率分配。另外，我們可以進一步設定選區層次模型，譬如以上次選舉各選區的政黨得票率，或此次選舉是否有在位者等變數，來區辨 73 個立委選區間的差異，以發揮多層次模型的效用。

具體而言，本研究採用貝氏定理的分析工具——馬可夫鏈蒙地卡羅法（Markov chain Monte Carlo，以下簡稱 MCMC）模擬事後機率分配——以建立選舉預測的模型。²資料方面，除了全國民調資料以外，我們還使用 2015 年內政部公佈的全國人口調查資料，透過計算每一選區的性別、年齡、教育程度等比例，將各個變數交叉後形成聯合分佈（joint distribution），供我們事後透過模型的估計係數進行分層加權之用（post-stratification）。最後，我們可以將各層的投票意向加總起來，推估每一選區的政黨（或候選人）得票率。以下我們根據 Park et al. (2004) 與俞振華、蔡佳泓 (2006) 的模型架構，說明如何透過 MRP 模式，來預測 2016 年區域立委選舉中，兩大黨（民進黨及國民黨）區域立委參選人的得票率：

首先，我們建立對數模型。依變項為受訪者就特定問題（譬如：是否支持國民黨的候選人）的反應 y ，自變項為受訪者的基本人口特徵和其戶籍地所屬之立委選區。基本特徵包括性別（男或女），年齡（五項分類：20-29 歲、30-39 歲、40-49 歲、50-59 歲、及 60 歲以上），和教育程度（五項分類：小學及以下、中學、高中高職、專科、及大學以上）。立委選區則包括所有 73 個區域立委單一選區。以上基本特徵和立委選區可組合出 3,650 種（ $2 \times 5 \times 5 \times 73$ ）類型 j (categories)，從「連江縣」（即本研究選區標號第 1 號）、「男性」、「20-29 歲」、且「不識字或小學以下」，到「桃園市第六選區」（即本研究選區標號第 73 號）、「女性」、「60 歲以上」、且「大學以上」。透過這個統計模型，我們旨在估計每一個類型就特定問題的反應 y 之平均值 π_j 。

其次，利用 2015 年內政部所公佈的全國人口統計資料，我們可以得知每一個類型 j 中有多少成年人口 N_j 。舉例而言，我們可以從人口資料推估，2015 年時有 10,253 位成年人為「男性」、年齡為「65 歲以上」、教育程度為「大學以上」、且戶籍在「台北是第一選區」。於是，每一個選區（ d ）就特定問題的反應 y 之平均值為該行政區內各基本特徵類型（50 種類型）就特定問題的反應 y 之加權平均值，其中權數為該選區內各類型成年人口占其總成年人口的比例：

$$\theta_c = \frac{\sum_{j \in c} N_j \pi_j}{\sum_{j \in c} N_j}$$

總之，對於任何一個全國性民調中二分的問題，我們皆可利用以上兩個步驟，即以對數模型的估計係數並配合母群多變數聯合分佈值（multivariate joint distribution of population）進行事後分層加權的方式，來求得各選區對該二分問卷調查題目的分區民調結果。

² MCMC 的基本原則為先假設一個樣本分布，然後從這個分布抽出一套樣本含有未知的參數，以 θ 表示。接下來按照我們已觀察到的樣本，做一定的修正，以極大化觀察到 θ 的機率。在每一次抽出（或模擬）樣本過程中，機率一直被極大化，而最後所得的樣本就是最佳的樣本。之所以稱為 Markov chain 是因為我們做一連串的模擬，每次模擬都是根據上一個模擬而來。而這正是貝氏定理的精神——根據所有已知的資訊預測下一次事件發生的條件機率。軟體可以模擬一連串的抽樣，例如五百次或一千次或更多，然後觀察最後達到一定的聚合時的樣本分配。

以 2016 年立委選舉為例，假設受訪者反應 $y_i = 1$ 表示受訪者 i 支持國民黨區域立委參選人（或支持民進黨區域立委參選人）， $y_i = 0$ 表示受訪者 i 不支持國民黨區域立委參選人（或不支持民進黨區域立委參選人），³ 未表態者則設為遺漏值，由模型推估（設為為 NA ）⁴。則個別受訪者支持國民黨區域立委參選人（或民進黨區域立委參選人）的機率為：

$$\begin{aligned} y_i &\sim \text{Binomial}(p_i) \\ p_i &= \Pr(y_i = 1) \\ &= \text{logit}^{-1}(X\beta)_i \end{aligned}$$

X 是 0 與 1 所構成的矩陣，代表受訪者各基本特徵分類項（性別、年齡、及教育程度），加上年齡與教育程度的交叉項。另外， X 還包括了 73 個選區項。 β 向量代表 X 的估計係數。於是，個體層級的對數迴歸模型可表示如下：

$$\Pr(y_i = 1) = \text{logit}^{-1}(\beta_0 + \beta^{\text{female}} \times \text{female}_i + \beta_{\text{age}(i)}^{\text{age}} + \beta_{\text{edu}(i)}^{\text{edu}} + \beta_{\text{age}(i) \times \text{edu}(i)}^{\text{age} \times \text{edu}} + \beta_{\text{district}(j)}^{\text{district}}) \quad (1)$$

進一步地，我們假設第 (1) 式中代表選區差異的截距項 β^{district} 係數呈常態分配，即所謂的 Random Effects。接著，我們分別針對國民黨及民進黨立委得票率的估計模型，利用選區內是否有現任者及 2014 年縣市長選舉時，國民黨縣市長參選人（或民進黨縣市長參選人）在這 73 個選區的得票率，建立選區層級的迴歸模型，以求更準確地掌握 73 個選區之間受訪者投票意向的普遍差異。以國民黨的模型為例，選區層級的模型表示如下：

$$\beta_j^{\text{district}} \square N(\beta_{\text{incumbent}(j)} + \beta^{2014\text{KMTvote}} \times 2014\text{KMTvote}_j, \sigma_{\text{district}}^2), \text{ for } j=1, \dots, 73 \quad (2)$$

選區模型置入兩個自變數（predictor）的原因有以下兩點：⁵ 第一、對於小選區的議員選舉而言，現任者是否參選對於選情有很大的影響。由於可以連選連任，且選區內只有一位現任者，使得現任者能夠經營所謂的「個人選票」（personal vote），有利其連任（Fiorina 1977; Cox and Katz 1996）。因此，選區模型第一個自變數即為選區內是否有現任者參選。第二、2014 年的選舉雖然是地方選舉，但由於台灣地狹人稠，縣市長選舉參選人往往也有過中央層級的歷練，甚至有全國知名度。因此即使 2014 年選舉是地方層級的競爭，但該次選舉的結果仍可視為是 2016 年大選前，政黨競爭態勢的最新發展。是故以 2014 年選舉結果來代表各個選區內政黨最新的實力，應該最能反映 2016 年大選前政黨競爭的樣態。

我們假設第 (1) 式中的年齡、教育程度、及年齡與教育程度交叉項等代表受訪者基本特徵分類項的 β 係數，和第 (2) 式中代表行政區差異的地理區域項之 β 係數各自呈獨立的常態分配，其中皆

³ 這裡將國民黨區域立委得票率與民進黨區域立委得票率分別估計的主要原因，是因為多數選區不只這兩大黨的參選人。倘若我們使用譬如多項對數迴歸模型，則除了需檢測 Independence of Irrelevant Alternative (IIA) 的問題外，又得面對許多選區只有兩大黨參選人的問題。由於本文的焦點是以全國樣本推估分區選舉結果，主要的目的是利用選前資料預測誰輸誰贏，因此我們認為用較簡單的方式，如果能清楚地呈現各個選區的競爭情形並預測輸贏，則利用兩個模型分別估計兩黨的支持度可能更簡單易懂些。

⁴ 透過貝式統計模型以多重插補的方式（multiple imputation）來模擬遺漏值。

⁵ 由於選區層次的樣本數只有 73 個，為顧及模型的自由度，我們認為不宜再置入超過兩個以上的自變數。

以 0 為平均數，各標準差包括 σ_{age}^2 、 σ_{edu}^2 、 $\sigma_{age \times edu}^2$ 、及 $\sigma_{district}^2$ 的估計值則取決於分析資料，即基於無資訊的先驗分佈 (non-informative prior)。⁶ 至於第 (1) 式中的性別項，我們將其估計係數定義為固定作用 (fixed effect)。這是由於性別只有兩類，當個體資料中的類別 $j < 3$ 時，設定其係數為隨機作用 (random effect) 並利用多層次模型估計所得的結果和定義其為固定作用的差別不大 (Gelman et al. 2004; Park et al. 2004)。

除了性別項的 β 係數以外，各 β 係數可表示如下：

$$\begin{aligned}\beta_j^{age} &\sim N(0, \sigma_{age}^2), \text{ for } j=1, \dots, 5 \\ \beta_j^{edu} &\sim N(0, \sigma_{edu}^2), \text{ for } j=1, \dots, 5 \\ \beta_{j,k}^{age \times edu} &\sim N(0, \sigma_{age \times edu}^2), \text{ for } j=1, \dots, 5; k=1, \dots, 5 \\ \beta_j^{district} &\sim N(0, \sigma_{district}^2), \text{ for } j=1, \dots, 73\end{aligned}$$

我們透過貝氏定理的分析工具，即 MCMC 模擬事後分配的樣本資料，以進一步估計以上的統計模型 (包括第 (1) 及第 (2) 式)。所使用的軟體為 WinBUGS (Spiegelhalter et al. 1999) 及 R 語言 (R Development Core Team 2003)。在模擬的過程中，可透過重覆性參數的技巧 (redundant parameterization) 來加速模擬收斂 (convergence) 的過程。至於如何評估收斂與否，我們採用 \hat{R} 值 (potential scale reduction) 是否接近 1 為標準，即所有估計參數的 \hat{R} 與 1 的差距皆小於 0.1。⁷

基於個人基本特徵變數及戶籍所在地選區別，以上對數模型可用來模擬任一類型受訪者，支持國民黨立委參選人及民進黨立委參選人的機率。接著，我們利用以上模型的估計係數，求取 73 個選區內各基本特徵類型 (50 種類型) 是否支持國民黨及民進黨立委參選人之加權平均值。

本研究的電話調查資料是由全國公信力民意調查公司協助搜集，電訪執行期間為 2015 年 12 月 26 日至 2016 年 1 月 6 日，共計完成 4,037 個有效樣本，遍布全國 73 個選區，因此平均每個選區約有 55 個有效樣本。其中，樣本數最少的是連江縣，只有 11 個有效樣本，最多的則是台南市第 2 選區，完成了 92 個有效樣本。

四、結論及未來方向

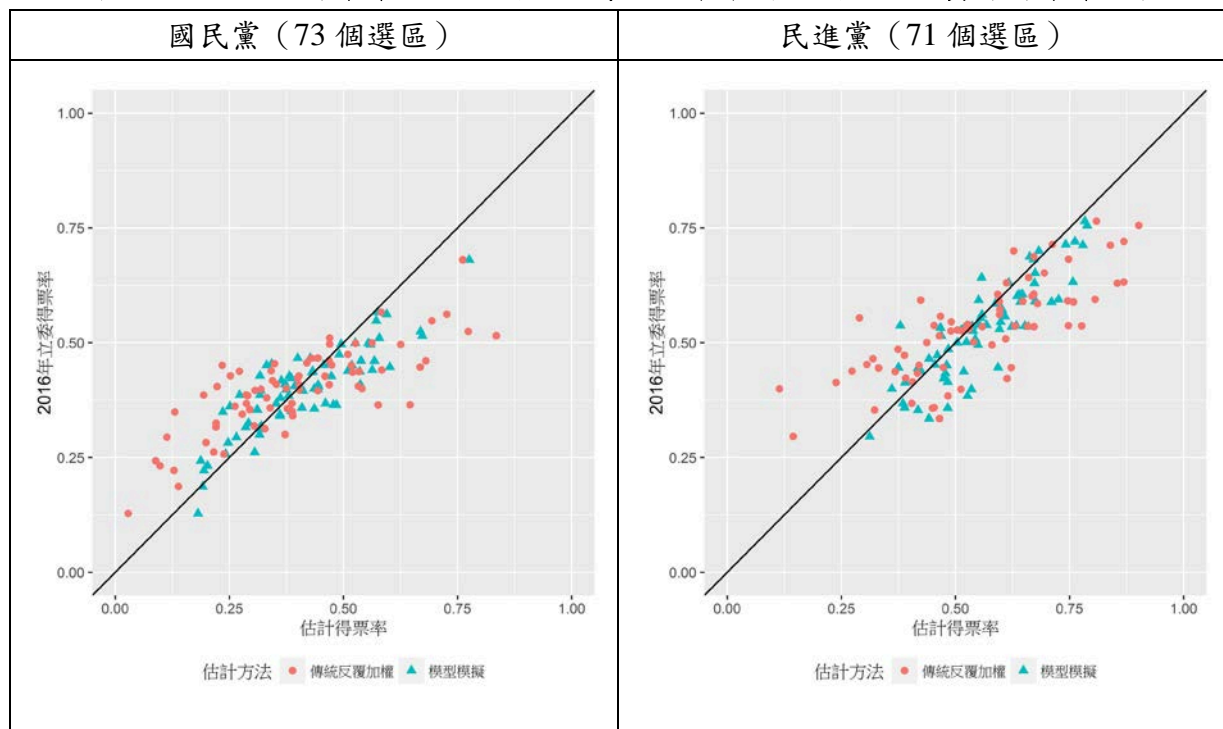
在只有約 4,000 個全國有效樣本的情況下，預測 73 個立委選區選情的結果究竟如何？如果沒有利用上述 MRP 模式，而是用傳統反覆加權 (依性別、年齡、教育程度、戶籍地選區) 的方式，預測結果是比較好還是比較差？以下圖一分別依 73 個選區，將國民黨 (及民進黨) 立委得票率的模型模擬值及傳統反覆加權平均值，與最後開票結果做比較：⁸

⁶ 各 σ 的無資訊先驗分佈設定為均等分配 (Uniform Distribution)。

⁷ 關於如何估計貝氏模型及評斷收斂與否的詳細說明請見 Gill 2008; Gelman et al. 2004; Park et al. 2004。

⁸ 民進黨於 2016 年立委選舉時有多個選區未提名候選人，本文使用其公開支持的特定候選人得票率替代；民進黨於花蓮縣未提名 2014 年縣長，本文以 2012 年立委候選人得票率替代。另外，由於民進黨在馬祖未提名 2012 年及 2016 年立委候選人，也未支持特定候選人；在馬祖與金門皆未提名 2014 年縣長候選人，無從替代。因此針對民進黨區域立委得票率的部份，排除金馬後，本研究只分析 71 個選區。

圖一、區域立委得票率模型預測值、傳統加權平均預測值、與實際得票率比較



備註：因為民進黨在外島地區未提名立委參選人，也不支持特定無黨參選人，因此其得票率估計的選區不包含金門縣與連江縣，總計 71 個。

圖一中左、右圖的橫軸是估計得票率，縱軸是實際得票率。如果估計得票率和實際得票率完全吻合，則各個選區的落點應該正好落在 45 度線上。換言之，如果估計得票率與實際得票率愈接近，則各選區的落點應該離 45 度線愈近，反之則愈遠。如果我們單純用目測衡量，可以發現圖一當中，不論是左圖（國民黨得票率）還是右圖（民進黨得票率），MRP 模型模擬值（三角型綠色落點）都比傳統反覆加權平均值（圓型紅色落點）更接近 45 度線。

接著，我們利用表一來說明得票率估計值和實際值的數值差異，即計算估計值與實際值差異的絕對值。以國民黨來說，MRP 模型模擬值與實際值的平均差異絕對值為 5.36 個百分點，遠較傳統加權平均值與實際值之間的平均差異絕對值 9.24 個百分點要好很多。至於民進黨的部份，MRP 模型模擬值與實際值的平均差異絕對值為 5.07 個百分點，同樣較傳統加權平均值與實際值差異之絕對值 9.18 個百分點來得優異。此外，如果我們以 ± 3 及 ± 5 個百分點做為標準，約有四成的模型模擬值與實際值的差異小於 ± 3 個百分點（國民黨 37%；民進黨 40%），兩成的模型模擬值與實際值的差異介於 3-5 個百分點之間（國民黨 21%；民進黨 18%），另外略超過四成的模型模擬值與實際值的差異大於 ± 5 個百分點（國民黨 43%；民進黨 41%）。儘管有超過四成的選區，模型模擬值與實際值的差異大於 ± 5 個百分點，但這樣的分佈仍遠優於傳統反覆加權估計值與實際值的差異——畢竟採傳統反覆加權後，不論是針對國民黨還是民進黨的得票率，估計值與實際值差異大於 ± 5 個百分點的選區數超過六成。總之，從表一的數值分析來評斷，MRP 模型模擬值是遠優於傳統反覆加權下所得的選區政黨得票率平均值。

表一 MRP 模擬估計與傳統反覆加權估計誤差比較（百分點）

誤差絕對值	國民黨		民進黨	
	傳統加權	模型模擬	傳統加權	模型模擬
平均值	9.24	5.36	9.18	5.07
標準差	7.21	4.17	7.18	4.18
<3	17(23.3%)	27(37.0%)	17(23.9%)	29(40.8%)
3-5	11(15.1%)	15(20.5%)	10(14.1%)	13(18.3%)
>5	45(61.6%)	31(42.5%)	44(62.0%)	29(40.8%)
總計	73(100.0%)	73(100.0%)	71(100%)	71(100.0%)

備註：

1. 民進黨的選區不包含金門縣與連江縣，總計 71 個。
2. 括號內為選區數比例。
3. 選區列表請參考附錄中的附表三和附表四。

如果我們跳脫點估計的思維，而是以估計區間的方式來評估模型模擬值的優劣，則表二顯示，在國民黨得票率的部份，73 個選區中有 49 個選區的 MRP 模型模擬值之 95% 區間包含實際國民黨參選人得票率，約佔了所有選區的 67%。而就民進黨得票率的部份，則有 40 個選區（總共 71 個選區，約佔了 56%）的 MRP 模型模擬值之 95% 區間包含實際民進黨參選人得票率。⁹從這個視角來評斷，模型模擬值的表現差強人意，即就算以 95% 的區間來預測，包含實際得票率的比例不過介於五成五到七成之間。

表二 模型模擬值 95% 區間預測準確性

選區數	包含/未包含實際得票率			
	國民黨		民進黨	
	包含	未包含	包含	未包含
	49	24	40	31
	(67%)	(34%)	(56%)	(44%)

備註：

1. 民進黨模擬選區不包含金門縣與連江縣，總計 71 個。
2. 括號內為選區數比例。

最後，如果我們以是否準確預測當選人為評斷基準，則 MRP 模型模擬值的表現又是如何？我們不彷拿選前媒體時常拿來評論的「未來事件交易所」的預測結果來做比較。該交易所在進行選舉預測時，依賴的不是民調數據，而是藉由類似期貨或股票交易所的概念，透過「供給-需求」的市場法則來探索參選人支持度的「均衡價格」，並依參選人當選可能性的「價格」，來評斷誰會當選（童振源等 2009）。該交易所強調其預測機制是「集眾人的智慧」，但所謂的「眾人」其實是指實際參與交易平台者的主觀判斷。當然，這些判斷不是漫天喊價，而是需要透過市場機制，找出一個均衡值。雖然未來事件交易所在其新聞中宣稱，其預測的準確度高達 91.67%，但這個結果是因為他們排除掉「選情膠著」

⁹ 個別選區 95% 的模型模擬值區間與實際值的關係，請見附錄中的附圖一。

選區的估計（即差距在 0.5% 內者）。倘若將這些「選情膠著」選區納入計算，則未來事件交易所的準確率和本文單純以民調資料 MRP 模式模擬的準確度相比，並沒有特別高。表三顯示，在國民黨獲勝的 20 個選區當中，民調 MRP 模擬值正確預測國民黨贏得 15 個特定選區，未來事件交易所則只正確預測到 14 席。民進黨的部份，民調模擬值正確預測到 46 席，未來事件交易所則略高，正確預測了 48 席。換言之，民調 MRP 模式模擬的預測正確率為 83.56% ($61/73 = 0.8356$)，和未來事件交易所的 84.93% ($62/73 = 0.8493$) 準確度相比，差距非常有限，73 席當中只差了 1 席。¹⁰

表三 民調 MRP 模擬 vs. 未來事件交易所（席次）

政黨	大選結果	民調 MRP 模擬		未來事件交易所	
		預測席次	正確預測	預測席次	正確席次
國民黨	20	22	15	19	14
民進黨	53	51	46	53	48
新黨	0	-	-	1	0
總計	73	73	61	73	62

備註：

1. 除了兩大黨以外，「未來事件交易所」預測金門縣選區為新黨參選人獲勝。
2. 參考資料：蘋果新聞，〈【未來事件交易所】區域立委預測準確度 91.67%〉，<http://www.appledaily.com.tw/realtimenews/article/new/20160119/778824/>，查閱時間：2016 年 10 月 6 日。

上述實證結果雖顯示，我們利用少數的全國性樣本，即可有效地預測 73 個區域立委選舉結果。但究竟多少的全國樣本數才足以讓我們正確地推估這麼多選區的選情？又或者，我們究竟需要多少的全國樣本，才足以讓我們更有效地掌握各個選區 MRP 模擬估計值的誤差範圍？以本文為例，如果以各個選區平均誤差值約 +/-5 個百分點來看，我們用了約 4,000 個全國有效樣本，即可獲得採傳統方式約需要 28,000 個樣本才能獲得的誤差範圍。不過，上述以每個選區平均約 55 個樣本所得出來的結果並不是很穩定，即選區樣本數與誤差值的關係還有待更進一步討論。以下我們就針對各個選區樣本數與估計誤差絕對值之間的關聯性，簡單討論「50 個樣本」是否足夠讓我們利用統計模型有效推估選情？

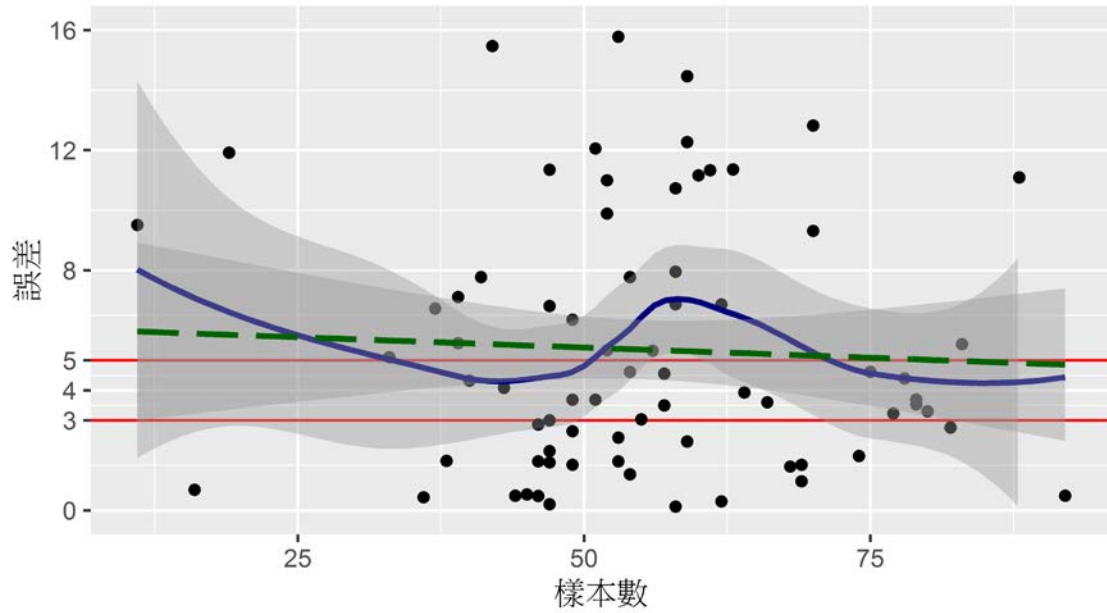
圖二分別是各選區之樣本數與 MRP 模型模擬誤差（百分點）絕對值的散佈圖（左圖為國民黨得票率模擬，右圖為民進黨得票率模擬），另輔以無母數迴歸 (LOWESS)¹¹ 與線性迴歸線 (linear prediction line)。此外，除了標上模擬誤差絕對值與樣本數的關係外，圖二還加上了誤差絕對值為 3 及 5 的輔助線（即 $y=3$ 與 $y=5$ ），協助我們說明樣本數與模擬誤差絕對值之間的關係。

圖二 模擬誤差絕對值與樣本數散佈圖（暨無母數迴歸曲線（藍色實線）與線性迴歸線（綠色虛線））

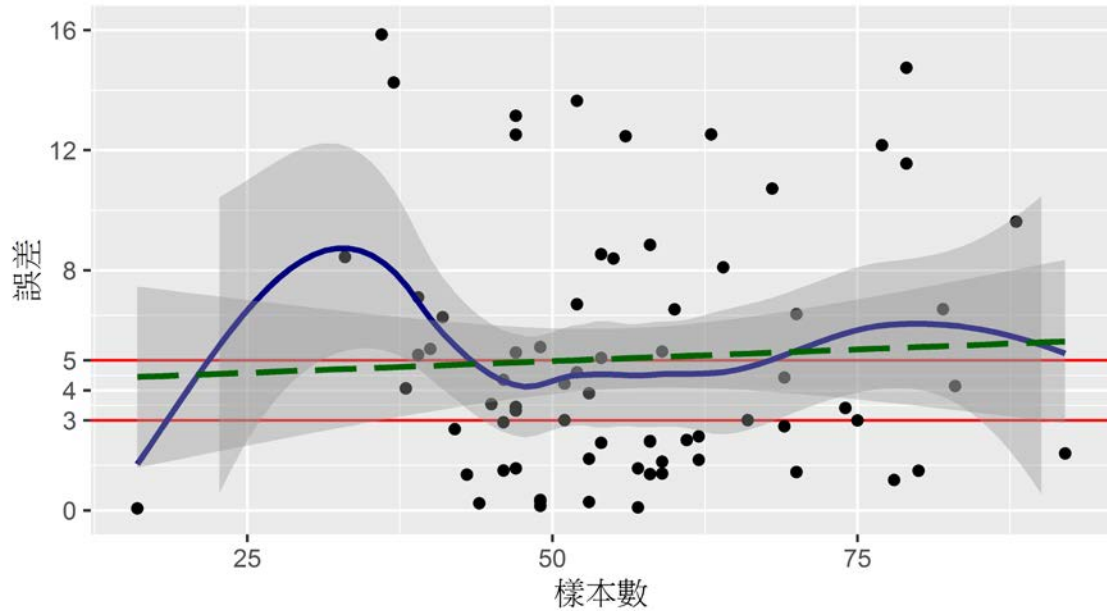
¹⁰ 有關兩種預測方式預測錯誤的選區列表，請見附錄中的附表五。

¹¹ LOWESS 係為 Locally Weighted Scatterplot Smoothing，其為使用 K-近鄰算法 (K-Nearest Neighbors algorithm) 建立模型的無母數迴歸 (non-parametric regression)。研究者透過 LOWESS 可以在建立線性迴歸模型前觀察 x 與 y 的關係，適於資料探勘上的研究。

國民黨



民進黨



備註：民進黨模型不包含金門縣與連江縣選區。

根據圖二中的線性迴歸線（即綠色虛線），首先我們發現，上圖中選區中的樣本數與國民黨得票率模擬誤差絕對值呈現負向關係，即樣本數愈大誤差愈小。另外，根據無母數迴歸曲線（即藍色的LOWESS實曲線），我們發現當樣本數在0-50之間時，大致呈現樣本數的上升，模擬誤差減少的情形，不過當樣本數在50-75之間時，誤差絕對值反而上升到達一個高峰，顯示有部分選區的模擬誤差並不隨著樣本數上升而下降，儘管最後仍處在震盪收斂的狀態。總之，對於國民黨得票率的模型模擬值而言，樣本數增加雖有助於模型模擬值與實際值之間的差異減少，但基本上用50個樣本數的結果仍然不錯，更多樣本數的選區，其模擬誤差值反而存有些許震盪。

反觀民進黨的部份，根據圖二中下圖的線性迴歸線（即綠色虛線），我們發現選區樣本數與民進

黨得票率模擬誤差絕對值的關係為正向關係，即樣本數愈大，誤差絕對值愈大，與預期不符。此外，根據無母數迴歸曲線（即藍色的 LOWESS 實線），我們發現模擬誤差值在樣本數介於 25-50 之間有下降的趨勢。然而在樣本數大於 50 後，則略呈現穩定上升的趨勢，且在樣本數約 75 左右又達到一個高峰，並未出現誤差持續下降或震盪收斂的情況。

總結來說，國民黨與民進黨得票率的模擬誤差值在 0-50 個樣本時，大致呈現下降的趨勢，符合樣本數提升有助於減少估計誤差的預期。不過，當樣本數超過 50 個後，誤差絕對值波動的趨勢並不一致。當然，上述樣本數與誤差絕對值之間的關係只是初步的觀察，我們仍舊無法歸納出一個妥適的經驗準則。但至少有一點可以確定的是：就本研究的資料而言，選區的樣本數愈大時，估計誤差值不見得會比較小。當然，未來還需要更精緻的模擬及更深入地分析，才有辦法確切找出樣本數與 MRP 模擬誤差絕對值之間的關係。

最後，我們提出未來改善本研究的幾個方向及本文結果可供延伸的研究議題：

- 第一、 針對本研究所採用的貝式多層次模型而言，各迴歸係數的先驗（事前）機率分配皆採用無資訊的先驗分佈。然而，我們可以在實際執行電訪民調預測前，先進行小規模的選情測試，譬如在選舉前兩個月，透過小規模（800 個樣本）的網路調查資料，先估計各自變數的係數分配。然後，將這些利用網路調查所得來的係數分配，當作是實際電訪選舉預測模型的先驗（事前）機率分配。倘若網路調查的受訪者和電訪受訪者的投票模式有差異，則利用這樣的方式等於也協助我們調合兩者間的落差。此外，由於我們將網路調查與電訪調查的時間點錯開（譬如相差一個月），這項作法也等同調合了不同時間點的結果（Tsai 2005）。
- 第二、 針對選舉預測的相關研究而言，近年來網路大數據分析在選戰中的應用愈來愈普及，但該類型分析結果仍受限於大眾對於所謂「數位落差」的疑慮，即並非所有人都能夠接觸到網路資訊，也非所有人都自願在網路上表態。也因為這項可能的「選擇性偏誤」，使得依靠網路數據所進行的選舉預測較欠缺公信力。不過，網路所能提供的資訊仍遠比民調多很多，資料搜集成本也較民調低，且往往更能呈現選戰的動態。因此如能將大數據分析結果與隨機抽樣的民調資料相結合，相信對於選舉預測的研究將更有助益。而貝式統計分析的邏輯強調正好適用於將不同類型的資料相結合，求得更全面的「事後分配」（張順全、莊文忠 2008）。
- 第三、 本研究的貢獻除了強調減少民調資料搜集成本外，另一項貢獻是建立整合不同層級或類型資料的模式，即同時以全國性的個體資料及地方層級的總體資料來推估地方民意。過去地方政治（縣市或立委選區層級）的研究受限於調查訪問樣本數的不足，使得我們不容易將全國性的民意調查資料解組（disaggregate）成地方民意，因此少有探討地方民意的研究，更遑論分析地方民意與政策的關係（蔡佳泓、俞振華 2011）。本研究再度利用有「底牌」的選舉預測研究，驗證多層次模型配合事後分層加權（MRP）模式在台灣的適用性，希冀有助於日後與地方（或選區）民意相關的研究。

參考文獻

中文部份：

- 李錦河、溫敏杰，1998，〈從行銷學「產品屬性」角度建構「選民需求指標」選舉預測模式—以1997年台南市市長選舉為例〉，《選舉研究》，5（2）：1-33。
- 李錦河、溫敏杰、陳盈太，2010，〈「品牌知名度」理念應用於選舉預測之探討—以台灣選舉民意調查資料為例〉，《選舉研究》，17（1）：1-20。
- 李皇萱，2014，〈民調預測失準 年輕人投票率衝高是關鍵〉，《聯合晚報》。
<http://udn.com/NEWS/BREAKINGNEWS/BREAKINGNEWS1/9104399.shtml>，查閱時間：2016年6月2日。
- 洪永泰，1994，〈選舉預測：一個以整體資料為輔助工具的模式〉，《選舉研究》，1（1）：93-110。
- 俞振華、蔡佳泓，2006，〈如何利用全國性民調推估地方民意？多層次貝式定理估計模型與分層加權的應用〉，《台灣政治學刊》，10（1）：5-38。
- 俞振華，2013，〈網路民意調查的理論與實務〉，陳陸輝主編，《民意調查新論》：89-110，台北：五南。
- 徐永明，2005，〈時序模型在選舉預測上的應用：以美國、台灣總統選舉為例〉，《調查研究：方法與應用》，17：111-49。
- 梁世武，1994，〈1994台北市長選舉之預測：「候選人形象指標」預測模型之驗證〉，《選舉研究》，1（2）：97-130。
- 莊文忠，2000，〈選舉預測與策略性投票：2000年總統大選之個案研究〉，《理論與政策》，14（2）：55-92。
- 張順全、莊文忠，2008，〈探索選民的投票行為變化：應用機率分配模型的預測方法〉，《選舉研究》，15（2）：91-117。
- 許勝懋，2015，〈「唯手機族」對未來電話調查的影響與啟發〉，《調查研究：方法與應用》，34：33-65。
- 黃紀、張佑宗，2003，〈樣本代表性檢定與最小差異加權：以2001年台灣選舉與民主化調查為例〉，《選舉研究》，10（2）：1-35。
- 黃智聰、程小綾，2005，〈經濟投票與政黨輪替—以台灣縣市長選舉為例〉，《選舉研究》，12（2）：45-78。
- 雷飛龍、陳義彥、丁庭宇，1985，〈民國七十二年台灣地區增額立委選舉之人文區位研究〉，《政大選研中心專刊第1集》，台北：政治大學選舉研究中心。
- 蔡佳泓，2008，〈反貪倒扁運動的支持度之多層次貝式定理分析〉，《政治學報》，45：67-93。
- 蔡佳泓，2014，〈現任縣市長支持之研究：多層次貝氏統計之應用〉，《應用經濟論叢》，96：69-104。
- 蔡佳泓、俞振華，2011，〈地方政府如何回應民意？以2006-2007年為例〉，《台灣政治學刊》，15（1）：73-136。
- 劉文卿，1995，〈台北市長選舉之基因預測模型〉，《選舉研究》，2（1）：1-16。
- 劉念夏，1996，〈一九九六年總統大選選舉預測：民意調查中未表態選民投票行為規則假設的提出與驗證〉，《選舉研究》，3（2）：131-56。
- 劉義周，1996，〈選舉預測：一組簡單理論的檢驗〉，《選舉研究》，3（2）：107-30。
- 劉義周，2009，〈緒論：選舉的預測〉，游清鑫、蔡佳泓主編，《選舉預測》：1-12，台北：五南。
- 鮑彤，1999，〈選舉地盤：候選人票源凝聚程度之分析〉。台北：國立政治大學政治學研究所碩士論文。
- 盛杏媛，1998，〈投票選擇與選舉預測〉，《選舉研究》，5（1）：37-76。
- 盛杏媛、周應龍，2008，〈選樣偏誤模型在調查研究中項目無反應問題的應用〉，《台灣政治學刊》，12

(1): 147-83。

盛治仁，2000，〈總統選舉預測探討-未表態選民與情感溫度計〉，《選舉研究》，7 (2): 75-107。

童振源、林馨怡、林繼文、黃光雄、周子全、劉嘉凱、趙文志，2009，〈台灣選舉預測：預測市場的運用與實證分析〉，《選舉研究》，16 (2): 131-66。

外文部份：

Abramowitz, Alan I. 1988. "An Improved Model for Predicting Presidential Election Outcomes." *PS: Political Science & Politics* 21: 843-847.

Buchanan, William. 1986. "Election Prediction: An Empirical Assessment." *Public Opinion Quarterly* 50: 222-27.

Fiorina, Morris P. 1977. *Congress: Keystone of the Washington Establishment*. Yale University Press.

Gelman, Andrew, and Thomas C. Little. 1997. "Postratification into Many Categories Using Hierarchical Logistic Regression." *Survey Methodology* 23: 127-35.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo," *American Journal of Political Science* 44(2): 369-398.

Jackman, Simon. 2009. *Bayesian Statistics for Social Science*. New York: John Wiley & Sons.

Cox, Gary W. and Jonathan N. Katz. 1996. "Why Did the Incumbency Advantage in U.S. House Elections Grow?" *American Journal of Political Science* 40(2): 478-97.

Kastellec, Jonathan P., Jeffrey R. Lax and Justin H. Phillips. 2010. "Public Opinion and Senate Confirmation of Supreme Court Nominees." *Journal of Politics* 72:767-84.

Kelley, Stanley, Jr. and Thad W. Mirer. 1974. "The Simple Act of Voting." *The American Political Science Review* 68(2): 572-591.

Kramer, Gerald. 1971. "Short-Term Fluctuations in U.S. Voting Behavior, 1896-1964." *American Political Science Review* 65(1):131-143.

Lewis-Beck, Michael S. and Tom W. Rice. 1992. *Forecasting Elections*. Washington, D.C.: CQ Press.

Lavine, Michael. 1999. "What Is Bayesian Statistics and Why Everything Else Is Wrong." *The Journal of Undergraduate Mathematics and Its Applications* 20: 165-174.

Lax, Jeffrey R. and Justin H. Phillips. 2009a. "Gay Rights in the States: Public Opinion and Policy Responsiveness." *American Political Science Review* 103(3): 367-86.

Lax, Jeffrey R. and Justin H. Phillips. 2009b. "How Should We Estimate Public Opinion in the States?" *American Journal of Political Science* 53(1): 107-21.

Lax, Jeffrey R. and Justin H. Phillips. 2012. "The Democratic Deficit in the States" *American Journal of Political Science* 56(1): 148-66.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimate from National Polls." *Political Analysis* 12(4): 375-385.

Spiegelhalter, D., A. Thmomas, and N. Best. 1999. *WinBugs Version 1.4*. Cambridge: MRC biostatistics Unit.

Rao, J. N. K. 2003. *Small Area Estimation*. New York: Wiley InterScience.

Rosenstone, Steven J. 1983. *Forecasting Presidential Elections*. New Haven : Yale University Press.

Tsai, Chia-hung. 2000. "American Voter Responses to International Political Events and Economic Conditions: 1920-1996," *Euramerica* 30(3): 143-191.

- , 2005. "Bayesian Inference in Binomial Logistic Regression: A Case Study of the 2002 Taipei Mayoral Election." *Journal of Social Sciences and Philosophy* 17(1) : 103-23
- Tufte, Edward R. 1978. *Political Control of the Economy*. Princeton, N.J. : Princeton University Press.
- Warshaw, C. and J. Rodden, 2012, "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74: 203-219.
- Western, Bruce. 1998. "Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modeling Approach," *American Journal of Political Science* 42: 1253-1259.

科技部補助專題研究計畫出席國際學術會議心得報告

日期：105 年 6 月 06 日

計畫編號	104-2410-H-004 -090 -		
計畫名稱	再探全國性民調推估地方民意的可行性：應用改良式多層次貝氏定理估計模型及事後分層加權預測立法委員選舉結果		
出國人員姓名	俞振華	服務機構及職稱	政治大學選舉研究中心副研究員
會議時間	105 年 5 月 14 日至 105 年 5 月 15 日	會議地點	日本東京日本大學
會議名稱	(中文)2016 日本選舉研究協會年會 (英文)2016 Annual Meeting of Japanese Association of Electoral Studies		
發表題目	(中文) 解析台灣主要政黨候選人選擇機制 (英文) Analyzing Candidate Selection Method in Taiwan's Major Parties		

一、參加會議經過

此次參與 2016 年日本選舉研究協會年會，主要是發表 “Analyzing Candidate Selection Method” 一文。此次研究者參與了該協會的 international panel，論文的評論人有三位，分別是國際基督教大學的石生義人教授、愛媛大學的梅田道生教授、及 University of South Carolina 的 John Hsieh 教授。三位評論人對於本文都提供了許多修改意見，並鼓勵日後採用比較研究的觀點。此外，三位學者都認為，台灣政黨的利用電話民調來候選人相當特別，對於政黨日後發展的影響，值得再更深入的探討。

二、與會心得

本人每年都會在五月期間，參與日本選舉研究協會年會的活動，並與日本學者互動。過去日本學者或許受限於英語交流能力，學術社群比較封閉。但現在日本留美的學者愈來愈多，國際化的程度也很快地提昇。此外，日本與台灣的國會及地方議會選舉制度相近，有很多可以比較的主題。過去本人就已有和日本學者共同探討選舉制度對於政黨候選人甄補的影響，政大選舉研究中心也出版了英文專書，其中也深刻地比較了台灣與日本選舉制度變遷後的現象。總之，在選舉研究方面，本人認為台灣學者與日本學者都還有很多可以展開的研究議題，值得持續深耕。

三、發表論文全文或摘要

Analyzing Candidate Selection Methods in Taiwan's Major Political Parties

Abstract

This study analyzes the transformation of candidate selection methods in Taiwan's major political parties for legislative elections. Particularly, it pays special attention on the current system, “polling primary,” and discusses its pros and cons with respect to party development. Our findings are threefold: first, the changes of the candidate selection methods for both parties have mainly followed Taiwan's democratic pace and become more and more decentralized. Specifically, more and more selectorates have been included in the selection process as the nomination systems have evolved. Second, polling primary, the way that a party delegates power to the general public to choose its candidates, serves two major purposes—that is, to select the most electable candidates as well as to resolve possible intra-party confrontations. Third, while the two major

parties cannot figure out a better way to solve their “fake membership” problems, polling primary will be regarded as the default system to nominate candidates for both parties in the near future even though it has its own problems.

四、建議

今年我們台灣學者是以組 panel 的方式參與日本選舉協會年會。為了擴大影響力，本人認為日後若能採與日本學者合寫英文論文，但以日文發表的方式，效果可能更好。畢竟用日文發表，還是能夠協助我們與更多日本學者對話。

五、攜回資料名稱及內容

無

六、其他

無

科技部補助計畫衍生研發成果推廣資料表

日期:2017/01/18

科技部補助計畫	計畫名稱: 再探全國性民調推估地方民意的可行性:應用改良式多層次貝氏定理估計模型及事後分層加權預測立法委員選舉結果
	計畫主持人: 俞振華
	計畫編號: 104-2410-H-004-090- 學門領域: 政治理論
無研發成果推廣資料	

104年度專題研究計畫成果彙整表

計畫主持人：俞振華			計畫編號：104-2410-H-004-090-			
計畫名稱：再探全國性民調推估地方民意的可行性:應用改良式多層次貝氏定理估計模型及事後分層加權預測立法委員選舉結果						
成果項目			量化	單位	質化 (說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等)	
國內	學術性論文	期刊論文		0	篇 本 章 篇 篇	2016年日本選舉協會年會、2016年台灣政治學會年會
		研討會論文		2		
		專書		0		
		專書論文		0		
		技術報告		0		
		其他		0		
	智慧財產權及成果	專利權	發明專利	申請中	0	件
				已獲得	0	
			新型/設計專利		0	
		商標權		0		
		營業秘密		0		
		積體電路電路布局權		0		
		著作權		0		
		品種權		0		
		其他		0		
	技術移轉	件數		0	件	
		收入		0	千元	
	國外	學術性論文	期刊論文		0	篇 本 章 篇 篇
			研討會論文		0	
			專書		0	
專書論文			0			
技術報告			0			
其他			0			
智慧財產權及成果		專利權	發明專利	申請中	0	件
				已獲得	0	
			新型/設計專利		0	
		商標權		0		
		營業秘密		0		
		積體電路電路布局權		0		
		著作權		0		

		品種權	0		
		其他	0		
	技術移轉	件數	0	件	
		收入	0	千元	
參與計畫人力	本國籍	大專生	0	人次	
		碩士生	1		
		博士生	0		
		博士後研究員	0		
		專任助理	0		
	非本國籍	大專生	0		
		碩士生	0		
		博士生	0		
		博士後研究員	0		
		專任助理	0		
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)					

科技部補助專題研究計畫成果自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否具有政策應用參考價值及具影響公共利益之重大發現）或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

達成目標

未達成目標（請說明，以100字為限）

實驗失敗

因故實驗中斷

其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形（請於其他欄註明專利及技轉之證號、合約、申請及洽談等詳細資訊）

論文： 已發表 未發表之文稿 撰寫中 無

專利： 已獲得 申請中 無

技轉： 已技轉 洽談中 無

其他：（以200字為限）

已在研討會發表，並已修改投稿學術期刊。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性，以500字為限）

學術成就

本研究只使用約4,000個全國有效樣本，預測73個立委選區選情。具體來說，不論是針對國民黨或民進黨候選人，本研究的模型模擬值與實際值的平均差距約為5個百分點，遠較使用傳統加權方法的9個百分點差距（與實際值相比）來得優異。

技術創新

台灣立委選舉有73個單一選區，倘若我們要以傳統分區民調方式預測73個選區的選舉結果，根據機率抽樣的理論，樣本數開根號後與抽樣誤差成反比，於是在特定誤差範圍的要求下（譬如 ± 5 個百分點），我們可能需要超過2萬8千筆樣本（ $385 \times 73 = 28105$ ）。這樣龐大的樣本數當然不是不可能達成，只是不論採用何種調查方式，皆肯定所費不貲。本研究只採用了約4000個樣本數，即得到不錯的預測效果，精進學界及實務界的選舉預測技術。

社會影響

由於本研究所使用的方法能將預測立委選舉的成本降低，自然有助於實務界或政黨採用此方法，掌握立委候選人選情。另外，這項方法也可適用於政策研究，特別是透過全國性的調查來分析地方民意。

4. 主要發現

本研究具有政策應用參考價值：否 是，建議提供機關中央政府機關
(勾選「是」者，請列舉建議可提供施政參考之業務主管機關)

本研究具影響公共利益之重大發現：否 是

說明：(以150字為限)

過去地方政治(縣市或立委選區層級)的研究受限於調查訪問樣本數的不足，使得我們不容易將全國性的民意調查資料解組成地方民意，因此少有探討地方民意或政策的研究。本研究再度利用有「底牌」的選舉預測研究，驗證多層次模型配合事後分層加權(MRP)模式在台灣的適用性，希冀有助於日後與地方(或選區)民意相關的研究。