

第四章 選樣偏誤模型在總統選舉的應用

在上一章中，已經證實選樣偏誤模型在勝負接近的地方首長選舉中，是可以適用的。而在本章中，筆者選擇二次總統選舉建立選樣偏誤模型，分別是2000年及2004年總統選舉，進一步檢驗選樣偏誤模型在總統選舉的表現。

第一節 2000 年總統選舉

2000年總統選舉是第二次的民選總統選舉，在該次選舉中，共有五組候選人角逐總統的寶座。國民黨推出的候選人是當時的副總統連戰以及行政資歷豐富的蕭萬長。民進黨籍的候選人是曾擔任台北市長的陳水扁與曾擔任桃園縣長的呂秀蓮，新黨則是推派李敖與馮滬祥參選。另外還有曾擔任過第一屆民選省長，脫離國民黨自行參選的宋楚瑜，其搭檔為張昭雄，最後一組候選人則是曾經擔任過民進黨主席的許信良，搭配朱惠良參選。不同於 1996年李登輝一枝獨秀的情況，在2000年總統選舉中，呈現連蕭、陳呂、宋張三強鼎立的局面，最後的選舉結果是，陳水扁、呂秀蓮獲得39.3%的選票，宋楚瑜、張昭雄得到36.84%的選票，連戰、蕭萬長的得票率則為23.1%，許信良、朱惠良只獲得0.63%的選票，李敖與馮滬祥的得票率更低，僅有0.13%，由陳水扁、呂秀蓮以相對多數當選。

表 4 - 1 是電話訪問中，選民投票意向的次數分配表，其中有將近七成（69.74%）的選民願意表態，而有略高於三成（30.26%）的受訪者不願意表態，相對於地方性的選舉來說，不表態的比例要高出許多，可能是受訪者尚未確定投票對象，以及有棄保效應的考量。三組候選人的實力相當接近，有23.74%

的受訪者要投票給宋楚瑜、張昭雄，願意投票給連戰、蕭萬長的受訪者有21.12%，而陳水扁、呂秀蓮的支持率則是24.88%，在該次電話訪問中，並沒有受訪者回答要投票給許信良、朱惠良或是李敖、馮滬祥這兩組候選人。由於研究方法上的限制，本研究只能在應變數是二分變數的情況下進行分析，但是三組人馬的實力又很接近，無法採取類似2001年台北縣長選舉及2002年高雄市長選舉的處理方式，將實力較弱的候選人予以排除，因此在不得已的情況下，筆者只能退而求其次，將連蕭、陳呂、宋張三組候選人分成「投陳呂」、「不投陳呂」兩類。在表態的受訪者當中，有三成五左右（35.68%）的比例表示會投給陳呂，有六成五左右（64.32%）的選民不會投票給陳呂。在看完初步的投票意向之後，接下來則進入選樣方程式的建構。

表 4 - 1 2000年總統選舉選民投票意向次數分配表

	次數	百分比		百分比	候選人支持率 相對百分比
宋楚瑜、張昭雄	353	23.74	不投陳呂	44.86%	64.32%
連戰、蕭萬長	314	21.12			
陳水扁、呂秀蓮	370	24.88	投陳呂	24.88%	35.68%
不表態	450	30.26	不表態	30.26%	
合計	1487	100.0		100.0%	100.0%

註：本次電訪訪問中，沒有受訪者回答要投票給其他候選人。

一、選樣方程式

在建構選樣方程式的變數上，筆者同樣選擇了性別、年齡、教育程度、媒體使用情形、是否有政黨認同、候選人評價問題回答情形等六個變數，在實際進行選樣方程式的模型建構之前，筆者先透過自變數與是否表態的交叉列表及卡方檢定，來初步觀察這六個自變數和選民是否表態之間的關係。

表 4 - 2 2000 年總統選舉選樣方程式自變數與是否表態交叉列表

自變數	不表態	表態	(樣本數)	卡方檢定結果
性別	女性	35.7	64.3	(802) $\chi^2=24.043$ d.f=1
	男性	23.9	76.1	(685) $p<.001$
年齡	20 至 29 歲	18.6	81.4	(269)
	30 至 39 歲	22.7	77.3	(458) $\chi^2=72.075$
	40 至 49 歲	34.3	65.7	(431) d.f=4
	50 至 59 歲	35.3	64.7	(167) $p<.001$
	60 歲以上	53.5	46.5	(142)
教育程度	低	44.0	56.0	(420) $\chi^2=59.793$
	中	27.5	72.5	(553) d.f=2
	高	21.2	78.8	(509) $p<.001$
媒體使用情形	報紙電視都不看	52.2	47.8	(46) $\chi^2=76.809$
	看其中一種	48.1	51.9	(308) d.f=2
	報紙電視都看	24.2	75.8	(1122) $p<.001$
是否具有政黨認同	無	52.6	47.4	(542) $\chi^2=201.338$ d.f=1
	有	17.5	82.5	(945) $p<.001$
候選人評價問題回答情形	三題都沒有回答	77.2	22.8	(136)
	回答一題	53.8	46.2	(93) $\chi^2=225.885$
	回答二題	42.0	58.0	(169) d.f=3
	三題都回答	20.6	79.4	(1089) $p<.001$

註：表中所列為橫列百分比（括弧內為樣本數）。

從表 4 - 2 的數據可以知道，女性不表態的比例(35.7%)要比男性(23.9%)來的高，與理論預期相符。年齡愈大的選民愈不願意表態，60歲以上的選民有超過一半(53.5%)的比例不願意表態，而20至29歲的選民僅有不到二成(18.6%)的比例不願意表態。教育程度在是否表態上的差異是，教育程度愈低的選民愈不願意表態，低教育程度選民不願意表態的比例為44.0%，中教育程度選民的不表態比例略低於低教育程度選民，有27.5%，而高教育程度選民的不表態比例最低，僅略超過二成一些(21.2%)。在媒體使用情形方面，同樣呈現與理論預期的方向，報紙、電視都不看的選民，其不表態率超過五成(52.2%)，而報紙、電視都看的選民，其不表態比例則只有二成五左右(24.2%)，比都不看

的選民低了許多。有無政黨認同在是否表態上也具有顯著的差異，沒有政黨認同的選民中，不表態的比例超過五成（52.6%），而有政黨認同的選民裡面，不表態的比例則不到一成八（17.5%），差異非常的大。候選人評價問題回答方面，三個題目都沒有回答的選民中，超過七成七（77.2%）沒有表達他的投票意願，反觀三個題目都回答的選民，則僅有二成左右的比例（20.6%）沒有表達其投票意願，差異也是相當的大。

在確定了各個自變數與是否表態之間的關係都如理論預期，且都達到統計上的顯著水準之後，接下來便是建立本年度選樣方程式的probit模型，模型的應變數為選民是否表態，表態為1，不表態為0。自變數中的性別、教育程度、是否具有政黨認同是類別變數，因此以虛擬變數的方式放入模型中，性別區分為男性與女性，比較基礎為女性；教育程度有高教育程度、中教育程度、低教育程度三類，比較基礎為低教育程度；是否具有政黨認同區分為有政黨認同與沒有政黨認同，比較基礎為沒有政黨認同；另外三個自變數：年齡、媒體使用情形與候選人評價問題回答情形，是以數字資料的方式放入模型中。為了比較各個自變數在模型中的影響力大小，筆者將不是虛擬變數的年齡、媒體使用情形、候選人評價問題回答情形這三個變數轉化為在0-1之間變動，以便能夠比較各個自變數的影響力大小。分析的結果詳見表4-3。

從表4-3各變數的係數可以發現，各變數的係數值與理論預期的方向完全相符，且都達到統計上的顯著水準。其中，對候選人評價問題的回答情形這個變數，可說是影響選民是否表態的最重要因素，對候選人評價問題回答愈多題，則表態的機率愈高；其次則是年齡的影響，年齡愈大的選民愈不願意表態；影響力第三大的是具不具有政黨認同，有政黨認同的選民比沒有政黨認同的選民更傾向表態。整個模型的正確預測率是78.55%。接下來，筆者便將以這個模型做為選樣偏誤模型中的選樣方程式，建構選樣偏誤模型。

表 4-3 2000 年總統選舉選樣方程式 probit 模型

模型 I	
性別（女性=0）	
男性	.310 (.081) ***
年齡	-1.245 (.238) ***
教育程度（低教育程度=0）	
中教育程度	.0674 (.107)
高教育程度	.240 (.113) *
媒體使用情形	.671 (.163) ***
是否具有政黨認同（無=0）	
有政黨認同	.749 (.083) ***
候選人評價回答情形	1.155 (.130) ***
常數	-1.325 (.210) ***
分析個數	1389
正確預測率（%）	78.55
-2 Log Likelihood	1320.1204

註：1.應變數編碼方式：1-表態，0-不表態。

2.表中數字為probit模型分析的估計值，括弧中的數字為標準誤。

3. ***表 $p < .001$ ，**表 $p < .01$ ，*表 $p < .05$ ，^s表 $p < .1$ 。

二、選樣偏誤模型

在2000年總統選舉的選樣偏誤模型中，結果方程式的應變數是選民投票對象（陳水扁=1、非陳水扁=0），自變數除了地方首長選舉中的候選人評價、政黨認同及省籍之外，還加上統獨立場，其中候選人評價、政黨認同、省籍、統獨立場都是類別變數，筆者將以虛擬變數的方式放入模型中，候選人評價區分為對連戰評價最高、對宋楚瑜評價最高、對陳水扁評價最高、宋扁同，對連戰評價最低、連扁同，對宋楚瑜評價最低、連宋同，對陳水扁評價最低、以及無法比較高低等七類，比較基礎為對陳水扁評價較高；政黨認同有認同泛藍、認同泛綠及中立無反應三類，比較基礎為認同泛綠；省籍區分為本省

人與外省人，比較基礎為外省人；統獨立場有傾向獨立、傾向統一、維持現狀三類，比較基礎為傾向獨立。由於這幾個變數在過去的研究中，都已經被證實對於投票行為具有理論上的影響力，因此筆者就不進行個別自變數對應變數的卡方檢定，直接進入模型建構。

模型分析的結果見於表 4 - 4，其中選樣偏誤模型是考慮選樣偏誤之後，依據Dubin與Rivers的方法，將選樣方程式與結果方程式同時進行估計的模型，而傳統probit模型則是不考慮選樣偏誤，直接以結果方程式所建立的probit模型。

首先我們觀察表 4 - 4 中，各變數的係數方向都與理論預期方向相同，而且除了省籍外，每個變數的影響都達到統計上的顯著水準。 ρ 等於-0.732， ρ 小於0的意思是在不考慮選樣偏誤的情況下，將低估結果方程式中， $Y=1$ 發生的機率，也就是說將低估選民投票給陳水扁的機率。對於 $\rho=0$ 的概似比檢定（likelihood ratio test, LR）結果也顯示，若只單純對結果方程式進行估計，會因為選樣偏誤而造成估計偏差的情況。

為了瞭解自變數對選民投票抉擇的影響，以及經過校正選樣偏誤之後，參數估計值變動所造成的影響，筆者計算出當自變數被控制在某些特定的情況下，選民投給每一個候選人的機率，詳見附錄一的表D。

觀察模型中參數估計值的變動，在候選人評價方面，傳統 probit 模型高估了給連戰較高評價者比給陳水扁較高評價者傾向不投票給陳呂的強度，校正選樣偏誤後的參數估計值從-2.344 減弱為-2.074，意味著在未校正選樣偏誤的情況下，模型高估給連戰較高評價者與給陳水扁較高評價者之間的差距；傳統 probit 模型也高估了給宋楚瑜較高評價者比給陳水扁較高評價者傾向不投票給陳呂的強度，而且高估的程度蠻多的，表示在校正選樣偏誤之後，兩者的差距其實沒有那麼大。在無法比較候選人評價高低者的影響上，未校正選樣偏誤的傳統 probit 模型高估了無法比較候選人評價高低者比給陳水扁最高

評價者更傾向不投票給陳呂的強度，係數由-1.594 減少為-1.285，表示在校正選樣偏誤之後，無法比較候選人評價者與給陳水扁最高評價者的距離減少了0.309，變動幅度還算蠻大的，由於這個變數是虛擬變數，因此我們可以利用表中變數的參數估計值，分別計算出無法比較候選人評價高低者與給連戰最高評價者、給宋楚瑜最高評價者之間的距離，若以給連戰最高評價者為比較基礎的話，在傳統 probit 模型中，無法比較候選人評價高低者的參數估計值是0.75 (-1.594 + 2.344)，在選樣偏誤模型中的參數估計值是 0.789 (-1.285 + 2.074)，表示在校正選樣偏誤之後，給連戰最高評價者與無法比較候選人評價高低者之間的差距增加了一點點，若以給宋楚瑜最高評價者為比較基礎的話，在傳統 probit 模型中，無法比較候選人評價高低者的參數估計值是 1.544 (-1.594 + 3.138)，在選樣偏誤模型中的參數估計值是 1.509 (-1.285 + 2.794)，表示在校正選樣偏誤之後，給宋楚瑜最高評價者與無法比較候選人評價高低者之間的差距減少了一點點，綜合來看，在校正選樣偏誤之後，連戰、宋楚瑜最高評價者與無法比較候選人評價高低者的距離沒有什麼變動，而無法比較候選人評價者與給陳水扁最高評價者的距離減少，表示在無法比較候選人評價高低的選民中，不表態選民的投票傾向比表態選民更接近給陳水扁評價較高者，也就是說會比較傾向投票給陳呂，因此若不校正選樣偏誤的話，將低估陳呂的得票率。

在政黨認同參數估計值的變化上，在沒有校正選樣偏誤的情況下，傳統 probit 模型高估了認同泛藍的選民比認同泛綠的選民更傾向不投票給陳呂的強度，未校正選樣偏誤時的參數估計值為-1.857，校正選樣偏誤之後的參數估計值則減少為 1.671，產生這樣的變化，與本研究的預期是相符的，因為我們僅以表態者建立傳統 probit 模型，而有政黨認同的選民又比沒有政黨認同的選民更傾向表態，因此使得傳統 probit 模型高估藍綠認同者之間的差異；接下來我們觀察到，未校正選樣偏誤的傳統 probit 模型高估了中立無反應選民比認同泛綠選民更傾向不投票給陳呂的強度，係數由-0.823 減少為-0.453，表示在校正選樣偏誤之後，中立無反應者與泛綠的距離減少了 0.37，是個不小的變動，由

於這個變數是虛擬變數，因此我們可以利用表中變數的參數估計值計算出中立無反應者與認同泛藍者的距離，若以認同泛藍為比較基礎的話，在傳統 probit 模型中，中立無反應者的參數估計值是 1.034 (-0.823+1.857)，在選樣偏誤模型中的參數估計值是 1.218 (-0.453+1.671)，表示在校正選樣偏誤之後，泛藍與中立無反應之間的差距變大了，綜合來看，校正選樣偏誤之後，泛藍與中立無反應的距離變大，泛綠與中立無反應的距離變小，表示在政黨認同是中立無反應的選民中，不表態者與表態者相比，其投票傾向與認同泛綠者比較接近，會比較傾向投票給陳呂，因此若不校正選樣偏誤的話，將低估陳呂的得票率。

在統獨立場方面，參數估計值的變動並不算太大，傳統probit模型高估傾向統一比傾向獨立的選民更傾向不投票給陳呂的強度，校正選樣偏誤後的參數估計值由-0.729減少為-0.619，而在校正選樣偏誤之後，傾向維持現狀的選民比傾向獨立的選民更傾向不投票給陳呂，其強度也是減弱的。

省籍參數估計值的變動相當小，在校正選樣偏誤之後，省籍的參數估計值從0.310減弱到0.240，僅減少了0.06。

從以上這幾個變數的係數變動程度來看，在2000年的總統選舉中，政黨認同是中立無反應者的變動程度是最多的，而政黨認同是中立無反應的選民也正是在訪問中非常不願意表態的一群人，這表示說如果我們將不願意表態的人忽略不計，便容易錯估政黨認同是中立無反應者的真正投票傾向，當然就會進一步去影響到候選人得票率的預測。無法比較候選人評價高低的變動程度也不小，由於對候選人評價問題的回答情形影響了是否表態，對候選人評價問題都回答的選民較傾向表態，在都回答的情況下，才有可能分出對候選人的評價高低，因此僅以表態者所建立的傳統probit模型，便可能高估給不同候選人較高評價者之間的差異，而省籍影響力的變動最少，可能是因為省籍本身並不太會去影響是否表態，因此他的參數估計值沒有太大的變化。

表 4-4 2000 年總統選舉選民投票預測模型

	選樣偏誤模型 (校正選樣偏誤)	傳統 probit 模型 (未校正選樣偏誤)
結果方程式		
候選人評價(陳水扁最高=0)		
連戰最高	-2.074 (.306) ***	-2.344 (.313) ***
宋楚瑜最高	-2.794 (.335) ***	-3.138 (.316) ***
宋扁同, 連戰最低	-1.267 (.207) ***	-1.357 (.223) ***
連扁同, 宋楚瑜最低	-.994 (.205) ***	-1.090 (.220) ***
連宋同, 陳水扁最低	-2.732 (.486) ***	-3.119 (.536) ***
無法比較高低	-1.285 (.199) ***	-1.594 (.199) ***
藍綠認同(泛綠=0)		
泛藍	-1.671 (.179) ***	-1.857 (.165) ***
中立無反應	-.453 (.173) **	-.823 (.153) ***
省籍(外省=0)		
本省	.240 (.201)	.310 (.153)
統獨立場(傾向獨立=0)		
傾向統一	-.619 (.198) **	-.729 (.220) **
維持現狀	-.361 (.154) **	-.415 (.173) *
常數	2.078 (.276) ***	2.059 (.313) ***
ρ	-.732 (.162)	
選樣方程式		
性別(女性=0)		
男性	.278 (.080) **	
年齡	-1.325 (.233) ***	
教育程度(低教育程度=0)		
中教育程度	.059 (.104)	
高教育程度	.257 (.109) *	
媒體使用情形	.590 (.162) ***	
是否具有政黨認同(無=0)		
有政黨認同	.773 (.083) ***	
候選人評價回答情形	1.126 (.129) ***	
常數	-1.24 (.213) ***	
分析個數	1389	963
-2 Log Likelihood	1800.8246	486.5767
LR test ($\rho = 0$)	5.87*	

註：1.應變數編碼方式：1-陳呂，0-非陳呂。

2.表中數字為probit模型分析的估計值，括弧中的數字為標準誤。

3.***表 $p < .001$ ，**表 $p < .01$ ，*表 $p < .05$ ，^s表 $p < .1$ 。

根據以上兩個模型的參數估計所計算出的預測結果，詳見表 4－5。從表 4－5 中可以得知，傳統 probit 模型的預測結果，在沒有考慮選樣偏誤的問題之下，果然低估了陳呂的得票率，與實際得票率的誤差達到 7.68%，表現比只看電訪表態者的投票對象比例更差，反觀選樣偏誤模型的表現則相當不錯，與實際得票率的誤差只有 0.59%，獲得比傳統 probit 模型更好的預測結果。

表 4－5 2000 年總統選舉選樣偏誤模型、傳統 probit 模型、電訪表態者比較

	實際得票率	電訪表態者	傳統 probit 模型	選樣偏誤模型
投陳呂	39.61	35.68	31.93	40.20
不投陳呂	60.39	64.32	68.07	59.80
與實際誤差	——	3.93	7.68	0.59

註：表中實際得票率為扣除許信良、朱惠良與李敖、馮滬祥票數後的得票率。

第二節 2004 年總統選舉

2004 年的總統選舉，共有兩組參選人，一組是尋求連任，民進黨籍的陳水扁、呂秀蓮，另一組則是由國民黨主席連戰與親民黨主席宋楚瑜所組合成的搭檔。

表 4－6 是選民投票意向的次數分配表，在選前所進行的電話訪問中，有七成左右（70.1%）的受訪者願意表達可能的投票對象，而有 29.90%的受訪者並不願意表態，不表態比例在本研究所分析的五次選舉中，是最高的一次。在表態的受訪者當中有 46.81%要投票給陳呂配，53.19%則是要投票給連宋配，從初步的電訪結果來看，連宋配是處於領先的狀態。接下來則進入選樣方程式的建構。

表 4 - 6 2004年總統選舉選民投票意向次數分配表

	次數	百分比	候選人支持率 相對百分比
陳水扁、呂秀蓮	338	32.82%	46.81%
連戰、宋楚瑜	384	37.28%	53.19%
不表態	308	29.90%	
合計	1030	100.0%	100.0%

一、選樣方程式

在建構選樣方程式的變數上，筆者同樣選擇了性別、年齡、教育程度、媒體使用情形、是否有政黨認同、候選人評價問題回答情形等六個變數，在實際進行選樣方程式的模型建構之前，筆者先透過自變數與是否表態的交叉列表及卡方檢定，來初步觀察這六個自變數和選民是否表態之間的關係。

從表 4 - 7 可以發現，每一個自變數在是否表態上的差異，都達到統計上的顯著水準。在性別方面，男性願意表態的比例要比女性略高一些，分別是 73.6%與 66.9%。年齡層在是否表態上的差異是，年齡愈大的選民愈不願意表態，60 歲以上的選民有超過四成（42.3%）的比例不願意表態，隨著年齡層的降低，不表態的比例也隨之降低，20 至 29 歲的選民僅有二成左右（20.8%）的比例不願意表態。教育程度愈高的選民，愈願意表達自己的投票對象，有將近八成（79.6%）的高教育程度選民願意表態，低教育程度選民表態的比例則僅略超過六成（60.5%）而已。在媒體使用情形方面，資訊接收愈多的選民，不表態的比例愈低，報紙和電視都看的受訪者有二成五左右（25.1%）的比例不願意表態，都不看的受訪者則有超過五成（53.2%）的比例不願意表態。是否具有政黨認同在是否表態上的差異相當的明顯，沒有政黨認同的受訪者中，不表態的比例超過六成以上（62.0%），而有政黨認同者，其不表態比例則僅有一成二左右（12.2%），差距了五成左右。候選人評價問題回答方面，四個

題目都沒有回答的選民有超過七成七(77.5%)的比例沒有表達他的投票意願，隨著回答的題數增多，不表態的比例也隨之降低，四個題目都回答的選民中，僅有不到二成三的比例(22.8%)沒有表達其投票意願，差異也是相當的大。

表 4 - 7 2004 年總統選舉選樣方程式自變數與是否表態交叉列表

自變數	不表態	表態	(樣本數)	卡方檢定結果
性別	女性	33.1	66.9	(541) $\chi^2=5.511$ d.f=1
	男性	26.4	73.6	(489) $p<.05$
年齡	20 至 29 歲	20.8	79.2	(192)
	30 至 39 歲	27.8	72.2	(255) $\chi^2=16.400$
	40 至 49 歲	30.8	69.2	(302) d.f=4
	50 至 59 歲	28.9	71.1	(149) $p<.01$
	60 歲以上	42.3	57.7	(111)
教育程度	低	39.5	60.5	(261) $\chi^2=30.519$
	中	33.1	66.9	(366) d.f=2
	高	20.4	79.6	(398) $p<.001$
媒體使用情形	報紙電視都不看	53.2	46.8	(47) $\chi^2=27.802$
	看其中一種	37.6	62.4	(287) d.f=2
	報紙電視都看	25.1	74.9	(692) $p<.001$
是否具有政黨認同	無	62.0	38.0	(366) $\chi^2=279.421$ d.f=1
	有	12.2	87.8	(664) $p<.001$
候選人評價問題回答情形	四題都沒有回答	77.5	22.5	(89)
	回答一題	76.9	23.1	(13) $\chi^2=144.141$
	回答二題	52.4	47.6	(21) d.f=4
	回答三題	50.0	50.0	(42) $p<.001$
	四題都回答	22.8	77.2	(865)

註：表中所列為橫列百分比(括弧內為樣本數)。

整體來看，所有變數與是否表態之間的關係都如理論所預期的，也都達到統計上的顯著水準，因此這六個變數都將進入選樣方程式的 probit 模型之中，做進一步的分析。模型的應變數為選民是否表態，表態為 1，不表態為 0。自變數中的性別、教育程度、是否具有政黨認同是類別變數，因此以虛擬變

數的方式放入模型中，性別區分為男性與女性，比較基礎為女性；教育程度有高教育程度、中教育程度、低教育程度三類，比較基礎為低教育程度；是否具有政黨認同區分為有政黨認同與沒有政黨認同，比較基礎為沒有政黨認同；另外三個自變數：年齡、媒體使用情形與候選人評價問題回答情形，是以數字資料的方式放入模型中。同樣為了比較各個自變數在模型中的影響力大小，筆者將模型當中以數字資料方式放入的自變數變動範圍轉化成在 0-1 之間，以便能夠比較各個自變數的影響力大小。選樣方程式的 probit 模型分析結果詳見表 4 - 8。

表 4 - 8 當中有兩個模型，在模型 I 當中，除了中教育程度這個虛擬變數的方向與研究預期不一致外，其餘的方向都與研究預期方向相同，不過在控制了其他變數的情況下，年齡及媒體使用情形對於是否表態的影響上，並沒有達到統計上的顯著水準。為了檢證這兩個不顯著的變數是否會造成其他變數估計值的偏誤，因此筆者將這兩個變數刪除之後，再估計模型 II。估計的結果令筆者滿意，觀察跨模型的變數可以發現，除了教育程度的變動稍微大一些外，其餘變數的係數大致上來說都還算穩定，而教育程度的係數變動是往研究預期的方向改變，使得所有變數的係數值與理論預期的方向完全相符，且仍然達到統計上的顯著水準，說明了這些自變數與應變數之間的估計關係應該是沒有問題的。

從表 4 - 8，模型 II 中各變數係數的大小可以發現，是否具有政黨認同這個變數，可說是影響選民是否表態的最重要變數，有政黨認同的選民比沒有政黨認同的選民更傾向表態；影響力居於第二位的則是對候選人評價問題回答情形這個變數，候選人評價問題回答愈多題的選民，其表態的機率愈高；整個模型的正確預測率是 80.82%。接下來，筆者將以模型 II 做為本年度總統選舉選樣偏誤模型中的選樣方程式，來建構選樣偏誤模型。

表 4-8 2004 年總統選舉選樣方程式 probit 模型

	模型 I	模型 II
性別 (女性=0)		
男性	.241 (.103) *	.218 (.099) *
年齡	-.256 (.318)	——
教育程度 (低教育程度=0)		
中教育程度	-.027 (.144)	.044 (.128)
高教育程度	.361 (.151) *	.466 (.132) ***
媒體使用情形	.290 (.184) ***	——
是否具有政黨認同 (無=0)		
有政黨認同	1.392 (.103) ***	1.398 (.101) ***
候選人評價回答情形	1.238 (.210) ***	1.279 (.195) ***
常數	-1.817 (.291) ***	-1.763 (.199) ***
分析個數	934	954
正確預測率 (%)	81.16	80.82
-2 Log Likelihood	801.9864	824.8860

註：1.應變數編碼方式：1-表態，0-不表態。

2.表中數字為probit模型分析的估計值，括弧中的數字為標準誤。

3.***表 $p < .001$ ，**表 $p < .01$ ，*表 $p < .05$ ，^s表 $p < .1$ 。

二、選樣偏誤模型

在 2004 年總統選舉的選樣偏誤模型中，結果方程式的應變數是選民投票對象（陳水扁、呂秀蓮=1、連戰、宋楚瑜=0），放入的自變數是候選人評價、現任者施政滿意度、政黨認同、省籍及統獨立場。其中候選人評價、政黨認同、省籍、統獨立場都是類別變數，筆者將以虛擬變數的方式放入模型中，候選人評價區分為對陳呂評價較高、對連宋評價較高、無法比較三類，比較基礎為對陳呂評價較高；政黨認同有認同泛藍、認同泛綠及中立無反應三類，比較基礎為認同泛綠；省籍區分為本省人與外省人，比較基礎為外省人；統獨立場有傾向獨立、傾向統一、維持現狀三類，比較基礎為傾向獨立；現任

者施政滿意度是以數字資料的方式放入模型中，爲了比較自變數的影響力大小，因此筆者將施政滿意度這個變數的變動範圍轉化爲 0-1 之間。模型結果詳見表 4 – 9。

從表 4 – 9 中可以發現，各變數的係數方向都與理論預期方向相同，也達到統計上的顯著水準。 ρ 等於-0.914， ρ 小於 0 的意思是在不校正選樣偏誤的情況下，將低估結果方程式中， $Y=1$ 發生的機率，也就是說傳統 probit 模型將低估選民投票給陳水扁、呂秀蓮的機率。對於 $\rho = 0$ 的概似比檢定（likelihood ratio test, LR）結果也顯示，若只單純對結果方程式進行估計，會因爲選樣偏誤而造成估計偏差的情況。

同樣爲了瞭解自變數對選民投票抉擇的影響，以及經過校正選樣偏誤之後，參數估計值變動所造成的影響，筆者計算出當自變數被控制在某些特定的情況下，選民投給每一個候選人的機率，詳見附錄一的表 E。

觀察模型中參數估計值的變動，可以發現，政黨認同參數估計值的變化相當大，在沒有校正選樣偏誤的情況下，傳統 probit 模型高估了認同泛藍的選民比認同泛綠的選民更傾向投票給連宋的強度，未校正選樣偏誤時的參數估計值爲-2.158，校正選樣偏誤之後的參數估計值則減少爲-1.762，在校正選樣偏誤之後，藍綠認同者的距離縮小 0.396；接下來我們觀察到，未校正選樣偏誤的傳統 probit 模型高估了中立無反應選民比認同泛綠選民更傾向投票給連宋的強度，係數由-1.587 減少爲-.713，表示在校正選樣偏誤之後，中立無反應者與泛綠的距離減少了 0.874，是個相當大的變動，由於這個變數是虛擬變數，因此我們可以利用表中變數的參數估計值計算出中立無反應者與認同泛藍者的距離，若以認同泛藍爲比較基礎的話，在傳統 probit 模型中，中立無反應者的參數估計值是 0.694（-1.587 + 2.281），在選樣偏誤模型中的參數估計值是 1.233（-0.713 + 1.946），表示在校正選樣偏誤之後，泛藍與中立無反應之間的差距增加了 0.539，綜合來看，校正選樣偏誤之後，泛藍與中立無反應的距離

變大，泛綠與中立無反應的距離變小，表示在政黨認同是中立無反應的選民中，不表態者與表態者相比，其投票傾向與認同泛綠者比較接近，會比較傾向投票給陳呂配，因此若不校正選樣偏誤的話，將低估陳呂配的得票率。

在候選人評價方面，傳統 probit 模型高估了給連宋較高評價者比給陳呂較高評價者傾向投票給連宋的強度，校正選樣偏誤後的參數估計值從-2.158 減弱為-1.762，意味著在未校正選樣偏誤的情況下，模型高估給連宋較高評價者與給陳呂較高評價者之間的差距；在無法比較候選人評價高低者的影響上，未校正選樣偏誤的傳統 probit 模型高估了無法比較候選人評價高低者比給陳水扁較高評價者更傾向投票給連宋的強度，係數由-1.151 減少為-0.711，表示在校正選樣偏誤之後，無法比較候選人評價者與給陳呂較高評價者的距離減少了 0.44，這個變動程度不算少，由於這個變數是虛擬變數，因此我們可以利用表中變數的參數估計值，計算出無法比較候選人評價高低者與給連宋較高評價者的距離，若以給連宋較高評價者為比較基礎的話，在傳統 probit 模型中，無法比較候選人評價高低者的參數估計值是 1.007 (-1.151+2.158)，在選樣偏誤模型中的參數估計值是 1.051 (-0.711+1.762)，表示在校正選樣偏誤之後，給連宋較高評價者與無法比較候選人評價高低者之間的距離是增加的，綜合來看，給連宋較高評價者與無法比較候選人評價高低者的距離增加，而給陳呂較高評價者與無法比較候選人評價高低者的距離減少，表示在無法比較候選人評價高低的選民中，不表態選民的投票傾向比表態選民更接近給陳呂評價較高者，也就是說會比較傾向投票給陳呂配，因此若不校正選樣偏誤的話，將低估陳呂配的得票率。

在陳水扁施政滿意度影響力的變動上，在未校正選樣偏誤的情況下，傳統 probit 模型也是高估此變數的影響力，校正選樣偏誤之後，參數估計值從 2.357 變成 1.785，減少了 0.572，減弱的程度頗大。

表 4-9 2004 年總統選舉選民投票預測模型

	選樣偏誤模型 (校正選樣偏誤)	傳統 probit 模型 (未校正選樣偏誤)
結果方程式		
候選人評價 (陳呂較高=0)		
連宋較高	-1.762 (.311) ***	-2.158 (.337) **
無法比較高低	-.711 (.279) *	-1.151 (.324) ***
陳水扁施政滿意度	1.785 (.528) **	2.357 (.630) ***
藍綠認同 (泛綠=0)		
泛藍	-1.946 (.320) ***	-2.281 (.360) ***
中立無反應	-.713 (.319) *	-1.587 (.316) ***
省籍 (外省=0)		
本省	.939 (.386) *	1.000 (.489) *
統獨立場 (傾向獨立=0)		
傾向統一	-.352 (.365)	-.483 (.485)
維持現狀	-.710 (.253) **	-.836 (.321) **
常數	1.013 (.486) *	.889 (.620)
ρ	-.914 (.119)	
選樣方程式		
性別 (女性=0)		
男性	.207 (.095) *	
教育程度 (低教育程度=0)		
中教育程度	.069 (.124)	
高教育程度	.512 (.127) ***	
是否具有政黨認同 (無=0)		
有政黨認同	1.408 (.101) ***	
候選人評價回答情形	1.163 (.194) ***	
常數	-1.683 (.197) ***	
分析個數	954	649
-2 Log Likelihood	935.9964	114.8409
LR test ($\rho = 0$)	3.73 ^s	

註：1.應變數編碼方式：1-扁呂，0-連宋。

2..表中數字為probit模型分析的估計值，括弧中的數字為標準誤。

3. ***表 $p < .001$ ，**表 $p < .01$ ，*表 $p < .05$ ，^s表 $p < .1$ 。

相較於政黨認同、候選人評價這兩個變數，統獨立場影響力的變動算是比較小的，傳統probit模型高估傾向統一比傾向獨立的選民更傾向投票給連宋的強度，校正選樣偏誤後的參數估計值由-0.483減少為-0.352，而在校正選樣偏誤之後，傾向維持現狀的選民比傾向獨立的選民更傾向投票給連宋的強度也是減弱的。

參數估計值變動最少的是省籍這個變數，在校正選樣偏誤之後，省籍的參數估計值從1.000減弱到0.939，僅減少了0.061。

綜合來說，在2004年的總統選舉中，變數係數變動程度較大的，仍然是政黨認同為中立無反應者與無法比較候選人評價高低者。

根據以上兩個模型的參數估計所計算出的預測結果，詳見表 4 - 1 0。從表 4 - 1 0 中可以得知，傳統 probit 模型的預測結果，由於沒有考慮選樣偏誤的問題，果然低估了陳水扁、呂秀蓮這組候選人的得票率，選樣偏誤模型的表現相當不錯，雖然沒有預測到正確的當選者，不過與實際得票率的誤差只有 0.65%，除了比傳統 probit 模型表現更好之外，也比表態者的投票意向百分比更貼近實際選舉結果。

表 4 - 1 0 2004 年總統選舉選樣偏誤模型、傳統 probit 模型、電訪表態者比較

	實際得票率	電訪表態者	傳統 probit 模型	選樣偏誤模型
扁呂	50.11	46.81	43.92	49.46
連宋	49.89	53.19	56.08	50.54
與實際誤差	——	3.30	6.19	0.65

第三節 小結

本章運用選樣偏誤模型對兩次總統選舉進行預測，以檢驗選樣偏誤模型在總統選舉的適用性。

研究結果發現，在兩次總統選舉當中，若只單純對結果方程式進行估計，都會因為選樣偏誤而造成估計的偏差情況，當我們校正選樣偏誤的問題之後，所得到的預測結果都相當不錯，選樣偏誤模型的表現比直接看表態者的投票意向百分比、傳統 probit 模型都要來的好，其中 2000 年的總統選舉中，選樣偏誤模型的誤差僅有 0.59%；而 2004 年總統選舉的預測結果，雖然沒有正確預測到陳呂配連任，不過與實際得票率的誤差也僅有 0.65%，筆者認為，已經具有相當準確的預測力了。

