

## 第三章 選樣偏誤模型在地方首長選舉的應用

在本章中，筆者選擇三次地方性選舉建立選樣偏誤模型，分別是 2001 年台北縣長選舉、2002 年台北市長選舉，以及 2002 年高雄市長選舉，檢驗選樣偏誤模型在不同地方首長選舉的表現。

### 第一節 2001 年台北縣長選舉

在 2001 台北縣長選舉中，有四位候選人參選，分別是民進黨籍尋求連任的蘇貞昌、國民黨、親民黨、新黨共同推薦的王建煊，以及兩位無黨籍的候選人，劭建興與石翊靖。雖然參選人有四位，不過劭建興和石翊靖的知名度並不高，在整個選舉過程中，幾乎都是呈現蘇貞昌與王建煊兩強對決的情勢，最後的選舉結果是，蘇貞昌獲得 51.31% 的選票當選連任，王建煊得到 48.16% 的選票，邵建興與石翊靖的得票率分別為 0.23% 及 0.29%，兩位主要候選人的得票率合計超過九成九。

表 3 - 1 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 78.89% 的受訪者願意表達可能的投票對象，而有 21.11% 的受訪者並不願意表態。在表態的受訪者當中有 46.99% 支持王建煊，53.01% 是支持蘇貞昌，沒有受訪者回答要投票給劭建興或石翊靖，因此本研究主要針對王建煊與蘇貞昌的相對得票率進行預測。接下來則進入選樣方程式的建構。

表 3 - 1 2001年台北縣長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
蘇貞昌	713	41.81%	53.01%
王建煊	632	37.08%	46.99%
不表態	360	21.11%	
合計	1705	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 3 - 2 中，我們可以發現，在性別方面，女性不表態的比例為 25.8%，而男性不表態的比例則為 16.1%，可見女性比男性更不願意表態。年齡層對於是否表態的影響，呈現了年齡愈大，愈不願意表態的情形，表態比例最高的是 20 至 29 歲這一個年齡層，有將近八成五（84.9%）的受訪者願意表達自己的可能投票對象，表態比例最低的則是 60 歲以上的受訪者，只有七成左右（70.1%）的受訪者願意表態。教育程度與是否表態之間的關係也是如預期一般，教育程度愈低的受訪者表態的比例愈低，低教育程度的受訪者表態比例比高教育程度者要低了將近 23 個百分點左右。在媒體使用情形方面，報紙和電視都看的受訪者有 84.2% 的比例願意表態，都不看的受訪者則僅有六成左右（61.5%）的比例願意表態，差距超過二成。是否具有政黨認同在是否表態上，也具有顯著的差異，沒有政黨認同的受訪者中，有將近一半的比例不願意表態，但是具有政黨認同的受訪者當中，則僅有一成左右（11.4%）的比例不願

意表態。在候選人評價問題回答情形與是否表態之間的關係上，我們可以發現，回答題目愈多的受訪者愈傾向表態，二題都回答的受訪者有超過八成（82.7%）的比例願意表態，反觀二題都沒有回答的受訪者，卻只有三成（30.9%）左右願意表態，差異可說是非常之大。

表 3 - 2 2001 年台北縣長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	25.8	74.2	(875)	$\chi^2=23.983$ d.f=1
	男性	16.1	83.9	(830)	$p<.001$
年齡	20 至 29 歲	15.1	84.9	(292)	$\chi^2=20.773$ d.f=4 $p<.001$
	30 至 39 歲	18.3	81.7	(514)	
	40 至 49 歲	20.0	80.0	(526)	
	50 至 59 歲	27.5	72.5	(207)	
	60 歲以上	29.9	70.1	(137)	
教育程度	低	35.6	64.4	(444)	$\chi^2=84.629$ d.f=2
	中	18.2	81.8	(609)	$p<.001$
	高	13.0	87.0	(637)	
媒體使用情形	報紙電視都不看	38.5	61.5	(91)	$\chi^2=42.383$ d.f=2
	看其中一種	26.1	73.9	(636)	$p<.001$
	報紙電視都看	15.8	84.2	(961)	
是否具有政黨認同	無	49.4	50.6	(437)	$\chi^2=282.813$ d.f=1
	有	11.4	88.6	(1268)	$p<.001$
候選人評價問題回答情形	二題都沒回答	69.1	30.9	(94)	$\chi^2=149.238$ d.f=2
	回答一題	29.5	70.5	(139)	$p<.001$
	二題都回答	17.3	82.7	(1472)	

註：表中所列為橫列百分比（括弧內為樣本數）。

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之後，接下來便是建立選樣方程式的 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，比較基礎是女性；

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。而為了比較各個自變數在模型中的影響力大小，筆者特別將年齡、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，由於其他三個自變數是虛擬變數，原本就是在 0 與 1 之間變動，如此一來，所有自變數每一個單位的變動都是從最小值到最大值，也就是全距的變動，便能夠比較各個自變數的影響力大小。分析的結果詳見表 3 - 3。

表 3 - 3 2001 年台北縣長選舉選樣方程式 probit 模型

	模型 I	模型 II
性別（女性=0）		
男性	.253 (.081) **	260 (.079) **
年齡	-.039 (.247)	——
教育程度（低教育程度=0）		
中教育程度	.346 (.103) ***	.342 (.097) ***
高教育程度	.477 (.110) ***	.450 (.100) ***
媒體使用情形	.266 (.132) *	.266 (.131) *
是否具有政黨認同（無=0）		
有政黨認同	.971 (.083) ***	.977 (.083) ***
候選人評價回答情形	1.041 (.157) ***	1.039 (.153) ***
常數	-1.348 (.205) ***	-1.371 (.165) ***
分析個數	1608	1630
正確預測率（%）	82.21	81.78
-2 Log Likelihood	1296.0528	1331.7215

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 $p < .05$ ，<sup>§</sup>表 $p < .1$ 。

表 3 - 3 中有兩個模型，這是因為在模型 I 中，年齡的係數值雖然方向與筆者預測的一致，但是很可惜並未達到統計上的顯著水準，為了檢證這個不顯著的變數是否會造成其他變數估計值的偏誤，因此將此變數去除之後，再估計模型 II。估計的結果令筆者頗為滿意，因為即使跨不同的模型，各變數的係數大致仍相當穩定，此外各變數的係數值與理論預期的方向完全相符，說明了自變數與應變數之間的估計關係應該相當可信。

從表 3 - 3，模型 II 中各變數的係數可以發現，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，其次則是是否具有政黨認同；有政黨認同的選民比沒有政黨認同的選民更傾向表態；影響力較小的則是性別及媒體使用情形；而整個模型的正確預測率是 81.78%。接下來，筆者將以模型 II 做為選樣偏誤模型中的選樣方程式，建構選樣偏誤模型。

## 二、選樣偏誤模型

在 2001 年台北縣長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（蘇貞昌=1、王建煊=0），自變數為候選人評價、現任者施政滿意度、政黨認同及省籍，其中候選人評價、政黨認同、省籍都是類別變數，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蘇貞昌評價較高、對王建煊評價較高、無法比較三類，比較基礎為對蘇貞昌評價較高；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；現任者施政滿意度是以數字資料的方式放入模型中，為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為 0-1 之間。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，直接進入模型建構的程序。

模型分析的結果見於表 3 - 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 3 - 4 中，各變數的係數方向都與理論預期方向相同，而且每個變數的影響都達到統計上的顯著水準。 $\rho$  等於 0.543， $\rho$  是代表  $u_1$  和  $u_2$  之間的相關係數，它大於 0 的意思是，在不校正選樣偏誤的情況下，將高估結果方程式中  $Y=1$  發生的機率，也就是說傳統 probit 模型將高估選民投票給蘇貞昌的機率。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。

接下來我們觀察各變數的影響力，由於 probit 模型屬於非線性的模型，因此詮釋估計值的方式與一般的線性迴歸模型不同。在 probit 模型中，自變數對於應變數的影響是看其他變數控制在哪一個水準上而定。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 A。

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不小，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王建煊的強度，未校正選樣偏誤時的參數估計值為 -1.657，校正選樣偏誤之後的參數估計值則減少為 -1.570，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 模型低估了中立無反應選民比認同泛綠選民更傾向投票給王建煊的

強度，係數由-0.665 增加為-0.918，表示在校正選樣偏誤之後，中立無反應者與泛綠的距離增加了 0.253，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 0.992 (-0.665 + 1.657)，在選樣偏誤模型中的參數估計值是 0.652 (-0.918 + 1.570)，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小了，綜合來看，校正選樣偏誤之後，泛藍與中立無反應的距離變小，泛綠與中立無反應的距離變大，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藍者比較接近，會比較傾向投票給王建煊，因此若不校正選樣偏誤的話，將高估蘇貞昌的得票率。

在候選人評價方面，傳統 probit 模型高估了給王建煊較高評價者比給蘇貞昌較高評價者傾向投票給王建煊的強度，校正選樣偏誤後的參數估計值從 -2.582 減弱為-2.446，意味著在未校正選樣偏誤的情況下，模型高估給王建煊較高評價者與給蘇貞昌較高評價者之間的差距，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；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蘇貞昌較高評價者更傾向投票給王建煊的強度，係數由-1.308 減少為-1.284，表示在校正選樣偏誤之後，無法比較候選人評價者與給蘇貞昌較高評價者的距離減少了 0.024，不過這個變動程度非常的小，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王建煊較高評價者的距離，若以給王建煊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 1.274 (-1.308 + 2.582)，在選樣偏誤模型中的參數估計值是 1.162 (-1.284 + 2.446)，表示在校正選樣偏誤之後，給王建煊較高評價者與無法比較候選人評價高低者之間的差距也是減少的，不過綜合來看，給王建

煊較高評價者與無法比較候選人評價高低者的距離減少比較多，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王建煊評價較高者，也就是說會比較傾向投票給王建煊，因此若不校正選樣偏誤的話，將高估蘇貞昌的得票率。

在蘇貞昌施政滿意度的影響上，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 2.116 變成 1.895，減少了 0.221，表示在校正選樣偏誤之後，對蘇貞昌施政滿意度愈高，愈傾向投票給蘇貞昌的強度是減弱的，而且減弱的程度不算小。

參數估計值變動較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 0.555 減少到 0.511，僅減少了 0.044，表示在校正選樣偏誤之後，本省人較外省人更傾向投票給蘇貞昌的強度減弱。

從以上這幾個變數的係數變動程度來看，在 2001 年的台北縣長選舉中，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。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 3-4 2001 年台北縣長選舉投票預測模型

	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(蘇貞昌較高=0)		
王建煊較高	-2.446 (.198) ***	-2.582 (.175) ***
無法比較高低	-1.284 (.148) ***	-1.308 (.149) ***
蘇貞昌施政滿意度	1.895 (.382) ***	2.116 (.374) ***
政黨認同(泛綠=0)		
泛藍	-1.570 (.168) ***	-1.657 (.163) ***
中立無反應	-.918 (.195) ***	-.665 (.184) ***
省籍(外省=0)		
本省	.511 (.166) **	.555 (.173) ***
常數	.411 (.335)	.458 (.355)
$\rho$	.543 (.207)	
選樣方程式		
性別(女性=0)		
男性	.254 (.078) **	
教育程度(低教育程度=0)		
中教育程度	.319 (.097) **	
高教育程度	.453 (.099) ***	
媒體使用情形	.270 (.129) *	
是否具有政黨認同(無=0)		
有政黨認同	.977 (.082) ***	
候選人評價回答情形	1.054 (.151) ***	
常數	-1.378 (.163) ***	
分析個數	1630	1285
-2 Log Likelihood	1848.0494	520.1994
LR test ( $\rho = 0$ )	3.87*	

註：1.應變數編碼方式：1-蘇貞昌，0-王建煊。

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  $p < .05$ ，<sup>§</sup>表  $p < .1$ 。

表 3 – 5 是根據以上兩個模型的參數估計值所計算出的預測結果。在預測得票率的計算方面，本研究採取相信表態受訪者所提供的投票意向，而對未表態受訪者的推估，只要受訪者可能投票給王建煊的機率超過五成，則判定該受訪者會投票給他，反之則判定會投票給蘇貞昌。對於模型的評估方面，則是比較候選人預測得票率與實際得票率的誤差，也就是模型預估的蘇貞昌得票率減去蘇貞昌實際得票率的絕對值。

從表 3 – 5 中的數據可以得知，傳統 probit 模型的預測結果，由於沒有考慮選樣偏誤的問題，果然高估了蘇貞昌的得票率，選樣偏誤模型的表現相當不錯，不但正確預測當選者，且與實際得票率的誤差只有 0.71%，除了比傳統 probit 模型表現更好之外，也比電訪表態者的投票傾向更貼近實際選舉結果。

表 3 – 5 2001 年台北縣長選舉選樣偏誤模型、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
蘇貞昌	51.58	53.01	56.24	52.29
王建煊	48.42	46.99	43.76	47.71
與實際誤差	——	1.43	4.66	0.71

註：表中實際得票率為扣除其他候選人票數後的得票率。

## 第二節 2002 年台北市長選舉

2002 年的台北市長選舉，只有兩位參選人，一位是現任市長馬英九，一位是由立委轉戰百里侯的李應元，雖然是三屆台北市長選舉中，第一次出現國民黨與民進黨兩黨對決的局面，但是在整個選舉的過程中，聲勢可說是一面倒的偏向馬英九，最後的選舉結果也是馬英九以 64.11% 的高得票率獲得連任。在這樣的情況之下，預測誰當選已經不是件困難的事，但是能否準確的預測候選人的得票率，仍然是項挑戰。

表 3 - 6 是選民投票意向的次數分配表，從表中數據可以得知，共有 88.77% 的受訪者願意表達可能的投票對象，僅有 11.23% 的受訪者不願意表態，可能是選情相當明朗的關係，所以選民比較願意表態，表態率可說是非常的高。在表態的受訪者當中有 27.19% 支持李應元，72.81% 支持馬英九。雖然這次選舉中，不表態的比率相當低，但是我們仍然先假設表態選民與不表態選民是有差異的，接下來便進入建構選樣方程式的程序。

表 3 - 6 2002 年台北市長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
李應元	245	24.14%	27.19%
馬英九	656	64.63%	72.81%
不表態	114	11.23%	
合計	1015	100.0%	100.0%

### 一、選樣方程式

在台北市長選舉的選樣方程式中，筆者選擇了性別、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及政治功效意識等七個變數，同樣地，在實際進行選樣方程式的模型建構之前，還是有必要先透過自變數與是否表態的交叉列表，來好好觀察這七個自變數和選民是否表態之間的關係。

從表 3 - 7 中的數據來看，首先觀察性別對於是否表態的影響，與研究預期的方向是一致的，男性表態的比例要比女性表態的比例略微高出一些，不過並沒有達到統計上的顯著水準，由此可見，在這次選舉中，性別對於是否表態應該是沒有影響的。年齡對於是否表態的影響，基本上還是呈現年齡層愈低，表態的比例愈高的趨勢，雖然經過卡方檢定的結果顯示，年齡層的差異具有統計上的顯著性，不過從細格內的百分比數字來看，差異並不算大。

接著要看的是教育程度對表態與否的影響，如研究所預期的，教育程度愈高的選民愈願意表態，高教育程度選民的表態比例較低教育程度選民表態比例高出十二個百分點左右。在媒體使用情形方面，亦如研究預期，報紙電視都看的選民，表態比例超過九成(91.1%)，都不看的選民則僅有不到八成(78.9%)的表態比例。

在政治態度方面，是否具有政黨認同對於表態與否的影響，從細格內百分比的數字來看，具有相當大的差異，具有政黨認同的選民中，有高達九成七(97.4%)的比例願意表態，不具有政黨認同的選民中，則僅僅只有六成(61.8%)左右的表態比例，兩者相差了三成五(35.6%)左右。對於候選人評價問題的回答情形，回答候選人評價問題的選民中，超過九成五(95.5%)願意表達投票意向，沒有回答候選人評價問題的選民，願意表達投票意向的比例下滑至74.1%。在政治功效意識的影響方面，呈現政治功效意識愈高，愈願意表態的情況。

綜合以上所述，在這七個變數之中，除了性別沒有通過初步的卡方檢定之外，其他變數與是否表態之間的關係都如理論預期，因此在建立選樣方程式的probit模型時，將先排除性別這個變數，僅放入其他六個變數。模型的應變數為選民是否表態，表態為1，不表態為0。自變數中的教育程度、是否具有政黨認同、候選人評價問題回答情形是類別變數，因此以虛擬變數的方式放入模型中，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；候選人評價問題回答情形區分為有回答與沒有回答，比較基礎為沒有回答；另外三個自變數：年齡、媒體使用情形與政治功效意識，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，年齡、媒體使用情形、政治功效意識這三個變數都將轉化為在0-1之間變動。分析的結果詳見表3-8。

表 3 - 7 2002 年台北市長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	11.5	88.5	(546)	$\chi^2 = .112$ d.f=1
	男性	10.9	89.1	(469)	$p > .1$
年齡	20 至 29 歲	7.7	92.3	(130)	
	30 至 39 歲	8.4	91.6	(250)	$\chi^2 = 12.285$
	40 至 49 歲	8.8	91.2	(296)	d.f=4
	50 至 59 歲	13.8	86.2	(181)	$p < .05$
	60 歲以上	17.4	82.6	(144)	
教育程度	低	20.7	79.3	(164)	$\chi^2 = 19.379$
	中	10.3	89.7	(272)	d.f=2
	高	8.6	91.4	(571)	$p < .001$
媒體使用情形	報紙電視都不看	21.1	78.9	(90)	$\chi^2 = 11.991$
	看其中一種	12.2	87.8	(384)	d.f=2
	報紙電視都看	8.9	91.1	(537)	$p < .01$
是否具有政黨認同	無	38.2	61.8	(246)	$\chi^2 = 237.059$ d.f=1
	有	2.6	97.4	(769)	$p < .001$
候選人評價問題回答情形	沒回答	25.9	74.1	(166)	$\chi^2 = 75.724$ d.f=1
	有回答	4.5	95.5	(651)	$p < .001$
政治功效意識	低	18.4	81.6	(201)	$\chi^2 = 27.296$
	中	14.8	85.2	(311)	d.f=2
	高	6.2	93.8	(503)	$p < .001$

註：表中所列為橫列百分比（括弧內為樣本數）。

表 3 - 8 有兩個模型，在模型 I 中，在控制了其他變數之後，年齡的係數方向與預期並不一致，係數的方向顯示年齡愈大愈傾向表態，不過並未達到統計上的顯著水準，在教育程度方面，中教育程度比低教育程度更傾向不表態，也與預期的方向不同，不過同樣並未達到統計上的顯著性，此外，雖然係數的方向顯示高教育程度比低教育程度更傾向表態，與理論相符，但是也沒有達到統計上的顯著水準，媒體使用情形的表現與預期方向相同，可是也是沒有達到統計上的顯著水準，在六個自變數當中，僅剩下是否具有政黨

認同、候選人評價問題回答情形及政治功效意識等三個變數，係數方向與理論預期一致，且達到統計上的顯著性。爲了檢證這些方向不正確，或是不顯著的變數是否會造成其他變數估計值的偏誤，因此將這些變數刪除之後，再估計模型 II。估計的結果尚可接受，觀察這三個跨兩個模型的變數，其係數大致穩定，此外各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，說明了這三個自變數與應變數之間的估計關係應該是可以相信的。

從表 3 - 8 模型 II 中各變數的係數可以發現，是否具有政黨認同對於表態與否的影響是最大的，其次是候選人評價問題回答情形，在三個變數當中影響力最小的則是政治功效意識，整個模型的正確預測率是 91.51%。接下來，筆者將以模型 II 做爲選樣偏誤模型中的選樣方程式，來建構選樣偏誤模型。

表 3 - 8 2002 年台北市長選舉選樣方程式 probit 模型

	模型 I	模型 II
年齡	.037 (.487)	——
教育程度 (低教育程度=0)		
中教育程度	-.117 (.273)	——
高教育程度	.050 (.267)	——
媒體使用情形	.007 (.242)	——
是否具有政黨認同 (無=0)		
有政黨認同	1.428 (.164) ***	1.424 (.158) ***
候選人評價回答情形 (無=0)		
有回答	.722 (.168) ***	.755 (.158) ***
政治功效意識	.407 (.211) §	.453 (.102) *
常數	-.189 (.386)	-.271 (.180)
分析個數	773	789
正確預測率 (%)	91.98	91.51
-2 Log Likelihood	303.0582	323.1264

註：1.應變數編碼方式：1-表態，0-不表態。

2.表中數字爲probit模型分析的估計值，括弧中的數字爲標準誤。

3. \*\*\*表 $p < .001$ ，\*\*表 $p < .01$ ，\*表 $p < .05$ ，§表 $p < .1$ 。

## 二、選樣偏誤模型

2002 年台北市長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（李應元=1、馬英九=0），自變數與 2001 年台北縣長相同，分別是候選人評價、現任者施政滿意度、政黨認同及省籍，其中候選人評價、政黨認同、省籍都是類別變數，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李應元評價較高、對馬英九評價較高、無法比較三類，比較基礎為對李應元評價較高；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；現任者施政滿意度是以數字資料的方式放入模型中，為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為 0-1 之間。模型分析的結果見於表 3 - 9。

在表 3 - 9 中，各變數的係數方向都與理論預期方向相同，除了省籍之外，其他變數的影響都達到統計上的顯著水準。 $\rho$  等於-0.700， $\rho$  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 模型會低估選民投票給李應元的機率。

同樣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 B。

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不小，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馬英九的強度，未校正選樣偏誤時的參數估計值為-2.147，校正選樣偏誤之後的參數估計值則減少為-2.075；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馬英九的強度，兩個模型之係數由-1.567 減少為-1.085，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 0.482，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 0.58 (-1.567 + 2.147)，在選樣偏誤模型中的參數估計值則是 0.99 (-1.085 + 2.075)，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則是縮小，表示政黨認同是中立無反應的選民中，不表態者的投票傾向比表態者更接近認同泛綠者，會比較傾向投票給李應元，因此若不校正選樣偏誤的話，將低估李應元的得票率。

在候選人評價方面，傳統 probit 模型高估了給馬英九較高評價者比給李應元較高評價者更傾向投票給馬英九的強度，校正選樣偏誤後的參數估計值從 -1.751 減弱為 -1.658，意味著在未校正選樣偏誤的情況下，模型高估給馬英九較高評價者與給李應元較高評價者之間的差距，不過變動程度並不是很大；對無法比較候選人評價高低者的影響，則是在未校正選樣偏誤的傳統 probit 模型中，高估了無法比較候選人評價高低者比給李應元較高評價者更傾向投票給馬英九的強度，係數由 -0.426 減少為 -0.329，表示在校正選樣偏誤之後，無法比較候選人評價者與給李應元較高評價者的距離減少了 0.097，不過變動程度也不是非常大，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馬英九較高評價者的距離，若以給馬英九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 1.325 (-0.426 + 1.751)，在選樣偏誤模型中的參數估計值則是 1.329 (-0.329 + 1.658)，表示在校正選樣偏誤之後，給馬英九較高評價者與無法比較候選人評價高低者之間的差距是增加的，綜合來看，給馬英九較高評價者與無法比較候選人評價高低者的距離增加，給李應元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，不表態者和表態者比起來，其投票傾向是與給李應元評價較高者比較接近，會比較傾向投票給李應元，因此若不校正選樣偏誤的話，將低估李應元的得票率。



表 3 - 9 2002 年台北市長選舉投票預測模型

	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 (李應元較高=0)		
馬英九較高	-1.658 (.394) ***	-1.751 (.399) ***
無法比較高低	-.329 (.238)	-.426 (.242) §
馬英九施政滿意度	-3.655 (.589) ***	-3.922 (.188) ***
藍綠認同 (泛綠=0)		
泛藍	-2.075 (.279) ***	-2.147 (.280) ***
中立無反應	-1.085 (.403) **	-1.567 (.257) ***
省籍 (外省=0)		
本省	.442 (.295)	.499 (.311)
常數	2.787 (.425) ***	2.915 (.430) ***
$\rho$	-.700 (.320)	
選樣方程式		
是否具有政黨認同 (無=0)		
有政黨認同	1.417 (.157) ***	
候選人評價回答情形 (無=0)		
有回答	.772 (.155) ***	
政治功效意識	.399 (.209) §	
常數	-.245 (.183)	
分析個數	789	717
-2 Log Likelihood	500.5092	178.5579
LR test ( $\rho = 0$ )	1.18	

註：1.應變數編碼方式：1-李應元，0-馬英九。

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 $p < .05$ ，§表 $p < .1$ 。

在馬英九施政滿意度的影響上，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-3.922 變成 -3.655，減少了 0.267，表示在校正選樣偏誤之後，對馬英九施政滿意度愈高，愈傾向投票給馬英九的強度是減弱的，而且減弱的程度不算小。

參數估計值變動較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 0.499 減弱到 0.442，僅減少了 0.057，表示在選樣偏誤模型中，本省人較外省人更傾向投票給李應元的強度減弱。

綜合來說，以上這幾個變數的係數變動程度大小，以政黨認同是中立無反應者的變動程度為最大，在校正選樣偏誤之後，可以讓我們對中立無反應者的投票傾向做出較為正確的判斷。

雖然模型整體看起來還算不錯，不過很可惜的是，對於  $\rho = 0$  的檢定並沒有達到統計上的顯著性，也就是說在這次選舉中，本研究所設定的結果方程式與選樣方程式之間的相關程度可能為 0。之所以會得到這樣的結果，可能的原因有下列兩個：

1. 在選樣方程式的設定上，可能沒有掌握到真正影響表態或不表態的原因，事實上在建構選樣方程式的過程中，我們可以發現，包括性別、年齡、教育程度、媒體使用情形等變數，在是否表態上的差異都不顯著，整個選樣方程式到最後僅剩下三個變數，這樣的模型設定，的確有可能無法掌握到真正影響表態或不表態的因素。
2. 表態者所給予的答案可能並非是真實的，在選舉勝負早已底定的情況下，某些可能傾向投票給李應元的受訪者，因為意見與主流趨勢不符，所以不好意思說，反而說他會投票給馬英九，造成估計模型的偏差。

不過儘管如此，我們還是利用兩個模型的參數估計值來計算候選人的預

測得票率，預測結果詳見表 3-10。

表 3-10 2002 年台北市長選舉選樣偏誤模型、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
李應元	35.89	27.19	27.43	27.43
馬英九	64.11	72.81	72.57	72.57
與實際誤差	——	8.70	8.46	8.46

從表 3-10 的數字來看，選樣偏誤模型所預估的得票率與傳統 probit 模型的預測值是一樣的，與表態者所回答的投票對象比例相當接近，當然這是因為表態的比例非常高的原因，比較模型預測得票率與實際得票率的誤差可以發現，傳統 probit 模型的預測力雖然比直接看電訪表態者的投票傾向來的好一些，但是與實際得票率的誤差相當的大，顯示模型的預測力非常差。筆者認為可能的原因有二，一是前一段所提到的，受訪者給了不真實的答案，所以造成研究者的誤判，第二個可能的原因則是投票率較低的關係，2002 年市長選舉的投票率是三屆民選市長選舉中，投票率最低的，1994 年的投票率是 78.5%，1998 年的投票率是 80.9%，而 2002 年的投票率僅有 70.6%，因此可能發生的情況是，原本傾向支持馬英九的選民，因為選情非常樂觀，因此最後就決定不去投票了，而傾向支持李應元的選民不希望選舉結果相差太懸殊，去投票的意願較強，所以造成模型低估了李應元的得票率，不過這只是筆者的假設，並沒有實際的經驗可以做為佐證。

### 第三節 2002 年高雄市長選舉

高雄市是台灣第二個直轄市，可說是整個南台灣的政治、經濟、文化和交通的重鎮，在 2002 年的市長選舉中，呈現參選爆炸的情況，共有五位候選人出來競選，其中代表民進黨參選的是現任市長謝長廷，而代表國民黨參選

者則是曾經擔任過副市長的黃俊英，另外無黨籍的候選人則是具有相當高知名度的前民進黨主席施明德，曾經擔任過嘉義市長及內政部長的張博雅，以及曾經擔任過立法委員的黃天生。雖然有五位參選人，但是在聲勢上，大致仍呈現民進黨與國民黨兩黨對決的情況。最後的選舉結果，謝長廷獲得 50.04% 的選票當選連任，黃俊英則是獲得 46.82% 的選票，其他三位無黨籍的候選人中，以張博雅的得票率最高，不過也僅有 1.75%，施明德的得票率為 1.13%，黃天生更只有 0.26% 的得票率。

表 3 - 1 1 是該次選舉前的電話訪問中，選民投票意向的次數分配表，其中表態要投票給謝長廷的受訪者有 33.34%，要投票給黃俊英的比例是 43.36，表態要投票給施明德的受訪者有 3.34%，要投給張博雅的受訪者有 4.62%，要投給黃天生的僅有 0.2%，合計約八個百分點左右，由於表態要投票給這三個候選人的受訪者次數過少，容易造成估計上的不穩定，因此本研究將之捨去，僅針對謝長廷與黃俊英的相對得票率進行預測。在表態支持謝長廷或黃俊英的受訪者當中有 43.39% 要投票給謝長廷，56.61% 要投票給黃俊英，從電訪的初步結果來看，是黃俊英呈現領先的情況。接下來便開始進入選樣方程式的建構。

表 3 - 1 1 2002年高雄市長選舉選民投票意向次數分配表

	次數	百分比	有效百分比	候選人支持率 相對百分比
謝長廷	338	33.34	36.19%	43.39%
黃俊英	441	43.36	47.22%	56.61%
不表態	155	15.24	16.59%	
施明德	34	3.34		
張博雅	47	4.62		
黃天生	2	0.20		
合計	1017	100.0	100.0%	100.0%

## 一、選樣方程式

在高雄市長選舉的選樣方程式中，與台北市長選舉相同，本研究選擇了性別、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及政治功效意識等七個變數，同樣地，在實際進行選樣方程式的模型建構之前，還是先透過自變數與是否表態的交叉列表及卡方檢定，來檢視這七個自變數和選民是否表態之間的關係。結果詳見表 3 - 1 2。

表 3 - 1 2 2002 年高雄市長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	19.8	80.2	(489)	$\chi^2=7.789$ d.f=1
	男性	13.0	87.0	(445)	$p<.01$
年齡	20 至 29 歲	10.2	89.8	(118)	$\chi^2=23.922$ d.f=4 $p<.001$
	30 至 39 歲	12.0	88.0	(233)	
	40 至 49 歲	14.2	85.8	(316)	
	50 至 59 歲	26.5	73.5	(162)	
	60 歲以上	24.0	76.0	(96)	
教育程度	低	32.4	67.6	(204)	$\chi^2=48.800$ d.f=2
	中	13.4	86.6	(358)	$p<.001$
	高	10.7	89.3	(366)	
媒體使用情形	報紙電視都不看	37.9	62.1	(87)	$\chi^2=43.526$ d.f=2
	看其中一種	18.6	81.4	(366)	$p<.001$
	報紙電視都看	10.4	89.6	(472)	
是否具有政黨認同	無	46.8	53.2	(220)	$\chi^2=189.919$ d.f=1
	有	7.3	92.7	(714)	$p<.001$
候選人評價問題回答情形	沒回答	64.1	35.9	(131)	$\chi^2=248.660$ d.f=1
	有回答	8.8	91.2	(803)	$p<.001$
政治功效意識	低	23.3	76.7	(305)	$\chi^2=22.301$ d.f=2
	中	17.5	82.5	(315)	$p<.001$
	高	9.2	90.8	(314)	

註：表中所列為橫列百分比（括弧內為樣本數）。

從表 3 - 1 2 可以發現，女性有將近二成（19.8%）的比例不願意表態，而男性不願意表態的比例是一成三，較女性稍低一些。年齡層方面，與理論所預期的方向一致，年齡層愈高愈不願意表態，年齡層最高的 60 歲以上這一群人有 24.0% 不願意表態，反觀最年輕的 20 至 29 歲這一層，則僅有一成左右（10.2%）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，高教育程度的選民有將近九成（89.3%）的比例願意表態，而低教育程度的選民則僅有六成左右（62.1%）的比例願意表態。媒體使用情形的影響上，資訊接收管道愈多的受訪者願意表態的比例愈高，報紙、電視都看的受訪者僅有一成左右的比例不表態，而這二者都不看的受訪者則有超過三成五（37.9%）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非常的明顯，沒有政黨認同的受訪者中，表態的比例只略微超過五成（53.2%）一些，而有政黨認同者，其表態比例超過九成以上（92.7%）。候選人評價問題回答情形對於是否表態也有影響，有回答候選人評價問題的受訪者當中，超過九成以上（91.2%）願意表達他的投票意願，反觀沒有回答候選人評價的受訪者當中，願意表態的比例連四成都不到（35.9%）。最後觀察的是政治功效意識，同樣也是符合理論的預期，政治功效意識最高的受訪者不表態的比例不到一成（9.2%），而政治功效意識最低的受訪者則有超過二成（23.3%）的比例不願意表態。

整體來看，在交叉列表中，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，因此這七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、候選人評價問題回答情形是類別變數，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；候選人評價問題回答情形區分為有回答與沒有回答，比較基礎為沒有回答；另外三個自變數：年齡、媒體使用情形與政治功效意識，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，筆者採取與台北市長選舉相同的處理方式，將年齡、媒體使用情形、政治功效意識這三個變數都轉化為在 0-1 之間變動。選樣方程式的 probit 模型分析結果詳見表 3 - 1 3。

表 3 - 1 3 2002 年高雄市長選舉選樣方程式 probit 模型

	模型 I	模型 II
性別 (女性=0)		
男性	.150 (.132)	——
年齡	-.648 (.381)	——
教育程度 (低教育程度=0)		
中教育程度	.234 (.172)	——
高教育程度	.166 (.178)	——
媒體使用情形	.458 (.191) *	.467 (.182) **
是否具有政黨認同 (無=0)		
有政黨認同	1.074 (.134) ***	1.089 (.131) ***
候選人評價回答情形 (無=0)		
有回答	1.200 (.153) ***	1.271 (.148) ***
政治功效意識	.425 (.159) **	.533 (.155) ***
常數	-1.239 (.344) ***	-1.376 (.184) ***
分析個數	758	771
正確預測率 (%)	87.20	87.16
-2 Log Likelihood	485.3835	501.7464

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 $p < .05$ ，<sup>§</sup>表 $p < .1$ 。

表 3 - 1 3 同樣有兩個模型，在模型 I 當中，所有變數的係數方向與研究預期都是一致的，但是性別、年齡及教育程度對於是否表態的影響上，並沒有達到統計上的顯著水準。同樣地，為了檢證這些不顯著的變數是否會造

成其他變數估計值的偏誤，因此筆者將這三個變數刪除之後，再估計模型 II。估計的結果還可以接受，觀察這四個跨二個模型的變數，其係數除政治功效意識的變動稍微多一點點以外，大致上來說都還算穩定，此外各變數的係數值與理論預期的方向也都完全相符，且仍然達到統計上的顯著水準，說明了這四個自變數與應變數之間的估計關係應該是沒有問題的。

從表 3-13 中，模型 II 各變數係數的大小可以發現，對候選人評價問題的回答情形這個變數，是影響選民是否表態的最重要因素，對候選人評價問題有回答的選民，其表態的機率較沒有回答的人來得高，其次則是是否具有政黨認同；有政黨認同的選民比沒有政黨認同的選民更傾向表態；影響力最小的則是媒體使用情形；整個模型的正確預測率是 87.16%。接下來，筆者將以模型 II 做為選樣偏誤模型中的選樣方程式，來建構選樣偏誤模型。

## 二、選樣偏誤模型

在 2002 年高雄市長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（謝長廷=1、黃俊英=0；投給其他候選人歸為遺漏值），自變數同樣是候選人評價、現任者施政滿意度、政黨認同及省籍。其中候選人評價、政黨認同、省籍都是類別變數，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黃俊英評價較高、對謝長廷評價較高、無法比較三類，比較基礎為對謝長廷評價較高；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；現任者施政滿意度是以數字資料的方式放入模型中，為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為 0-1 之間。模型結果詳見表 3-14。

從表 3-14 中可以發現，各變數的係數方向都與理論預期方向相同，



除了省籍之外，其他變數的影響都達到統計上的顯著水準。 $\rho$  為-0.836，而  $\rho$  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，也就是說我們將低估選民投票給謝長廷的機率。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。

同樣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 C。

觀察模型中參數估計值的變動，在候選人評價方面，傳統 probit 模型高估了給黃俊英較高評價者比給謝長廷較高評價者傾向投票給黃俊英的強度，校正選樣偏誤後的參數估計值從-2.643 減弱為-2.410；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謝長廷較高評價者更傾向投票給黃俊英的強度，係數由-1.342 減少為-1.051，表示在校正選樣偏誤之後，無法比較候選人評價者與給謝長廷較高評價者的距離減少了 0.291，變動程度不算小，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黃俊英較高評價者的距離，若以給黃俊英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 1.301 (-1.342 + 2.643)，在選樣偏誤模型中的參數估計值是 1.359 (-1.051 + 2.410)，表示在校正選樣偏誤之後，給黃俊英較高評價者與無法比較候選人評價高低者之間的差距也是增加的，綜合來看，在校正選樣偏誤之後，給黃俊英較高評價者與無法比較候選人評價高低者的距離增加，而無法比較候選人評價者與給謝長廷較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謝長廷評價較高者，也就是說會比較傾向投票給謝長廷，因此若不校正選樣偏誤的話，將低估謝長廷的得票率。

表 3 - 1 4 2002 年高雄市長選舉投票預測模型

	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 (謝長廷較高=0)		
黃俊英較高	-2.410 (.291) ***	-2.643 (.290) ***
無法比較高低	-1.051 (.267) ***	-1.342 (.282) ***
謝長廷施政滿意度	1.873 (.464) ***	2.132 (.502) ***
藍綠認同 (泛綠=0)		
泛藍	-1.475 (.251) ***	-1.610 (.264) ***
中立無反應	.122 (.261)	-.296 (.278)
省籍 (外省=0)		
本省	.329 (.311)	.352 (.357)
常數	.677 (.408)	.530 (.489)
$\rho$	-.836 (.156)	
選樣方程式		
媒體使用情形	.404 (.180) *	
是否具有政黨認同 (無=0)		
有政黨認同	1.113 (.131) ***	
候選人評價回答情形 (無=0)		
有回答	1.223 (.145) ***	
政治功效意識	.624 (.151) ***	
常數	-.1353 (.184) ***	
分析個數	771	621
-2 Log Likelihood	657.371	159.4415
LR test ( $\rho = 0$ )	3.82 <sup>s</sup>	

註：1.應變數編碼方式：1-謝長廷，0-黃俊英。

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 $p < .05$ ，<sup>s</sup>表 $p < .1$ 。

政黨認同參數估計值的變化也不小，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黃俊英的強度，未校正選樣偏誤時的參數估計值為-1.610，校正選樣偏誤之後的參數估計值則減少為-1.475，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較傾向表態，因此使得傳統 probit 模型高估藍綠認同者之間的差異；中立無反應的選民在傳統 probit 模型中，比認同泛綠的選民更不傾向投票給謝長廷，但是在選樣偏誤模型中，係數的方向改變了，變成比認同泛綠的選民更傾向投票給謝長廷，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  $1.314(-0.296 + 1.610)$ ，在選樣偏誤模型中的參數估計值是  $1.597(.122 + 1.475)$ ，表示在校正選樣偏誤之後，中立無反應者與泛藍的距離增加了 0.283，是個不小的變動，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，會比較傾向投票給謝長廷，因此若不校正選樣偏誤的話，將低估謝長廷的得票率。

在謝長廷施政滿意度的影響上，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 2.132 變成 1.873，減少了 0.259，表示在校正選樣偏誤之後，對謝長廷施政滿意度愈高，愈傾向投票給謝長廷的強度是減弱的，而且減弱的程度不少。

參數估計值變動較少的仍然是省籍這個變數，在校正選樣偏誤之後，省籍的參數估計值從 0.352 減弱到 0.329，僅減少了 0.023，表示在選樣偏誤模型中，本省人較外省人更傾向投票給謝長廷的強度減弱，不過減弱的程度很小。

從以上這幾個變數的係數變動程度來看，在 2002 年的高雄市長選舉中，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。此外，無法比較候選人評價高低這個變數的變動也不小，而它變動的方向與政黨認同中立無反應變數變動的方向可說是相當一致。省籍影響力的變動是最少的，可能是因為省籍本身並不太會去影響是否表態，因此它的參數估計值沒有太大的變化。

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 3 - 1 5。從表 3 - 1 5 中可以得知，電訪表態者所宣稱的投票對象比例與實際得票率之間的差距非常之大，達到 8.27%。傳統 probit 模型的預測結果，由於沒有考慮選樣偏誤的問題，低估了謝長廷的得票率，雖然與實際得票率的差距僅有 1.94%，不過並沒有預測出正確的當選人。表現最好的是選樣偏誤模型，預測結果與實際得票率的誤差只有 1.16%，也正確的預測謝長廷當選連任。

表 3 - 1 5 2002 年高雄市長選舉選樣偏誤模型、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
謝長廷	51.66	43.39	49.72	52.82
黃俊英	48.34	56.61	50.28	47.18
與實際誤差	——	8.27	1.94	1.16

註：表中實際得票率為扣除其他候選人票數後的得票率。

## 第四節 小結

本章運用選樣偏誤模型對三次地方首長選舉進行預測，以檢驗選樣偏誤模型在地方首長選舉的適用性。

研究結果發現，在 2001 年台北縣長選舉、2002 年高雄市長選舉中，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，因此，

使用選樣偏誤模型所計算出的預測結果都比傳統probit模型來得準確。

而在2002年台北市長選舉中，選樣偏誤模型預測不佳，最可能的原因是發生在選樣方程式的設定上，由於選樣方程式中僅有三個變數，可能並沒有掌握到真正影響表態或不表態的變數，再者，在這種勝負早已底定的選舉中，表態受訪者所宣稱的投票對象不一定是真實的，原本傾向投票給落後一方的受訪者，有可能爲了順應主流意見，反而告訴研究者他要投票給領先的一方，而造成研究上的錯誤判斷。

整體來說，本研究所設定的選樣偏誤模型在勝負較爲接近的地方首長選舉是可以適用的，至於在勝負差距懸殊的地方首長選舉中，選樣偏誤模型並沒有發生效用，究竟是什麼因素去影響選民在勝負懸殊的選舉中表態與否，還需要更深的思考。

