# 判定預測市場之準確度：單一與合併鑑別模型之比較

戴中擎．池秉聰．林鴻文．童振源*

預測市場是近年來新發展出的預測方法，許多實證研究均證明預測市場能有效整合資訊並提出準確的預測。然而在大多數預測市場研究中，研究者只能由過去的歷史準確率來衡量市場預測的可靠性，無法針對單一市場合約預測的正確與否進行事前的評估。本文提出一個植基於市場交易特徵的合併鑑別方法，藉由整合迴歸模型、多變量分析、決策樹、及支持向量機等四種模型來擷取與市場預測準確率有關的潛在資訊。本文使用未來事件交易所自 2006 年至 2011 年共 650 個選舉合約作為資料，經實證分析後驗證合併鑑別模型可以非常準確地於事前對任一合約預測的正確與否提出評斷。本文所提出的合併鑑別方法不但比單一鑑別模型更為可靠，而且可依決策者不同的目標函數提出不同的評斷以進行風險控管。

**關鍵詞**：預測市場、事前鑑別、合併預測、邏輯迴歸、主成份分析、判別分析、決策樹、支持向量機

**JEL 分類代號**：C38, C53, G13, G14

## 1 緒論

最近二十年來，預測市場 (Prediction Markets) 的研究方法已經逐漸成為一種實用的預測工具，並廣泛地被用以預測各國的選舉活動、體育賽事、政經事件、乃至於企業活動，並獲得了相當的成功 (Hahn and Tetlock, 2006)。預測市場又稱作「資訊市場」(Information Markets)，是一個讓參與者買賣「未來事件合約」的市場。其運作類似一般的「期貨市場」，透過市場機制來彙整各方面的資訊以預測未來事件的結果。在預測市場中，每一個交易標的都是一個

* 作者分別為東海大學經濟學系助理教授、淡江大學產業經濟學系副教授、中山大學（大陸）南方學院經濟學與商務管理系助理教授與國立政治大學國家發展研究所特聘教授。戴中擎為通訊作者。本文審稿過程中三位匿名評審委員所提出之諸多建設性意見與指正均讓本研究能更加完整，作者在此特表謝忱。本文若有任何謬誤，當屬作者之責。

未來的事件，而市場參與者會依據對該事件發生可能性的判斷來衡量該合約的價格，並據以進行買賣。合約的價格可以作為該事件是否發生或如何發生的預測參考 (Forsythe et al., 1992)。

以市場作為預測工具的理論源自於所謂的「海耶克假說」(Hayek hypothesis) (Hayek, 1945; Smith, 1982)。該假說認為如果市場是有效率的，則市場價格將充分反映所有可以獲得的資訊。利用市場來預測選舉事件的正式研究始於自 1988 年開始運作的美國愛荷華大學「愛荷華電子市場」(Iowa Electronic Markets, IEM)。[1] 由於該市場預測歷屆美國總統大選結果之準確度超過傳統民意調查甚多 (Berg et al. 2001)，因而開啟了以市場進行預測的研究領域 (Hahn and Tetlock, 2006)。在世界各國的預測市場中，幾乎所有的研究都指出預測市場比傳統民調更為準確，例如 Allen et al. (2004)、Walker (2006)、Berg, Nelson, and Rietz (2008)、Erikson and Wlezien (2008) 針對美國選舉的預測市場資料進行的分析，還有 Forsythe et al. (1995) 針對加拿大選舉、Jacobsen et al. (2000) 針對荷蘭選舉、Wolfers and Leigh (2002) 以及 Leigh and Wolfers (2006) 針對澳洲選舉、Brüggelambert (2004) 針對德國選舉、童振源等（2009）針對臺灣選舉的預測市場資料進行的分析。

自愛荷華電子市場以降，預測市場不但自成了一個研究議題，也逐漸被政府、學界、及企業接受作為一個重要的決策支援系統 (Berg and Rietz, 2003; Hahn and Tetlock, 2006; Bergfjord, 2008)。然而，預測市場的高準確率是根據過去的市場案例來計算的歷史準確率，對於當前未知的預測事件，若全憑市場的價格作為決策參考仍具相當風險。舉例而言，吾人可利用預測市場對臺灣將於 2016 年舉行的總統大選進行預測，而我們也知道在選舉預測上預測市場的歷史準確率接近 95%。[2] 在此情況下若某候選人的合約價格顯示將會當選，則這個預測到底可不可信呢？若按歷史準確度而言，不論此候選人合約的交易人數多寡、交易量大小、價格動態為何，都只能作出同樣的結論：若相信市場的預測，會有 95% 的機率可以得到正確的結果。但若該合約從頭到尾只有一筆成交記錄，我們是否還是認為該合約的預測可信呢？尤有甚者，對於「贏者全拿型」(winner-take-all) 的合約而言，決策者往往需要在事前針對不同結果採取迥然不同的行動。在這種情況下歷史準確度提供的只是一個相對頻率 (relative frequency) 的概念，且此頻率並不會隨著該合約市場交易情況的變化而有所不同，因此在實際決策上的參考價值便有所限制。

由以上例子可知，歷史準確度只能反映過去預測市場的整體表現，只適合用來評估「預測市場」這個方法的可靠性。再者，個體發生與否的機率並不一定等同於整體的比例 (Gigerenzer, 2002)。雖然在欠缺個體資訊的情況下，一般作法常常是用整體比例權充個體發生機率的估計，但在預測市場這個例子上，以整體正確比例作為個別合約預測正確機率的作法是可議的：因為每個預測市場所預測的事件迥異，且其交易過程也全然不同，因此不可將每個合約的預測視為隨機實驗的結果，而只以過去正確的相對頻率作為個體合約是否會正確的機率評估。相反地，若研究者能正確地持有並分析個體的詳細資訊，理論上應可以作出更

[1] 該市場原名為「愛荷華政治股票市場」(Iowa Political Stock Market)。
[2] 本文所使用的資料來自於未來事件交易所 (http://xfuture.org/) 自 2008 年以來之交易資料。根據歷史資料，未來事件交易所的預測準確度達 94.62%。

準確的判定。[3]

既然以過去的整體準確率作為個別合約預測正確與否的判斷是不適合的，就必須對每一個合約進行獨立判斷，那麼哪些資訊可以作為個體判斷的依據呢？預測市場只是一個資訊彙整的機制，而資訊的來源為市場參與者的價值判斷以及交易的動態過程。每次選舉事件中預測市場的參與者可能不盡相同，且市場的交易過程和價格動態也是獨一無二的。因此，預測市場是否準確其實取決於一系列的市場因子。[4] 所幸在預測市場的文獻上已知有多項影響市場準確度的因子，這些知識足以幫助研究者在事件發生前對預測市場個別合約的準確度進行一定的評估。在過去的研究中，Chen et al. (2011) 以及林鴻文、童振源、與葉家興（2014）便是在這樣的基礎上捨棄歷史準確度，利用預測市場文獻中所提及的六大類市場交易特徵因子來鑑別合約預測的正確性。

而本文有別於現今預測市場研究之處，在於提出一個基於市場交易特徵的**合併鑑別法**來評估市場的預測準確度。本文所提出之合併鑑別法相對於傳統以歷史準確度作為決策參考的作法有以下幾點優勢：

1. 可針對每一個預測市場合約提出預測正確與否的判斷。
2. 可動態評估任何一個合約價格的預測準確度。

不僅如此，本文相較於 Chen et al. (2011) 以及林鴻文、童振源、與葉家興（2014）所提出的單一鑑別模型更有著以下兩個優勢：

1. 合併鑑別模型之判斷正確率較單一鑑別模型高。
2. 合併鑑別法允許研究者依不同的研究目標挑選出不同的鑑別模型。

本文具體安排如下：在第二節中，本文將對影響預測市場準確度的因子以及與本文相關的鑑別模型文獻進行介紹；在第三節中，本文將提出一系列具體方法來建構合併鑑別模型並介紹本文所使用的資料；在第四節中，我們將檢視單一鑑別模型事前鑑別的結果，而第五節則呈報合併鑑別模型之績效，並與單一模型進行比較；第六節則為結論。

---

[3] 在醫學上一個實際的成功例子是新生兒卡介苗的接種。在過去，人們知道新生兒若接種卡介苗會有十萬分之一的「發生率」會感染結核菌引發疾病。但事實上此十萬分之一的「發生率」不是「機率」，只是具備特殊體質新生兒的「比例」而已。現在人們已透過研究而知悉此特殊體質—「嚴重複合型免疫缺乏症」。不但如此，現今每一位新生兒都可以透過篩檢確認是否可以接種卡介苗。

[4] 舉例而言，若某事件的預測市場合約從頭到尾只有兩人參與交易，而另一預測市場合約則有上千人進場交易，則兩合約價格的可信度自然有可能不同。原因在於前者市場中單一交易者較容易操控價格，除非少數交易者的判斷完全正確，市場預測才會與事實相符；相反地，在後者市場中價格則由為數眾多的交易者決定，除非市場中完全沒有判斷正確的交易者，市場才會失準。

# 2 文獻回顧

在本節中，我們將分別介紹影響預測市場準確度之因子、以及如何利用這些因子來進行事前的鑑別，接著介紹本文所使用的四項鑑別模型，最後則簡介合併預測模型的文獻。

## 2.1 影響預測市場準確度之因子

根據現有之文獻，預測市場準確度共有幾類影響變數。本文所採用的變數包含邊際交易者比例、市場共識程度、市場交易特徵、預測困難度、分身變數及價格變數。這幾類變數分別有實證資料或理論的支持，茲詳細說明如下。

### 2.1.1 邊際交易者比例

在預測市場的文獻中，邊際交易者 (marginal trader) 被認為是預測市場準確的最重要原因。Forsythe et al. (1992) 以及 Forsythe, Rietz, and Ross (1999) 分別利用 IEM 於 1988 至 1998 年美加選舉的市場預測資料以及實驗室證據進行分析，發現「邊際交易者」是預測準確的主要原因。Forsythe et al. (1992) 將邊際交易者定義為在當天成交買賣委託單的交易者，或是在當天最後一筆交易價格上下 2% 以內遞交委託單進行交易的交易者；Forsythe, Rietz, and Ross (1999) 則將邊際交易者定義為遞交價格趨近於市場價格的限價委託單的交易者。他們雖然發現有政黨偏好的市場參與者會因偏好而有非理性交易行為，但市場價格卻不會受到這些交易行為的影響。換言之，邊際交易者主導了市場價格，他們是預測市場準確的主因。

Oliven and Rietz (2004) 將邊際交易者定義為遞交可以在市場成交的限價委託單的交易者，並用 1992 年美國總統大選資料來探討市場價格是否會發生偏誤的問題。他們的研究結果顯示，市場價格主要是由邊際交易者所決定，而這些邊際交易者在預測市場進行交易的過程中，投資行為產生偏差的機率遠低於市場價格接受者。

為了充份衡量邊際交易者在市場中所占的比重，本文使用兩個指標作為邊際交易者活躍程度的代理變數：

1. 限價成交口數占全部成交口數的比例：邊際交易者的特徵是會遞交接近於當前市場價格的限價委託單，因此本文以限價成交單數量占全部成交單數量的比例來表示邊際交易者占全部交易者的比例。

2. 最佳交易者占該合約全部交易者的比重：根據 Forsythe et al. (1992)、Forsythe, Rietz and Ross (1999) 及 Oliven and Rietz (2004)，邊際交易者的報酬較一般交易者高。本研究將所有參與未來事件交易所的使用者進行歷史獲利排名，並定義賺取報酬較

高者為「最佳交易者」。因此，本文使用「最佳交易者占該合約全部交易者的比重」來衡量該合約邊際交易者的比例。

### 2.1.2 市場共識程度

Gruca, Berg, and Cipriano (2005) 由 IEM 預測電影票房的結果，發現交易者之間會藉由訊息的傳遞使得市場漸漸達成共識，因此價格也趨於穩定。他們利用交易者於進入市場交易前所填答的個人預測值之標準差與市場價格分配之標準差進行比較，發現交易者個人預測差異愈大，市場價格波動也就愈大。[5] 在無法直接得知交易者個人預測的情況下，本文使用市場買賣單的價與量來推論交易者的共識程度。本研究以財務市場實證分析常用之變數，建構預測市場共識程度變數如下：

1. 最後一天未成交委託單的加權買賣叫價價差：價差愈小代表該合約在最後一天市場共識愈高，則正確預測的機率應該愈高。

2. 委買單口數相對於委賣單口數的比例（兩者相除）：一般而言，若交易者的預測值較低，在市場中應會站在賣方，反之則為買方。故交易者的預測若有相當的差距，買賣單數量便應旗鼓相當；反之，若交易者之個人預測集中於某個小範圍內，則其下單行為應該也會趨於一致。根據此概念本研究使用「委買單口數對委賣單口數的比例」來推論市場的共識程度，並認為有兩種情況可推論市場具高度共識：(1) 比例接近零：交易市場的委賣口數非常多、委買口數非常少，表示「大部分的交易者」認為目前的市場價格偏高，「事件不會發生的共識」相當一致；(2) 比例接近無窮大，交易市場的委買口數非常多、委賣口數非常少，表示「大部分的交易者」認為目前的市場價格過低，「事件會發生的共識」相當明顯。

### 2.1.3 市場價量特徵

Kambil and van Heck (2002) 指出預測市場能夠成功預測的關鍵因素有三：市場有足夠的交易者、交易者對預測的事件有充份的資訊、有足夠的誘因使參與者進行交易與產生有用的資訊。據此，交易人數、交易量及合約交易天數均可能與市場準確度有關：

1. 交易人數：Gruca, Berg, and Cipriano (2005) 研究顯示發現愈多交易者進行交易，預測市場愈能準確預測電影票房結果。Ledyard (2006) 及 Ho and Chen (2007) 認為眾多的交易者或是交易者熱絡程度都是影響預測市場準確度的因素。

[5] 若利用 Gruca, Berg, and Cipriano (2005) 中表 4 所呈報的數據計算兩者之相關性，可以得到高達 0.715 的相關係數。

2. 該合約開始進行交易到合約清算前一天的交易天數：Snowberg, Wolfers, and Zitzewitz (2005) 發現愈接近選舉投票日預測市場的準確度愈高，因為愈多資訊將會被揭露而反映在價格上。Ho and Chen (2007) 發現經由價格逐步調整的過程將會提高正確訊息所占有的權重，並且亦可以逐漸排除一些偏頗的訊息。交易時間愈長，市場可能彙整更多資訊，導致正確預測的機率提高。

3. 成交口數：Berg, Forsythe, and Rietz (1997) 分析 IEM 從 1988-1994 年十六場得票率預測市場的相對準確度後，發現預測準確度與交易量成正向關係。Luckner, Weinhardt, and Studer (2006) 針對 2005 年德國足球聯賽的資料，以及 Ho and Chen (2007) 在科技產品或季節性流感的預測中均發現，交易量愈高，其市場較能有效地整合資訊，使得其預測較為準確。參與該合約的成交合約口數愈大，表示預測市場能收集到更多資訊，因此正確預測的機率應該愈高。

### 2.1.4 預測困難度

Forsythe, Rietz, and Ross (1999)、Wolfers and Zitzewitz (2004)、Leigh, Wolfers, and Zitzewitz (2003) 以及 Rhode and Strumpf (2008) 對選舉預測市場的分析指出，預測市場在預測發生機率低的事件上可能表現不佳。這是因為市場交易者未必是理性而客觀的，對於預測較困難的合約，由於心理學上的認知偏誤影響其交易行為，導致資訊的彙整困難。基於此種看法，本文衡量合約預測困難度的方式如下：

1. 連動合約組的合約數量：連動合約數量愈高，表示在該選舉事件中有愈多候選人在該選區競選，則理論上預測困難度愈高。

2. 該選區第一名候選人與第二名候選人的價格差距：第一名候選人相對於第二名候選人的價差愈高，推測該選區競爭程度愈小，則預測正確的機率應該愈高。

### 2.1.5 價格操縱程度

諸多實證研究都顯示出預測市場的確具有優秀的預測能力，但預測市場相較於其他預測方法還是有一個顯而易見的可能缺失，那就是價格操縱 (manipulation) 的問題。市場本身的機制理論上會讓企圖操縱價格的交易者蒙受損失，從而迫使操縱者停止操弄甚至因破產而退出市場 (Wolfers and Zitsewitz, 2004)。但在某些預測市場案例中，即使操縱行為會損及場內的利潤，相關的利益人士可能會有非常強烈的場外動機來操弄價格。這些利益人士可能會不惜成本以高額資金投入市場，或耗費大量時間使用多重帳號來影響市場。例如由 Deck, Lin, and Porter (2013) 的研究得知，在某些特定的條件下，市場還是有可能會受到有心人士的操縱。

由於未來事件交易所使用虛擬貨幣，所以可能會有某些會員利用操縱價格獲取市場外的利益，實際上我們也觀察到諸多分身帳號的存在。因此，為了衡量某個合約在交易過程中遭受價格操縱的嚴重程度，在本研究中我們計算分身帳號使用者在該合約交易過程中所占的比重。[6]

### 2.1.6 市場價格

在每位交易者願意承擔的最大損失相同的假設之下，Manski (2006) 指出「認為事件發生的機率大於該價格的交易者占所有交易者的比例恰好等於該價格」。如果將所有交易者對事件發生機率的預測取平均值，則這平均值會落在一個以均衡價格為中點的區間。此外，Wolfers and Zitzewitz (2006) 指出在交易者效用為對數函數的假設之下，均衡價格等於所有交易者對事件發生機率預測值的加權平均。如果所有交易者的財富相同，均衡價格就等於所有交易者的平均預測。在大多數的狀況下，即使放寬對效用函數的假設，Wolfers and Zitzewitz (2006) 發現結論也大致相同，所以預測市場的觀察者可以使用價格預測事件發生的機率。因此，本研究亦將合約的市場價格的高低作為影響準確性的因素之一。

## 2.2 事前鑑別分析

若市場預測的準確性受到一系列市場因子的影響，要如何才能將合約的市場特徵轉化為對該合約是否會準確的事前判斷呢？首先，一個未來即將發生的事件可能有數種結果，而預測市場通常會設立數個合約來預測這些結果的發生機率。舉例而言，一場選舉中有三位候選人，則預測市場將會發行三位候選人的合約供交易者買賣。由於三個合約各自有其市場價格，而根據 Manski (2006) 之理論模型，某位候選人合約的市場價格愈高，表示交易者認為該候選人當選的機率愈高。三者之中，市場價格最高者即為市場之預測結果，表示該人最有可能當選，而其他人將會落選。然而這個預測卻有可能是錯誤的，亦即市場預測將當選者最後落選，而市場預測不會當選者反而最後勝選。因此，每一個合約的價格預測有「正確」及「不正確」兩種可能，而正確與否其實又受到前述一系列因子的影響。

在文獻上，Chen et al. (2011) 使用統計模型與機器學習法 (Machine Learning) 對合約的預測是否會正確進行鑑別分類。其所使用的分類方法包含邏輯迴歸 (Logistic Regression)、決策樹 (Decision Trees, DT)、以及支持向量機 (Support Vector Machine, SVM) 三種，而這三

[6] 在本研究中，相同密碼的交易者人數達三人以上即被歸類為分身帳號。依常理判斷，不同帳號卻有一模一樣密碼的概率非常低，根據經驗使用多重帳號者往往會利用多個 email 申請帳號，但會使用相同的密碼以方便快速登入及切換帳號，故以此方法辨別分身交易者。

種鑑別方法都是常用於資料分類的統計或機器學習方法。Chen et al. (2011) 所使用的資料為未來事件交易所 (http://xfuture.org/) 由 2006 至 2010 年的選舉預測市場交易資料，以其間 575 個合約的市場因素為解釋變數，以合約預測正確與否為被解釋變數。在比較過三種鑑別方法後，Chen et al. (2011) 發現三種鑑別方法所辨識出的重要解釋變數不盡相同。換言之，不同的鑑別方法會利用不同的市場特徵變數對合約預測是否正確進行判斷。

林鴻文、童振源、與葉家興（2014）進一步使用四種統計模型來進行鑑別分類：邏輯迴歸 (Logit)、判別分析 (Discriminant Analysis, DA)、主成分分析 (Principal Component Analysis, PCA) 與邏輯迴歸整合模型 (PCA-Logit)、以及主成分分析與判別分析整合模型 (PCA-DA)。該研究發現在此四種鑑別方法中，以 Logit 模型分類的準確度最高，但該模型卻不能有效地辨別出「預測未正確」的市場合約。換言之，Logit 模型雖然鑑別正確率高，但仍不完美。

有鑑別於各項分類鑑別工具採用的交易特徵不盡相同，而鑑別的成效也不一，因此在總結 Chen et al. (2011) 以及林鴻文、童振源、與葉家興（2014）的經驗後，除了經濟分析中經常使用的 Logit 模型外，本文亦額外採用了以下三種模型來進行鑑別，並以合併預測的方式來截長補短：支持向量機 (SVM)、主成分分析整合判別分析模型 (PCA-DA)、及決策樹 (DT)。

支持向量機在分類問題上已被公認為一個很好的工具，且能克服非線性的分類規則，特別是當樣本量小且資料屬性具有高維度的特性時。在文獻上，SVM 已被用來預測破產 (Segovia-Vargas, Salcedo-Sanz, and Bousoño-Calzón, 2004; Mahdi and Neda, 2013)、信用評等 (Doumpos and Zopounidis, 2009)、價格預測 (Gao et al., 2008)、成本預測 (Deng and Yeh, 2011)、通膨預測 (Marček and Marček, 2006)、財務時間序列預測 (Tay and Cao, 2001; Pai and Lin, 2005) 等，其餘文獻可見 State, Cocianu, and Fusaru (2010) 之整理。

主成份分析在經濟問題的應用上，Blaskowitz and Herwartz (2011) 首先用以判別經濟變數的正向或反向趨勢。但誠如 Anzanello and Fogliatto (2011) 以及 Kessler and Scherer (2011) 對主成份分析的批評，主成份分析在預測分組上不如判別分析法的明確。而在多變量理論中，DA 可以得到一組判別函數作為分組的準則，並將預測對象的樣本作鑑別分組。在經濟與社會議題的應用上，Robertson and Kennedy (1968) 利用判別分析準確地將消費者區分為創新者及非創新者；Crask and Perreault (1977) 討論小樣本的情況下，判別分析的有效性；Blin and Whinston (1975) 應用 DA 在多數決的投票研究；Mitchneck (1995) 運用 DA 在地理經濟的議題；Mitra, Thomakos, and Ulubasoglu (2002) 使用 DA 討論小型開放國家貿易保護措施的影響。

主成份分析結合判別分析的應用相當廣泛，最早出現在行銷學領域。例如：Wilton and Pessemier (1981) 鑑別出訊息對消費者接受新產品程度的影響力；Rosen and Grandbois (1983) 使用 DA 來驗證及說明 PCA 萃取出來的因素是否真的可以決定家庭所得配置；Newman and Sheth (1985) 以 PCA 萃取使用「部分特徵根大於一」的主成份，再把加權平均 PCA 分數引入 DA，預測美國總統候選人的可能得票數。

決策樹 (DT) 為一種以階層樹狀表現的迴歸或分類方法，當預測目標為類別時，其將預測或分類表示成一連串特徵屬性的邏輯組合。決策樹產生的結果相當容易解讀，而新的樣本也可以很快根據決策樹的路徑得到分類的結果。決策樹的方法源於 Morgan and Sonquist (1963) 提出的自動交互檢視法 (Automatic Interaction Detector, AID)，後續則有 Hartigan (1975) 提出之卡方自動交互檢視法 (Chi-square Automatic Interaction Detector, CHAID)。早期決策樹僅適用於類別變數的分類。目前的決策樹演算法已經可適用連續變數且大量使用在商業授信中，例如幫助銀行預測潛在的不良資產 (Tam and Kiang, 1992)、對消費需求的預測 (Tso and Yau, 2007)、與金融授信預測 (Takieddine and Andoh-Baidoo, 2014)。

### 2.3 合併預測方法

多年來已有相當多的證實研究證明合併預測 (combination of forecasts) 可以顯著地降低預測誤差，而簡單的合併方法往往又表現的比複雜的合併方法要穩定且佳。Bates and Granger (1969) 是最早開始討論合併預測理論的文獻，同時也發展並測試相當多的合併預測技術。在他們之前 Crane and Crotty (1967) 建議可以用迴歸模型來組合預測指標。Zarnowitz (1967) 則是指出國民所得的預測可以透過幾個不同預測的平均得到更佳的結果。實務上這樣的證據已經累積相當多，光是 Clemen (1989) 就已蒐集了 200 篇以上的文獻支持合併預測的優勢。Armstrong (2001) 審視了大量的實證文獻，發現在進行預測時可以透過合併不同的預測方法或資料集合得到更準確的預測。

在實際的作法上，Armstrong (2001) 建議研究者在結合個別的預測時，可以使用相同權重，或是在有證據充分顯示特定工具有較優異的表現時，可根據預測工具的過去表現來調整權重。Stock and Watson (2004) 提到簡單平均通常表現比複雜的方法出色，主要原因是個別預測準確度變化相當大，導致過去出色的權重過度依賴某一預測方法，隨著時空改變出現比簡單平均更大的誤差。由於其他的時空因素，例如：經濟景氣、政治環境、國際情勢等變化並非個別預測值能完全掌握，因此，均衡的攝取各種預測工具成了最可靠的做法。

本文在預測市場文獻上的貢獻，便是採用了不同的模型來盡可能捕捉一切與預測準確性有關的資訊。然而由於不同模型採用資料的範圍與方式不盡相同，造成在預測時無法具備全面性的優勢。為了結合不同模型的優勢，本文將各模型的預測予以加總，並依照 Armstrong (2001) 之建議比較不同的合併方法。

## 3 研究方法

在本節中，我們將在 3.1 節對選舉合約的預測結果可能性進行分類與定義，並據此提出兩種正確率指標。在 3.2 節中，本文將介紹解釋與被解釋變數，以及各鑑別模型的參數及訓

練方法。最後在 3.3 節，我們提出一套整合四個鑑別模型的合併預測方法。

### 3.1 事前鑑別與準確度之衡量

預測市場的最主要功能是於選前預測選舉結果。然而，並非每個合約都可以收集足夠的有效資訊。例如某些合約交易的過程可能十分清淡，其最後交易價格便可能不具參考價值。因此，我們希望利用鑑別模型，在選舉結果公告之前先行判讀預測市場交易價格所呈現的意涵（Chen et al., 2011; 林鴻文、童振源、與葉家興，2014）。

本文擬利用四種鑑別模型，在**選舉結果公告之前**將所有選舉預測事件合約區分成兩組：「鑑別正確預測組」與「鑑別未正確預測組」合約。「鑑別正確預測組」代表該合約的預測被鑑別模型判定為正確：亦即若該合約為所有候選合約中市價最高者，則預期該候選人會當選；反之，非最高價合約的候選人不會當選。另一方面，若某合約在選前被鑑別模型判定為「鑑別未正確預測組」合約，則表示按照該合約市場價格所作出的預測不會正確，亦即最高價合約的候選人不會當選，反而非最高價合約的候選人會當選。

在**選舉過後**，本文會比對每一合約選前的「最後一個交易日價格」所作出的預測與事後選舉結果是否相符。[7] 若選前的預測與選後公告結果一致，則該合約會被歸類為「事實正確預測組」合約。反之，若某合約按其最終日交易價格所進行的預測與選舉結果不符，該合約會被定義為「事實未正確預測組」合約。根據上述定義，本文將選舉預測事件分類如表 1：

表 1 鑑別模型對選舉預測事件的分類

| | 鑑別模型分類情況 | |
|---|---|---|
| **事後比對結果** | 鑑別正確預測組 | 鑑別未正確預測組 |
| 事實正確預測組 | TP | FN |
| 事實未正確預測組 | FP | TN |

表 1 即為資料探勘 (Data Mining) 研究中的分類矩陣 (classification matrix) 或混淆矩陣 (confusion matrix)，而其中的 TP, FN, FP, TN 分別代表以下意義：

- TP (True Positive) 組：選前落在「鑑別正確預測組」且選後為「事實正確預測組」。
- FN (False Negative) 組：選前落在「鑑別未正確預測組」但選後其實為「事實正確預測組」。

[7] 本文不使用最後一個月或其它期間來計算平均價格，其原因為選情往往不斷變化，一個月之間的實際選情可能會有相當的波動起伏。我們的目標是預測最後投票結果，因此採用投票前一日的作為計算價格的時間區間。

- FP (False Positive) 組：選前落在「鑑別正確預測組」但選後其實為「事實未正確預測組」。
- TN (True Negative) 組：選前落在「鑑別未正確預測組」且選後為「事實未正確預測組」。

據此，我們可以對所有鑑別模型的分類結果分別定義兩種正確率指標：[8]

- **鑑別正確準確率** (Accuracy Rate of Correct Identification, **ARCI**)：*TP/(TP+FP)* ，即分類問題 (classification problem) 研究中的陽性預測值 (Positive Predictive Value, PPV) 或 precision 指標，用以衡量在被鑑別模型認定為正確的合約中，事實上確實正確的比例。
- **鑑別不正確準確率** (Accuracy Rate of Incorrect Identification, **ARII**)：*TN/(FN+TN)*，即分類問題研究中的陰性預測值 (Negative Predictive Value, NPV) 指標，用以衡量在被鑑別模型認定為不正確的合約中，事實上的確不正確的比例。

ARCI 的意義為「選前」被鑑別模型判定市場預測會準確的合約中，「選後」證實預測無誤的比例。ARII 的意義為：選前被鑑別模型認定為不會準確的合約中，選後結果確實失準之合約比例。而一般的資料探勘研究中常用的基本指標如 recall (sensitivity) 和 f1-score，本文則呈報於附錄一中。[9]

## 3.2 解釋變數與鑑別模型

根據文獻，本文採用六大類變數作為解釋變數，被解釋變數則是個別合約事後到底是「事實正確預測組」還是「事實未正確預測組」。此六大類解釋變數分別為邊際交易者比例、市場共識程度、市場價量特徵、預測困難度、分身變數及價格變數共 40 個變數如表 2：

表 2 解釋變數類別與定義

| 變數類別 | 變數名稱 | 變數描述 |
|---|---|---|
| 邊際交易者 | *GP_share_lyc_R* | 前一年度總績效排名前 *R* 名當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *R*=100, 200, 300。 |

[8] 在 data mining 的領域中有非常多可用來評估鑑別模型的準則，每一種指標的設計都有不同的考量，其所適用的問題領域也不盡相同。要如何在特定的問題中選用合適的指標來評估鑑別模型端視研究者的需求而定 (Ferri, Hernández-Orallo, and Modroiu, 2009)，甚至同一種問題都會因為資料來源不同而需要使用不同的評估指標（如 Sokolova and Lapalme, 2009）。

[9] 就實務上而言，目前預測市場對選舉事件的預測已成為諸多媒體報導的對象，若在重要選舉事件上發生預測失準的情況不但會對預測市場的公信力造成相當負面的影響，而且也會產生相當高的社會成本。因此，如何在事前鑑別出即將失準的合約便是非常重要的目標，而這也是本文採用 ARII (NPV) 為績效評估指標的原因。

| | | |
|---|---|---|
| | *GP_share_lyc_S%* | 前一年度總績效排名前 *S*% 當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *S*=1, 5, 10。 |
| | *GP_share_365d_T* | 在最近 365 天清算合約的總績效排名前 *T* 名當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *T*=100, 200, 300。 |
| | *GP_share_365d_U%* | 在最近 365 天清算合約的總績效排名前 *U*% 當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *U*=1, 5, 10。 |
| | *GP_share_30d_V* | 在最近 30 天清算合約的總績效排名前 *V* 名當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *V*=100, 200, 300。 |
| | *GP_share_30d_W%* | 在最近 30 天清算合約的總績效排名前 *W*% 當中參與該合約的人數，占該合約清算日前一天總參與人數的比例, *W*=1, 5, 10。 |
| | *Limit_ratio_volume* | 限價成交口數占全部成交口數之比例。 |
| 市場共識程度 | *WBAS2_all* | 該合約所有未成交委託單的加權買賣叫價差距，計算公式：$\frac{\sum 賣單叫價\times賣單口數-\sum 買單叫價\times買單口數}{\sum 賣單口數+\sum 買單口數}$。 |
| | *Buy_sell* | 委買單口數相對於委賣單口數的比例。 |
| 市場價量特徵 | *Trades* | 參與該合約交易的交易量。 |
| | *Traders* | 參與該合約交易者的人數。 |
| | *Days* | 從該合約開始進行交易到合約清算前一天的交易天數。 |
| | *Volume* | 參與該合約的交易合約口數。 |
| | *Two_way* | 雙向交易者占全部交易者的比例。 |
| | *IP_share* | 臺灣交易者相對於全部交易者的比例，臺灣交易者是指帳號的 IP 註冊地點在臺灣。 |
| | *Traded_order_ratio* | 全部成交的委託單數量，除以全部的委託單數量。 |
| 預測議題困難程度 | *Highest-price* | 該合約組內，所有加權平均價中最高者。 |
| | *NC* | 合約組的合約數量。 |
| | *Price_gap* | 該合約組內，第一高價與第二高價的差距。 |
| 分身變數 | *Avatar_ratio_3* | 以 3 人同密碼即視為分身的假設，分身交易者人數占該合約全部交易者的比率。 |
| | *Avatar_Xd_ratio_3* | 以 3 人同密碼即視為分身的假設，結清日前 *X* 天參與該合約的所有交易者當中，其分身交易者人數占該合約全部交易者的比率, *X*= 15, 30, 365。 |
| | *Avatar_volume_ratio_3* | 以 3 人同密碼即視為分身的假設，分身所完成「交易口數」占 |

| | | |
|---|---|---|
| | | 該合約全部成交口數的比率。 |
| | *Avatar_volume_Yd_ratio_3* | 以 3 人同密碼即視為分身的假設，結清日前 *Y* 天該合約的成交筆數當中，分身所完成「交易口數」占該合約全部成交口數的比率, *Y*=15, 30, 365。 |
| 選舉特徵 | $P^w$ | 該合約的最後一天成交日的加權平均價格。 |

資料來源：未來事件交易所 (2011)，網址：http://xfuture.org。

### 3.2.1 SVM 鑑別模型作法

支持向量機 (SVM) 為一種機器學習演算法，經常被使用在二元分類問題上 (Cortes and Vapnik, 1995)。Vapnik and Lerner (1963) 提出的原始版本為線性分類，然而資料不一定能由一個線性的超平面完美地進行分割。到了 Boser, Guyon, and Vapnik (1992) 則建議用 Aizerman, Braverman, and Rozonoer (1964) 的非線性轉換或核心函數技巧 (kernel trick) 來處理非線性的問題。SVM 利用核心函數 (kernel function) 把屬性特徵座標轉換到另一個更高維度的特徵空間 (feature space)。在此特徵空間中，SVM 將可以找到將資料分成兩群的最適超平面以進行資料分類。

在實際的計算上，我們以線性支持向量機為例，假設有一組樣本 $\mathrm{D} = \{(\mathbf{x}_i, y_i) | \mathbf{x}_i \in \mathrm{R}^N, y_i \in \{-1, 1\}\}_{i=1}^{n}$，$\mathbf{x}_i$ 為 $N$ 維的向量，共有 $n$ 個樣本點。SVM 則是在向量空間中找到一個超平面，如圖 1 的實線所示 $\mathbf{w} \cdot \mathbf{x} - b = 0$，其中 $\mathbf{w}$ 為法向量，其到原點的距離為 $b/\|w\|$。因此，如果樣本資料是線性可分的，則我們可以產生距離最大的兩個超平面，$\mathbf{w} \cdot \mathbf{x} - b = 1$ 與 $\mathbf{w} \cdot \mathbf{x} - b = -1$ 的距離，亦即極大化 $2/\|w\|$。

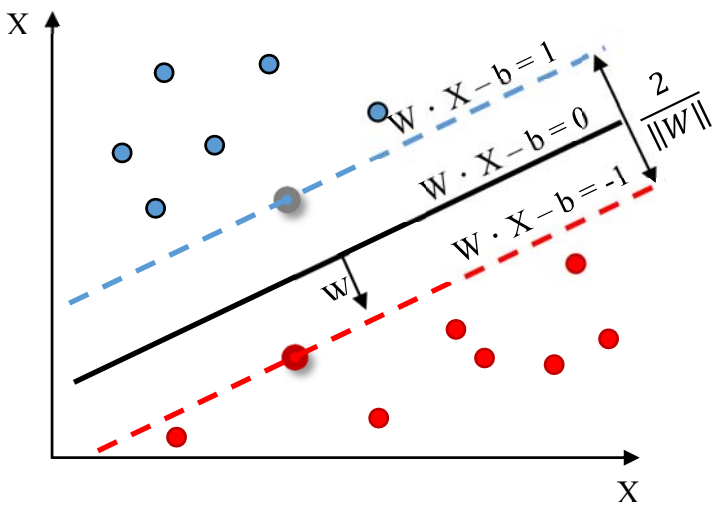


圖 1 支持向量機分類示意圖

SVM 會先利用核心函數將資料點由原始維度（在本文中為 40 維）的空間中轉換到一個更高維度的特徵空間，以利於超平面的找尋工作。文獻上常用的非線性轉換的核心函數有多項式 (polynomial)、S 型 (sigmoid)、及高斯 (Gaussian) 函數，後者也稱為徑向基函數 (radial basis function, RBF)。本文所採用的核心函數為 RBF，因為線型及 S 型核心都是 RBF 的特例。為了克服過度配適 (overfitting) 的問題，本文在訓練模型時使用 10 次交叉驗證法 (10-fold cross-validation)。本文所使用的軟體為 DTREG，並採用 Chang and Lin (2005) 所提出的 LIBSVM 程序。

### 3.2.2 PCA-DA 鑑別模型作法

本文採用主成份分析 (PCA) 的目的有二：(1) 將全面且大量的可用變數全部納入模型體系—本文希望運用 PCA 的特質，在眾多變數中歸納出這些變數的重要特徵，這些資料的主要特徵即稱為「主成份變數」；(2) 避免共線性的問題—主成份的個數不僅會少於原始變數個數，且主成份之間不會有相關性。本文對選舉候選人預測當選與否，必需區分「正確預測」、「未正確預測」兩種預期結果。有鑒於此，兩個多變量分析理論的結合，有助於 DA 對樣本的區分能力及預測準確度提升，其詳細理論可參閱 Loh and Vanichsetakul (1988)。

主成份分析結合判別分析 (PCA-DA) 鑑別模型是將主成份變數當作 DA 的自變數，且將「事實正確預測組」和「事實未正確預測組」，視為 DA 的分組變數。首先 PCA 將會由 40 個解釋變數中，濃縮萃取出較少數的主成份來描述整個資料型態及性質，每一主成份皆為所有原始解釋變數標準化後的線性組合，因此一個主成份被視為一個新的自變數：

$$X_1^* = W_{11}X_1 + W_{12}X_2 + W_{13}X_3 + \cdots + W_{1p}X_p \text{，} \tag{1}$$

其中 $X_1^*$ 為該合約的第一個主成份分數，$W_{1i}$為該合約的第一主成份之第 $i$ 個因素負載量，$X_j$ 為該合約的第 $j$ 個原始變數標準化數值。

接下來，本文將「事實正確預測組」和「事實未正確預測組」作為 DA 之分類變數 (grouping variable)，即因變數，並以 PCA 所得到的主成份變數作為判別變數 (discriminant variable)，即自變數，以建立判別函數 (discriminant function) 如下：

$$f^1 = \alpha_0 + \alpha_1 X_1^* + \alpha_2 X_2^* + \cdots + \alpha_k X_k^* \text{，} \tag{2}$$

$$f^2 = \beta_0 + \beta_1 X_1^* + \beta_2 X_2^* + \cdots + \beta_k X_k^* \text{。} \tag{3}$$

本文先分別對「事實正確預測組」及「事實未正確預測組」合約分別估計函數之係數值，所得到的判別函數分別為 $f^1$及 $f^2$。在實際鑑別某個合約依市場之預測是否會正確時，將該合約之主成份分別代入 $f^1$ 及 $f^2$，取其值大者並將合約認定為該類。[10] 本文使用的為

[10] 例如若某合約之主成份代入 $f^1$ 後所得之值較代入 $f^2$ 後所得之值為大，則該合約即被歸類為「鑑別正確預測組」，亦即 PCA-DA 模型認為依其市價所作出的預測將會是正確的；反之若該合約之主成份代入 $f^1$ 後

Fisher 判別分析，並使用 SPSS 進行資料處理與模型運算。完整的 PCA-DA 鑑別流程可見於圖 2。

### 3.2.3 Logit 鑑別模型作法

在採用 Logit 迴歸模型來鑑別每一合約是否滿足最高價準則，本文所採用的標準格式為：

$$\mathrm{E}(y) = \frac{e^{a+b_1x_1+b_2x_2+\cdots+b_nx_n}}{1+e^{a+b_1x_1+b_2x_2+\cdots+b_nx_n}}, \tag{4}$$

其中 $x_n$ 為解釋變數。迴歸機率配適值可透過 Logit 轉換過程（勝算比再取自然對數）得到：

$$\ln\left(\frac{\hat{y}}{1-\hat{y}}\right) = \hat{a} + \widehat{b_1}x_1 + \widehat{b_2}x_2 + \cdots + \widehat{b_n}x_n。 \tag{5}$$

本文使用 STATA 計量軟體得到 $\widehat{b_1}$, $\widehat{b_2}$, $\cdots$ , $\widehat{b_n}$，再運用文獻常使用的樣本平均數 (sample mean)，作為我們 Logit 鑑別分組門檻。舉例而言，將某一合約之解釋變數 ($x_1$ , $x_2$ , $\cdots$ , $x_n$) 輸入式 (5) 後可得一個配適值，此一數值若比樣本平均數大時，則我們判定依據該合約市價所作出的預測將會正確。相反地，若該合約的配適值小於樣本平均數時，則本文模型判定依據此合約市價所作出的預測將不會正確。

### 3.2.4 DT 鑑別模型作法

決策樹 (DT) 是機器學習中一種常用於資料探勘的演算法，它包含三個主要成分，根節點、條件與終端節點。決策樹的建立方法主要有三個原則：(1) 以全部樣本資料為根節點；(2) 找出變異量最大的變數作為分岔的依據；(3) 若分類結果的正確率或涵蓋率未滿足條件時，則再依據條件下最大變異量的變數長出分岔。目前常被使用的決策樹演算法包括 Classification And Regression Tree (CART)、C4.5、CHAID (Hastie et al., 2001)。其中 C4.5 和 CART 相當類似，差別在於其節點分支數目可以不受二分法的限制，可產生不同數目的分支。本文使用 DTREG 軟體進行決策樹演算法，其所使用的分類決策樹為 CART 演算法。

在實際進行分類時，決策樹演算法會檢視每一個解釋變數的類別，並計算每一個解釋變數類別下目標變數無法分類的程度。例如，我們的目標變數有 $m$ 種分類標籤，而 $f_i$ 為第 $i$ 種標籤所占的比例，而計算無法分類程度的幾種準則有三種：

---

所得之值較代入 $f^2$ 後所得之值為小，即代表 PCA-DA 模型認為依其市價所作出的預測將不會正確。

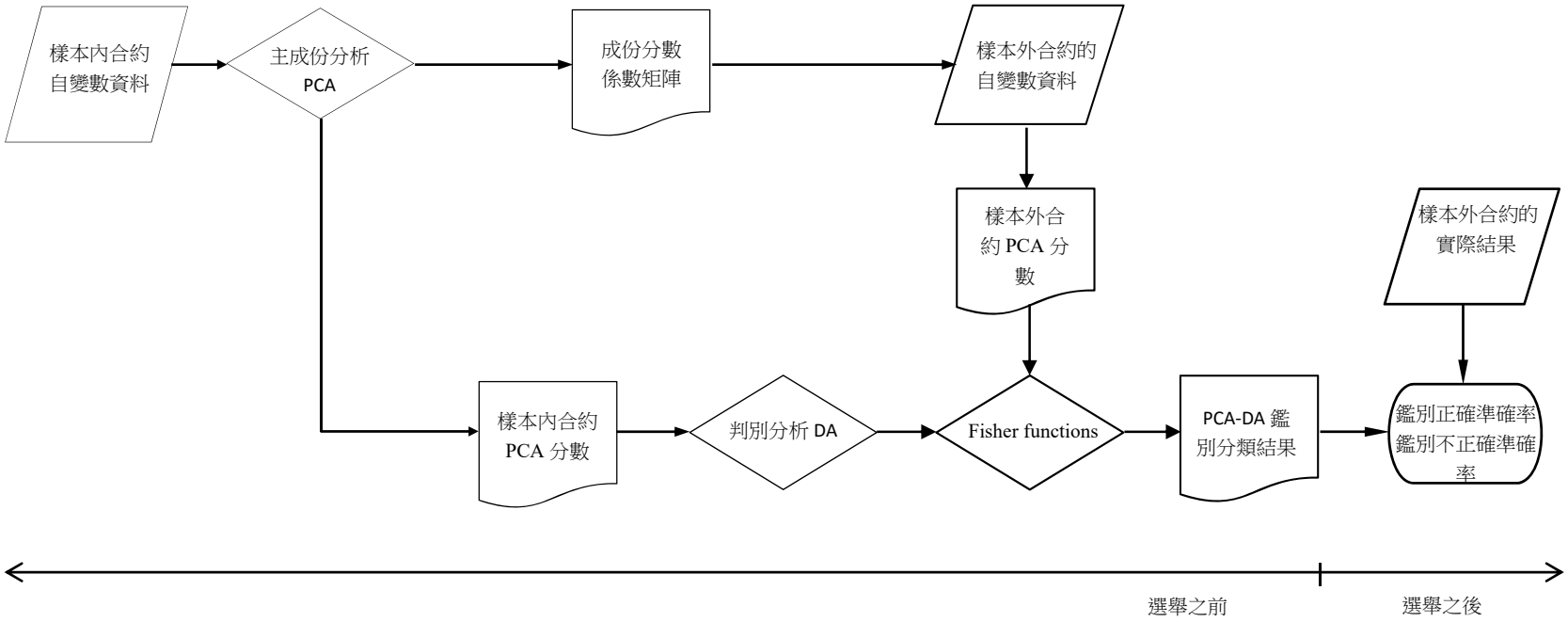


圖 2 主成份分析結合判別分析之鑑別流程圖

1. 吉尼雜質 (Gini impurity)：

$$\mathrm{I_G}(f) = \sum_{i=1}^{m} f_i(1-f_i) = 1 - \sum_{i=1}^{m} f_i^2 \text{。} \tag{6}$$

2. 熵 (entropy)：

$$\mathrm{I_E}(f) = -\sum_{i=1}^{m} f_i \log_2 f_i \text{。} \tag{7}$$

3. 分類錯誤率 (classification error)：

$$\mathrm{I_C}(f) = 1 - \max\{f_i\} \text{。} \tag{8}$$

不論使用何種方式衡量，其值愈大就代表無法分類的情況愈嚴重。決策樹會嘗試利用每一個解釋變數來分類，並據以計算利用該變數作為分類變數後所得到的資訊增益（information gain）。某一個解釋變數的資訊增益越大，就代表使用該變數進行分類後目標變數無法分類程度降低越大。決策樹演算法在逐一比較各解釋變數後，會採用資訊增益最大的變數率先進行分類，並重覆以上的步驟直到無法改善分類情況為止。在實際的研究問題中，通常無法只用一個變數就達到完美的分類，所以還需要再往下做子集合的分割，直到資訊增益不符合效益為止。[11]

## 3.3 合併鑑別法

Armstrong (2001) 建議研究者結合個別預測的方式，可以使用相同權重，或是在有證據充分顯示特定工具有較優異的表現時，可根據預測工具的過去表現來調整權重。本文將嘗試以下三種加權機制以及不同的組合方式，以找出最適的合併方法：

1. **均等權重法**：每項工具的權重相等。
2. **ARCI 加權法**：根據每項工具 ARCI 的相對大小來給定權重。
3. **ARII 加權法**：根據每項工具 ARII 的相對大小來給定權重。

我們設四種鑑別模型的判定值分別為 $f_{SVM}$、$f_{DA}$、$f_{Logit}$、及 $f_{DT}$。個別模型對每一個選舉合約的判斷皆為兩種結果：1 表示判定預測市場的結果將準確預測選舉結果（鑑別正確預測組）；0 則表示市場結果未能準確預測（鑑別未正確預測組）。因此，合併預測指標可以寫成：

$$f = w_1 f_{SVM} + w_2 f_{DA} + w_3 f_{Logit} + w_4 f_{DT}, \quad \sum_i w_i = 1 \text{。} \tag{9}$$

[11] 決策樹的應用亦有可能陷入過度配適的困境。經由後續改良，其中最受到關注的相關演算法之一稱為隨機森林（Random Forests）(Breiman, 2001)。隨機森林根據不同樣本的隨機抽樣產生對應的隨機決策樹，並以這些決策樹來驗證剩下的樣本，產生的森林透過投票的機制選出有潛力的分類樹來判斷未知的樣本 (Liaw and Wiener, 2002)。Pal (2005) 指出隨機森林的優點是可以處理遺漏資料的變數，實證上其表現優於傳統決策樹，但與 SVM 的表現不分上下。本文曾比較隨機森林演算法和決策樹在相同的樣本測試中的績效，發現以 precision, NPV, recall rate, 或 f1-score 等指標來評量，決策樹的表現在大部分的測試中均等於或優於隨機森林演算法。

此外，由於鑑別模型的預測值（判定值）不是實數，而是類別的概念，因此如何解讀便是本研究與一般預測研究的最大不同點。為了產生二元預測結果，在合併的過程中本文提出了兩種規則來彙整各個工具的預測：

1. 共識決規則 (consensus rule)
2. 多數決規則 (majority rule)

在共識決的規則下，當所有的工具皆有相同的見解便可直接將此共識作為結論。但若市場的預測無法從所有的工具得到一致正向或負向的支持，亦即各工具鑑別的結果不一，則在共識決的規則下我們的彙整結果便會是「無結論」。

在多數決原則下，我們依第 (9) 式來計算合併預測值。若合併預測值大於 0.5 則彙整的結果便是市場的預測將被視為「正確」，反之則是「不正確」。但若正面或負面意見皆未超過半數（等於 0.5），則彙整的結果便會是「無結論」。茲以下列式子來表達前述的概念：

$$F^M = \begin{cases} 1, & \text{if } f > 0.5 \\ 0, & \text{if } f < 0.5 \\ \text{un-consensus}, & \text{if } f = 0.5 \end{cases}; \tag{10}$$

$$F^C = \begin{cases} 1, & \text{if } f = 1 \\ 0, & \text{if } f = 0 \\ \text{un-consensus}, & \text{if } 0 < f < 1 \end{cases}, \tag{11}$$

其中 $F^M$ 為多數決預測指標，只要加權後的指標超過 0.5 則判定為準確；而 $F^C$ 為共識決預測指標，需要全體判定為 1 才判定為準確。由此可知，$F^M$將會和權重有關，而 $F^C$ 則不受權重影響。最後我們以表 3 來呈現合併預測使用的各種權重組合，並羅列所有加權組合可能方式下的權重計算法則。為了找出最佳的組合預測方法，本文將嘗試不同的組合規模和加權方法。

由表 3 中可明確得知共有三種加權方法：均等權重法、ARCI 加權法、以及 ARII 加權法。模型組合的規模則可由表中的第一欄得知。[12] 例如，模型規模為 1 代表單一模型（因此被選取的模型權重為 100%，而其餘模型權重為 0），而如果模型規模為 3 則代表從四個鑑別模型中選出三個來進行組合預測。若是組合方式為均等權重的話則每個模型的權重均為 33%，否則可依 ARCI 或 ARII 來制定權重。在 ARCI 加權法中，本文以四種鑑別模型在樣本內測試的 ARCI 來決定權重值：

$$w_i = \frac{ARCI_i}{\sum_{j=1}^{4} ARCI_j}, \quad i = 1,2,3,4。 \tag{12}$$

而在 ARII 加權法中，本文以四種鑑別模型在樣本內測試的 ARII 來決定權重值：

[12] 因此事實上會有 $C_1^4 + C_2^4 + C_3^4 + C_4^4$ 共 15 種組合方法，每種組合方法又有「均等權重法」、「ARCI 加權法」、及「ARII 加權法」三種可能的加權方式，在每種加權方式下又有「共識決」及「多數決」兩種合併預測法。

$$w_i = \frac{ARII_i}{\sum_{j=1}^{4} ARII_j},\ i = 1,2,3,4， \tag{13}$$

其中 $w_1$、$w_2$、$w_3$、$w_4$ 分別表示 SVM、DA、Logit、及 DT 四種鑑別模型的權重值。

表 3 四種鑑別模型的使用以及權重分配

| 鑑別模型規模 | 均等權重法 | | | | ARCI 加權法 | ARII 加權法 |
|---|---|---|---|---|---|---|
| | SVM | DA | Logit | DT | | |
| 1 | **100%** | 0% | 0% | 0% | 單一鑑別模型不受權重影響 | |
| 1 | 0% | **100%** | 0% | 0% | | |
| 1 | 0% | 0% | **100%** | 0% | | |
| 1 | 0% | 0% | 0% | **100%** | | |
| 2 | **50%** | **50%** | 0% | 0% | 以各模型在訓練樣本期間的 ARCI 加權 | 以各模型在訓練樣本期間的 ARII 加權 |
| 2 | **50%** | 0% | **50%** | 0% | | |
| 2 | **50%** | 0% | 0% | **50%** | | |
| 2 | 0% | **50%** | **50%** | 0% | | |
| 2 | 0% | **50%** | 0% | **50%** | | |
| 2 | 0% | 0% | **50%** | **50%** | | |
| 3 | **33%** | **33%** | **33%** | 0% | | |
| 3 | **33%** | **33%** | 0% | **33%** | | |
| 3 | **33%** | 0% | **33%** | **33%** | | |
| 3 | 0% | **33%** | **33%** | **33%** | | |
| 4 | **25%** | **25%** | **25%** | **25%** | | |

### 3.4 資料來源與樣本外測試方法

本文的資料來源為「未來事件交易所」(http://xfuture.org/) 發行的期貨合約交易記錄，選舉事件資料包括：2008 年立委選舉前、2008 年立委選舉、2008 年總統大選、2009 年縣市長選舉、立委補選和 2010 年五都選舉至 2011 年共 650 個合約之資料。

為了檢驗各鑑別模型的判斷準確度，每個單一模型或合併模型都需進行樣本外測試。在進行樣本外測試時，我們將所有的選舉合約切割為訓練樣本 (training sample) 以及測試樣本 (testing sample) 兩部分。一般也會將「訓練樣本」稱作「內樣本」(in-sample)，將「測試樣本」稱作「外樣本」(out-of-sample)。也就是說，我們將以訓練樣本來估計模型的參數值或訓練模型以求得模型結構，再用所得到的模型來對測試樣本進行測試，以評估此模型的預測

能力。[13] 在本文所進行的樣本外測試方法中，我們共有四個切分樣本的方法，具體切分方式如表 4 所示。其中 A、B、C 三種樣本切分方式的目的是針對不同年份及性質的選舉事件進行測試，而 D 樣本切分法則是最後為了進行大樣本外測試所採取的作法。

表 4 訓練樣本及測試樣本切分方式

<table>
<tr><th rowspan="2">選舉事件</th><th rowspan="2">合約數</th><th colspan="4">樣本切分方法</th></tr>
<tr><th>A</th><th>B</th><th>C</th><th>D</th></tr>
<tr><td>2008 年立委選舉前</td><td>45</td><td rowspan="2">訓練樣本</td><td rowspan="4">訓練樣本</td><td rowspan="6">訓練樣本</td><td rowspan="2">訓練樣本</td></tr>
<tr><td>2008 年立委選舉</td><td>288</td></tr>
<tr><td>2008 年總統大選</td><td>78</td><td>測試樣本</td><td rowspan="7">測試樣本</td></tr>
<tr><td>2008 年總統大選後，<br>2009 年縣市長選舉前</td><td>58</td><td rowspan="6"></td></tr>
<tr><td>2009 年縣市長選舉</td><td>48</td><td>測試樣本</td></tr>
<tr><td>2009 年立委補選</td><td>25</td></tr>
<tr><td>2010 年五都選舉</td><td>86</td><td>測試樣本</td></tr>
<tr><td>2010 年五都選舉後</td><td>22</td></tr>
<tr><td>合計</td><td>650</td><td></td><td></td><td></td><td></td></tr>
</table>

## 4 單一模型之鑑別績效

在本節中，我們將先使用最大資訊集合—即全部 650 個合約，對四個個別模型作測試，以瞭解這些個別模型本身在樣本內（4.1 小節）及樣本外（4.2 小節）的鑑別能力。而在後續使用合併模型進行樣本外鑑別時，將參考個別鑑別模型在樣本內測試的表現作為加權依據。

### 4.1 樣本內測試（全樣本）

在此測試中，每一個模型皆以 650 個選舉合約來估計參數，接著便對同樣的 650 個合約分別進行鑑別，而其結果則整理於表 5。由表 5 中我們可以發現四種工具的 ARCI 都相當地高，而且彼此間的表現也十分接近。若從 ARII 的角度來比較，我們可以清楚地看到所有被 SVM 模型認定為預測不會正確的合約（鑑別未正確預測組），選舉後事實上也全數都是預測不正確的合約（事實未正確預測組），因此其 ARII 為 100%。在這四種模型中，我們發現僅 PCA-DA 的 ARII 較低（雖然其 ARCI 亦約有 95%的準確率）。若評比四種工具在 ARCI 與 ARII 的綜合表現，則 SVM 的鑑別力是最好的，但與 Logit 及 DT 的鑑別結果差距不大。四種鑑別工具在此處的表現（表 5），將按式 (12) 及 (13) 計算權重值，並用於式 (9) 中

[13] 此即所謂的「樣本外測試」(out-of-sample test)。

作為之後合併模型時的加權依據。

表 5 個別模型之樣本內測試：全部選舉合約

| 模型 | TP | FP | FN | TN | ARCI | ARII |
|---|---|---|---|---|---|---|
| (1) SVM | 615 | 1 | 0 | 34 | 0.9984 | 1.0000 |
| (2) PCA-DA | 611 | 33 | 4 | 2 | 0.9488 | 0.3333 |
| (3) Logit | 601 | 10 | 14 | 25 | 0.9836 | 0.6410 |
| (4) DT | 612 | 14 | 3 | 21 | 0.9776 | 0.8750 |

## 4.2 樣本外之鑑別績效

在上一小節已確認四個模型在 ARCI 上有相當不錯的準確率，不過高準確率有可能是因為預測對象是模型曾經處理過並據以找出最適參數的合約。因此在本節中我們針對四種模型作樣本外測試，以瞭解個別模型在面對陌生合約時的鑑別能力。

A. 2008 年總統大選

在 2008 年總統大選中，未來事件交易所總共發行了 78 個相關合約。因此在此測試中，我們先以 2008 年總統大選前的 333 個合約作為訓練樣本 (training sample)，在鑑別模型參數確立或模型架構決定後，再對 2008 年總統大選的 78 個合約進行樣本外鑑別，鑑別結果如表 6 所示。

表 6 個別模型之樣本外測試：2008 年總統大選

| 模型 | TP | FP | FN | TN | ARCI | ARII |
|---|---|---|---|---|---|---|
| 樣本內 | | | | | | |
| (1) SVM | 307 | 20 | 2 | 4 | 0.9388 | 0.6667 |
| (2) PCA-DA | 304 | 15 | 5 | 9 | 0.9530 | 0.6429 |
| (3) Logit | 303 | 7 | 6 | 17 | 0.9774 | 0.7391 |
| (4) DT | 306 | 12 | 3 | 12 | 0.9623 | 0.8000 |
| 樣本外 | | | | | | |
| (1) SVM | 73 | 3 | 1 | 1 | 0.9605 | 0.5000 |
| (2) PCA-DA | 72 | 4 | 2 | 0 | 0.9474 | 0.0000 |
| (3) Logit | 63 | 0 | 11 | 4 | 1.0000 | 0.2667 |
| (4) DT | 66 | 0 | 8 | 4 | 1.0000 | 0.3333 |

由上表觀察可知，四個方法的樣本外鑑別正確準確率，與樣本內的鑑別正確準確率皆相近，而 SVM、Logit、DT 甚至在樣本外可以得到更高的鑑別正確準確率。而樣本外的鑑別未正確準確率，則比樣本內的測試結果更低。

B. 2009 年縣市長選舉

在 2009 年縣市長選舉中，未來事件交易所總共發行了 48 個相關合約。因此在此測試中，我們納入過去 469 個合約（含 2008 年立委選舉）作為樣本內訓練資料，在模型參數確立或模型架構決定後再對 2009 年縣市長選舉的 48 個合約進行樣本外鑑別，鑑別結果如表 7 所示。

此次的樣本內資料數量，明顯超越 2008 年總統大選樣本內 333 個合約，故在 2009 年縣市長選舉的鑑別準確率上，除了 PCA-DA 鑑別模型仍有 3 個合約鑑別錯誤外，其它三個模型的鑑別準確率已有相當程度的改善，不但在 ARCI 上四種模型的樣本外測試都能達到 100%，而且由於在 2009 年縣市長選舉中，市場合約的最高價預測事後來看都是正確的，也就是事實上沒有「事實未正確預測組」的合約，而除了 PCA-DA 外，其餘三個鑑別模型也沒有將任何一個 2009 縣市長選舉合約判定為不正確（鑑別未正確預測組），所以該三種模型的 ARII 無法計算（因此標為“-”）。

雖然欠缺 ARII 值表面上看起來讓我們無法評估模型的表現，但這點其實印證本文所採用的模型是相當可靠的。我們需要注意的是由於在測試樣本中（樣本外）事實上沒有任何一個合約屬於「事實未正確預測組」，所以在 ARII 的表現上是不可能出現 100% 的。相反地，此時最佳的 ARII 結果正應當是“-”（因為不該將任何一個合約判定為鑑別未正確預測組）。我們注意到雖然四種模型在樣本內的訓練時都曾遇過「事實未正確預測組」的合約，也都曾將合約判定為「鑑別未正確預測組」，但在面對樣本外的測試合約時卻沒有將事實上屬於「事實正確預測組」的合約誤判為「鑑別未正確預測組」。由這點，我們可以間接推論這些模型在訓練樣本內應該都有學習到部分的正確知識，所以雖然樣本內的 ARII 不高，但除了 PCA-DA 外的三個模型在樣本外的 ARII 皆為“-”。

C. 2010 年五都選舉

在 2010 年五都選舉中，未來事件交易所總共發行了 86 個相關合約。因此在此測試中，我們以五都選舉前的 542 個合約（含 2008 年立委選舉及 2009 年縣市長選舉）作為訓練樣本，在模型參數確立或模型架構決定後再對 2010 年五都選舉的 86 個合約進行樣本外鑑別，鑑別結果如表 8 所示。由表中我們得到與之前分析類似的結果：四種工具在樣本外的 ARCI 反而比樣本內的表現還好，而在 ARII 上除了 PCA-DA 外的三種模型都能夠意識到所有合約都是正確的，所以無法計算 ARII。

表 7 個別模型之樣本外測試：2009 年縣市長選舉

| 模型 | TP | FP | FN | TN | ARCI | ARII |
|---|---|---|---|---|---|---|
| | | | 樣本內 | | | |
| (1) SVM | 436 | 28 | 2 | 3 | 0.9397 | 0.6000 |
| (2) PCA-DA | 428 | 21 | 10 | 10 | 0.9532 | 0.5000 |
| (3) Logit | 424 | 8 | 14 | 23 | 0.9815 | 0.6216 |
| (4) DT | 438 | 17 | 0 | 14 | 0.9626 | 1.0000 |
| | | | 樣本外 | | | |
| (1) SVM | 48 | 0 | 0 | 0 | 1 | - |
| (2) PCA-DA | 45 | 0 | 3 | 0 | 1 | 0 |
| (3) Logit | 48 | 0 | 0 | 0 | 1 | - |
| (4) DT | 48 | 0 | 0 | 0 | 1 | - |

表 8 個別模型之樣本外測試：2010 年五都選舉

| 模型 | TP | FP | FN | TN | ARCI | ARII |
|---|---|---|---|---|---|---|
| | | | 樣本內 | | | |
| (1) SVM | 505 | 30 | 2 | 5 | 0.9439 | 0.7143 |
| (2) PCA-DA | 497 | 30 | 10 | 5 | 0.9431 | 0.3333 |
| (3) Logit | 493 | 10 | 14 | 25 | 0.9801 | 0.6410 |
| (4) DT | 501 | 17 | 6 | 18 | 0.9672 | 0.7500 |
| | | | 樣本外 | | | |
| (1) SVM | 86 | 0 | 0 | 0 | 1 | - |
| (2) PCA-DA | 85 | 0 | 1 | 0 | 1 | 0 |
| (3) Logit | 86 | 0 | 0 | 0 | 1 | - |
| (4) DT | 86 | 0 | 0 | 0 | 1 | - |

總結以上三組樣本外測試的結果，我們發現在 2008 年總統選舉以後，本文所提出的四種模型之樣本外鑑別正確準確率 (ARCI) 皆已達 100%。在比例上，在兩次選舉的鑑別過程，分別於 48 個與 86 個合約都能連續正確判定，表示此四個模型已有能力區隔出未來將會符合最高價準則的合約。

但我們也發現隨著資料數量的增加，樣本內的鑑別未正確準確率 (ARII) 不但沒有因為樣本數增加而有所改善（一直呈現波動現象），且在第一組測試的樣本外 ARII 上表現也相當不理想。因此，我們將在下一節中介紹四個模型重新組合後的合併預測方法，希望能結合各種鑑別工具的特性，截長補短以找出更好的鑑別模型。

## 4.3 解釋變數之邊際貢獻

文獻上與預測市場準確度有關的變數與討論相當多（詳見 2.1 節之文獻回顧），到底哪些變數較為關鍵呢？本文使用了 40 個可能的解釋變數（表 2），在本小節中我們將整理出此 40 個變數之邊際貢獻度，藉此來檢視四種鑑別模型在鑑別預測市場準確度時各變數的重要性。

在探討各變數之邊際貢獻度時，須注意各鑑別模型使用資料的方式不盡相同，有的是線性方式（如 Logit），有的則為非線性（如 DT）；有的模型直接使用原始變數，有的則間接使用（如 PCA-DA）。因此，我們採取一種最直觀的作法來討論變數的重要性：若剔除掉某一變數後鑑別模型的表現變差，即代表該變數具一定程度之重要性；反之若剔除某一變數後鑑別模型準確度不變（甚至上升），則代表該變數可有可無（可考慮忽略之）。[14]

首先，針對 Logit, SVM, DT 三個模型，本文在 10 次交叉驗證過程的每組訓練及測試樣本中，分別計算採用全部變數及剔除某一特定變數後的表現差異（以 ARCI 及 ARII 衡量之）。最後再找出剔除變數後表現落差最大者，推論其邊際貢獻較為顯著。分析的詳細步驟與結果詳列於附錄二中。由附錄二的表 19 中我們可以觀察到不同的模型及衡量指標會挑出不同的重要變數。就 SVM 而言，若由 ARCI 衡量，則最重要的三個變數依序為邊際交易者中的 *GP_share_365d_10*、*Limit_ratio_volume*，和分身變數中的 *Avatar_volume_ratio_3*、*Avatar_volume_30d_ratio_3*、與 *Avatar_volume_365d_ratio_3*；若由 ARII 衡量，則依序為 *Price_gap*、*Avatar_15d_ratio_3*、與 *Avatar_365d_ratio_3*。對 DT 而言，若由 ARCI 衡量則 *GP_share_365d_10* 是唯一不可缺少之變數；若由 ARII 衡量則為 *Highest-price* 及 *GP_share_365d_10*。就 Logit 而言，若由 ARCI 衡量則前三重要之變數分別為 *GP_share_lyc_300*、*Avatar_volume_30d_ratio_3*、及 *Volume*；若由 ARII 衡量則前三重要者為 *Avatar_volume_30d_ratio_3*、*GP_share_30d_5*、與 *GP_share_365d_10*。就整體頻率而言，吾人可歸納出邊際交易者（*GP_share* 等）是為最重要的變數，分身變數（*Avatar* 等）居次，而預測議題困難程度為第三（變數分類詳見表 2）。

就 PCA-DA 而言，由於此模型不直接使用原始變數，而是先由原始變數中找出主成份變數，再以主成份變數進行判別分析，因此宜採取的作法是一次剔除一個主成份（而非原始變數），再將其表現與未剔除前比較。藉此我們可以找出影響最大之主成份，再探討該主成份之內容。葉家興等（2012）已就 PCA-DA 剔除各主成份後之 ARCI 及 ARII 進行比較。由該文得知在十三個主成份中最重要者依序為第五、最一、及第六主成份，而由於主成份是由原始解釋變數所組成，因此可以由三個主成份占比最重者各自反推出最重要的解釋變數分別為 *Days*、邊際交易者（*GP_share*）、及分身變數（*Avatar*）。

[14] 感謝本文在審查過程中某匿名審查委員之建議，本文即依循該委員所提之作法來分析變數之邊際貢獻度。

# 5 合併模型之鑑別績效與比較

在本節中，我們將在 5.1 節探討合併鑑別模型的表現。在 5.2 節中我們將訂定一系列的目標函數來對預測市場的事前判斷進行風險控管。最後，在 5.3 節中我們將比較合併鑑別模型與單一鑑別模型的正確性，並比較不同合併權重法之效果。

## 5.1 合併模型鑑別結果

在合併個別的鑑別模型時，我們可以將本文所採用的四個模型視為四位預測「專家」，此四位專家在事前對每一合約，會進行投票表決該合約應歸類於「鑑別正確組」或「鑑別未正確組」。在表決的過程中，本文設計了兩種投票規則可供遵循—「共識決規則」 (consensus rule) 及「多數決規則」 (majority rule)。此外，本文亦嘗試讓過去具有較佳準確率的「專家」，得到較高的投票權重。而權重值則是使用它們在樣本內的鑑別正確準確率 (ARCI) 或鑑別未正確準確率 (ARII) 來配置。若再加上原先權重相同的配置方法，我們總共有三種權重配置法。

在接下來的分析中，5.1.1 節首先從全部樣本的開始逐一測試各種合併鑑別模型的鑑別表現，接著在 5.1.2 節我們探討的則是各合併模型的樣本外鑑別能力。

### 5.1.1 樣本內測試（全樣本）

在樣本內測試中，每一個模型皆以 650 個選舉合約來估計參數，接著便對同樣的 650 個合約分別進行判定。我們使用三種不同的加權方法進行合併，結果整理於表 9。在表 9 中，由於在產生合併模型時可以從四種專家中任意抽取其中兩位、三位、或四位全選來進行投票，因此總計有 11 種表決組合。[15] 值得一提的是，在均等權重法中，兩位專家在多數決與共識決規則具有相同的鑑別結果，因此在表中我們僅需列出三位及四位專家共識決之鑑別結果。表中的符號 M 代表多數決機制，符號 C 代表共識決機制，而其後的數字代表合併的模型代碼，其中 1 代表 SVM，2 代表 PCA-DA，3 代表 Logit，4 代表 DT。例如 M124 即代表將 SVM、PCA-DA、DT 以多數決方法合併而成的鑑別模型。

在 ARCI 加權法下，我們在合併模型時參照了個別鑑別模型在 4.1 小節中的 ARCI 表現來給定其權重（表 5）。在 ARII 加權法下，我們同樣參照了個別鑑別模型在表 5 的 ARII 表現來決定其權重。而表 9 中的 ARCI 和 ARII 值則是合併完成後的模型在對 650 個選舉合約進行鑑別時的鑑別正確及未正確準確率。由於共識決需要 100% 一致通過才能鑑別，

[15] $C_2^4 + C_3^4 + C_4^4 = 11$。

所以只要鑑別模型之間的意見不同，不論加權方法為何結果都會無法判定，因此採用共識決的合併模型鑑別結果在三種加權方法下將會完全相同。[16]

在均等權重法中，若以 ARCI 及可鑑別出最多的正確組合約 (TP) 來評估模型，SVM 模型最具預測優勢。若以 ARII 及可鑑別出最多的未正確組合約 (TN) 為比較前提，同樣是由 SVM 模型勝出。若以 ARCI 作為評量標準來比較合併鑑別模型，則 SVM、M12、M13、M14 表現最佳—它們不但都可以鑑別出 615 個合約，而且 ARII 亦達 100%。在 ARII 加權法下我們得到的數據與 ARCI 加權法的結果相似，SVM 的表現仍能與許多合併鑑別模型相匹敵。

在整合三個權重機制的數據後，我們可以得知 SVM 模型表現皆為最佳，而各種合併鑑別模型的表現都劣於、或至少不高於 SVM 這個單一鑑別模型的表現。根據這個結果，再考量到合併模型所牽涉的分析複雜度及成本，似乎可以讓人們得到一個結論：簡單的 SVM 模型即可成功鑑別預測市場預測之正確性，沒有必要花費額外成本進行合併鑑別。但事實上真的如此嗎？

理論上合併預測較單一模型預測為優的原因之一，在於每一個個別模型都可能利用了資料中不同層次或屬性的資訊來進行預測，而合併模型在結合不同模型時，便等於更全面地考量資料中所潛藏的資訊。是故，研究者往往發現某模型在一筆資料中預測表現最佳，但使用另一筆資料時該模型卻無法維持原本的良好表現 (Armstrong, 2001)。

更為重要的是，在樣本內測試中模型預測的對象即為訓練模型時所使用的資料，因此容易發生過度配適 (overfitting) 的問題，而 SVM 能在樣本內測試勝出也許就是這個原因。合併模型雖然表現未能擊敗 SVM，但由於四種單一鑑別模型都可能有過度配適的問題，結合了單一模型的合併模型是否在過度配適這個問題上反而較不嚴重呢？為了檢驗這點，我們必須比較各種單一或合併模型在樣本外的測試表現，以檢視其強健性 (robustness)。

### 5.1.2 樣本外之鑑別績效

在合併鑑別模型的樣本外測試上，我們同樣進行了三組不同的樣本測試（樣本的切分方式請見表 4）。在以下的分析中，我們將以各種樣本外測試為討論單位，並在其中分別檢視不同合併權重法則下合併鑑別模型的表現。

[16] 因此在本文往後的分析中只會在均等權重法表格中列出共識決模型之數據。

表 9 合併模型的樣本內測試結果

| | 均等權重法 | | | | | | | ARCI 加權法 | | | | | | | ARII 加權法 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 |
| (1) SVM | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 |
| (2) PCA-DA | 611 | 33 | 4 | 2 | 0.95 | 0.33 | 0 | 611 | 33 | 4 | 2 | 0.95 | 0.33 | 0 | 611 | 33 | 4 | 2 | 0.95 | 0.33 | 0 |
| (3) Logit | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 |
| (4) DT | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 |
| M12 | 611 | 1 | 0 | 2 | 1.00 | 1.00 | 36 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 |
| M13 | 601 | 1 | 0 | 25 | 1.00 | 1.00 | 23 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 |
| M14 | 612 | 1 | 0 | 21 | 1.00 | 1.00 | 16 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 |
| M23 | 598 | 10 | 1 | 2 | 0.98 | 0.67 | 39 | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 |
| M24 | 608 | 13 | 0 | 1 | 0.98 | 1.00 | 28 | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 |
| M34 | 599 | 7 | 1 | 18 | 0.99 | 0.95 | 25 | 601 | 10 | 14 | 25 | 0.98 | 0.64 | 0 | 612 | 14 | 3 | 21 | 0.98 | 0.88 | 0 |
| M123 | 614 | 10 | 1 | 25 | 0.98 | 0.96 | 0 | 614 | 10 | 1 | 25 | 0.98 | 0.96 | 0 | 615 | 1 | 0 | 34 | 1.00 | 1.00 | 0 |
| M124 | 615 | 13 | 0 | 22 | 0.98 | 1.00 | 0 | 615 | 13 | 0 | 22 | 0.98 | 1.00 | 0 | 615 | 13 | 0 | 22 | 0.98 | 1.00 | 0 |
| M134 | 614 | 7 | 1 | 28 | 0.99 | 0.97 | 0 | 614 | 7 | 1 | 28 | 0.99 | 0.97 | 0 | 614 | 7 | 1 | 28 | 0.99 | 0.97 | 0 |
| M234 | 613 | 16 | 2 | 19 | 0.97 | 0.90 | 0 | 613 | 16 | 2 | 19 | 0.97 | 0.90 | 0 | 613 | 16 | 2 | 19 | 0.97 | 0.90 | 0 |
| M1234 | 613 | 7 | 0 | 19 | 0.99 | 1.00 | 11 | 614 | 7 | 1 | 28 | 0.99 | 0.97 | 0 | 614 | 7 | 1 | 28 | 0.99 | 0.97 | 0 |
| C123 | 598 | 1 | 0 | 2 | 1.00 | 1.00 | 49 | | | | | | | | | | | | | | |
| C124 | 608 | 1 | 0 | 1 | 1.00 | 1.00 | 40 | | | | | | | | | | | | | | |
| C134 | 599 | 1 | 0 | 18 | 1.00 | 1.00 | 32 | | | | | | | | | | | | | | |
| C234 | 596 | 7 | 0 | 1 | 0.99 | 1.00 | 46 | | | | | | | | | | | | | | |
| C1234 | 596 | 1 | 0 | 1 | 1.00 | 1.00 | 52 | | | | | | | | | | | | | | |

說明：(1)共識決方法下使用 ARCI 或 ARII 加權法的結果會與均等權重法的鑑別結果完全相同，因此我們只列出共識決模型在均等權重法下的數據。(2)合併專家意見時，如果專家數目為二，則使用多數決或共識決的鑑別結果會完全相同，因此我們只列出多數決方法下的數據。

A. 2008 年總統大選

在 2008 年總統大選中，未來事件交易所總共發行了 78 個相關合約。因此在此測試中，我們先以 2008 年總統大選前的 333 個合約作為訓練樣本，在模型參數確立或模型架構決定後再對 2008 年總統大選的 78 個合約進行樣本外鑑別，鑑別結果可見於表 10。

在均等權重法的樣本外測試結果中，若以 ARCI 評斷模型之樣本外鑑別績效，則 Logit、DT、M13、M14、M23、M24、M34、M134、M234、M1234 及共識決下的五個合併預測法之 ARCI 皆為 100%，但以 M234 可以命中 73 (69+4) 個合約為最多（但其 ARII 僅 44%）。另外，在這 13 個模型中，M13、M1234 與 C134 的 ARII 為最高 100%。且 M1234 可正確預測的 69 個合約數量最多（但有 9 個無法判定的合約）。若我們比較均等權重法中採用共識決（C 開頭）及多數決（M 開頭）的合併模型，可以發現共識決的預測方法並沒有表現得比較好。

我們接著比較利用 ARCI 加權法的合併鑑別模型表現。由表中資料可以得知多個（合併）預測模型的 ARCI 皆為 100%，其中又以 M234 與 M1234 可以命中的合約數量最多（73 個）、ARII 最高 (44%)。換言之，透過 ARCI 來加權所產生的合併模型在樣本外測試的 ARCI 表現都不錯，但是這些模型的 ARII 卻明顯地普遍過低。若我們改以 ARII 來比較模型績效，則可發現以 M123 為最佳：其 ARII 可達 100%（雖然僅命中 1 個合約），且在整體總統選舉 78 個合約中可以命中 75 個合約，為全部方法可正確預測的合約最多。

若我們比較利用 ARII 加權法的合併鑑別模型表現，可以發現多個（合併）鑑別模型的 ARCI 皆可達 100%，其中以 M234 命中的合約數量最多（73 個），不過其 ARII 仍然偏低 (44%)。此外，若我們改以 ARII 作為預測效率的標準，則以 M123 為最佳（雖然僅命中 1 個合約）：整體總統選舉 78 個合約，可以命中 75 個合約，為全部方法可正確預測的合約最多。

最後，將 ARCI 加權法及 ARII 加權法的結果作一個比較後，我們發現採用不同加權法的模型在樣本外測試的 ARCI 及 ARII 上表現沒有太大差異，僅有少數幾個合併模型的表現在以 ARII 加權時有些微的改善。

B. 2009 年縣市長選舉

在 2009 年縣市長選舉中，未來事件交易所總共發行了 48 個相關合約。在此測試中，我們以過去 469 個合約（含 2008 年立委選舉）作為樣本內訓練資料，在模型參數確立或模型架構決定後再對 2009 年縣市長選舉的 48 個合約進行樣本外鑑別，鑑別結果可見於表 11。

由於預測市場在 2009 年縣市長選舉中的 48 個合約全部正確預測，大部份採用均等權重法的合併鑑別模型都將所有合約鑑別為「鑑別正確組」，且其樣本外測試的 ARCI 皆達到 100%。由於大部份的模型都沒有將合約鑑別為「鑑別未正確組」，因此無法計算其 ARII。欠缺 ARII 值的現象與單一鑑別模型在 4.2 小節的樣本外測試結果一樣，雖然表面上讓我

們無法評估模型的表現，但其實卻印證了合併模型是相當可靠的。因為事實上沒有任何一個合約屬於「事實未正確預測」組，所以理論上本來 ARII 就不可能為 100%。相反地，合併模型雖然在訓練樣本中都曾遇過「事實未正確預測」組的合約，但在面對全然陌生的測試樣本合約時卻沒有將「事實正確預測」組的合約歸類為「鑑別未正確預測」組。準此，我們可間接推論這些合併鑑別模型應該都有從訓練樣本中學習到正確知識的能力，所以除了 PCA-DA 之外的所有（合併）鑑別模型在樣本外的 ARII 都是無法計算的。

在採用 ARCI 加權法來建立合併模型時的結果與「均等權重法」下的結果接近：大部份的鑑別模型 ARCI 皆為 100%，僅少數鑑別模型無法區分全部 48 個合約（PCA-DA、M12 及表中的四個共識決鑑別模型）。在採用 ARII 加權法來建立合併模型時，僅有 PCA-DA、C123、C124、C234、C1234 無法正確區分全部 48 個合約，其餘合併模型的樣本外測試 ARCI 皆可達到 100%。

簡言之，三種權重機制在 2009 年縣市長選舉的樣本外測試中的鑑別結果十分接近，且絕大部份的鑑別模型皆具有 100% 的 ARCI。如此高準確率現象其中一個可能原因來自於鑑別模型的學習效果：在 2008 年總統大選樣本外測試中訓練樣本大小為 333 個合約，而 2009 年縣市長選舉樣本外測試時的訓練樣本大小則達到了 469 個。

### C. 2010 年五都市長選舉

在 2010 年五都選舉中，未來事件交易所總共發行了 86 個相關合約。因此在此測試中，我們以五都選舉前的 542 個合約（含 2008 年立委選舉及 2009 年縣市長選舉）作為訓練樣本，在模型參數確立或模型架構決定後再對 2010 年五都選舉的 86 個合約進行樣本外鑑別，鑑別結果可見於表 12。

藉由比較表 11 與表 12，我們發現 2010 年五都市長選舉的鑑別結果與 2009 年縣市長選舉鑑別數據雷同（僅在 ARCI 加權法中，M12 從無法完全辨視正確變成可以全部命中）。綜合 2009 年縣市長選舉與 2010 年五都市長選舉的樣本外鑑別結果，我們可以得知採用 ARII 加權法時能夠正確地鑑別全部合約的合併模型數量最多。我們推測這是由於要成功地判斷某個合約不會準確是較困難的（因此合併鑑別模型的 ARII 準確率普遍不高）。如果某一個合併鑑別模型具有較高的 ARII，代表該模型具有較佳的鑑別效率，因此在 ARCI 的表現上也會不錯。因此，以 ARII 作為加權的依據是一個較佳的作法。

最後，本文所討論的重點問題就是：到底採用單一鑑別工具就足夠，還是合併鑑別模型的表現較好？畢竟由表 9 至表 12 中我們可以看到單一鑑別工具模型如 SVM 的表現常常可與合併鑑別模型相比擬。因此，如果要對這個問題提供一個明確的答案，則我們需要對所有的樣本外測試結果作一個全面而且系統性的比較，而我們也將在 5.2 和 5.3 節中正式對這個問題提出解答。

表 10 合併鑑別模型在 2008 年總統大選之樣本外測試結果

| | 均等權重法 | | | | | | | ARCI 加權法 | | | | | | | ARII 加權法 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 |
| (1) SVM | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 |
| (2) PCA-DA | 72 | 4 | 2 | 0 | 0.95 | 0.00 | 0 | 72 | 4 | 2 | 0 | 0.95 | 0.00 | 0 | 72 | 4 | 2 | 0 | 0.95 | 0.00 | 0 |
| (3) Logit | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 |
| (4) DT | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 |
| M12 | 71 | 3 | 0 | 0 | 0.96 | - | 4 | 72 | 4 | 2 | 0 | 0.95 | 0.00 | 0 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 |
| M13 | 62 | 0 | 0 | 1 | 1.00 | 1.00 | 15 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 |
| M14 | 66 | 0 | 1 | 1 | 1.00 | 0.50 | 10 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 |
| M23 | 61 | 0 | 0 | 0 | 1.00 | - | 17 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 |
| M24 | 64 | 0 | 0 | 0 | 1.00 | - | 14 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 |
| M34 | 60 | 0 | 5 | 4 | 1.00 | 0.44 | 9 | 63 | 0 | 11 | 4 | 1.00 | 0.27 | 0 | 66 | 0 | 8 | 4 | 1.00 | 0.33 | 0 |
| M123 | 74 | 3 | 0 | 1 | 0.96 | 1.00 | 0 | 74 | 3 | 0 | 1 | 0.96 | 1.00 | 0 | 74 | 3 | 0 | 1 | 0.96 | 1.00 | 0 |
| M124 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 | 73 | 3 | 1 | 1 | 0.96 | 0.50 | 0 |
| M134 | 68 | 0 | 6 | 4 | 1.00 | 0.40 | 0 | 68 | 0 | 6 | 4 | 1.00 | 0.40 | 0 | 68 | 0 | 6 | 4 | 1.00 | 0.40 | 0 |
| M234 | 69 | 0 | 5 | 4 | 1.00 | 0.44 | 0 | 69 | 0 | 5 | 4 | 1.00 | 0.44 | 0 | 69 | 0 | 5 | 4 | 1.00 | 0.44 | 0 |
| M1234 | 68 | 0 | 0 | 1 | 1.00 | 1.00 | 9 | 69 | 0 | 5 | 4 | 1.00 | 0.44 | 0 | 68 | 0 | 6 | 4 | 1.00 | 0.40 | 0 |
| C123 | 60 | 0 | 0 | 0 | 1.00 | - | 18 | | | | | | | | | | | | | | |
| C124 | 64 | 0 | 0 | 0 | 1.00 | - | 14 | | | | | | | | | | | | | | |
| C134 | 60 | 0 | 0 | 1 | 1.00 | 1.00 | 17 | | | | | | | | | | | | | | |
| C234 | 58 | 0 | 0 | 0 | 1.00 | - | 20 | | | | | | | | | | | | | | |
| C1234 | 58 | 0 | 0 | 0 | 1.00 | - | 20 | | | | | | | | | | | | | | |

說明：同表 9 之說明。

表 11 合併鑑別模型在 2009 年縣市長選舉之樣本外測試結果

| | 均等權重法 | | | | | | | ARCI 加權法 | | | | | | | ARII 加權法 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 |
| (1) SVM | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| (2) PCA-DA | 45 | 0 | 3 | 0 | 1.00 | 0.00 | 0 | 45 | 0 | 3 | 0 | 1.00 | 0.00 | 0 | 45 | 0 | 3 | 0 | 1.00 | 0.00 | 0 |
| (3) Logit | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| (4) DT | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M12 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | 45 | 0 | 3 | 0 | 1.00 | 0.00 | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M13 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M14 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M23 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M24 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M34 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M123 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M124 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M134 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M234 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| M1234 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | 48 | 0 | 0 | 0 | 1.00 | - | 0 |
| C123 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | | | | | | | | | | | | | | |
| C124 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | | | | | | | | | | | | | | |
| C134 | 48 | 0 | 0 | 0 | 1.00 | - | 0 | | | | | | | | | | | | | | |
| C234 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | | | | | | | | | | | | | | |
| C1234 | 45 | 0 | 0 | 0 | 1.00 | - | 3 | | | | | | | | | | | | | | |

說明：同表 9 之說明同。

表 12 合併鑑別模型在 2010 年五都市長選舉之樣本外測試結果

| | 均等權重法 | | | | | | | ARCI 加權法 | | | | | | | ARII 加權法 | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 | TP | FP | FN | TN | ARCI | ARII | 無共識 |
| (1) SVM | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| (2) PCA-DA | 85 | 0 | 1 | 0 | 1.00 | 0.00 | 0 | 85 | 0 | 1 | 0 | 1.00 | 0.00 | 0 | 85 | 0 | 1 | 0 | 1.00 | 0.00 | 0 |
| (3) Logit | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| (4) DT | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M12 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M13 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M14 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M23 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M24 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M34 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M123 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M124 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M134 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M234 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| M1234 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | 86 | 0 | 0 | 0 | 1.00 | - | 0 |
| C123 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | | | | | | | | | | | | | | |
| C124 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | | | | | | | | | | | | | | |
| C134 | 86 | 0 | 0 | 0 | 1.00 | - | 0 | | | | | | | | | | | | | | |
| C234 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | | | | | | | | | | | | | | |
| C1234 | 85 | 0 | 0 | 0 | 1.00 | - | 1 | | | | | | | | | | | | | | |

說明：同表 9 之說明。

## 5.2 鑑別模型之績效比較

鑑別分析的最後一步驟，便是提出一套明確的評量標準來比較各模型的良窳，我們必須要有一套方法可同時且全面地比較各樣本下模型的表現。舉例而言，單一鑑別模型 SVM 在 650 個樣本的樣本內訓練表現可與合併鑑別模型相比擬（見 5.1.1 節表 9），而且在 2009 年及 2010 年選舉中的樣本外表現也和合併模型一樣好，只有在 2008 年選舉的樣本外測試中略遜於少數合併模型（見 5.1.2 節表 10 至表 12），這是否就代表 SVM 表現不如合併模型呢？

為了回答這個問題，在本節中我們不但將各模型在 2008、2009、2010 三個樣本外測試的結果以同樣的標準進行評估，並且又引進另一個樣本外測試—以 2008 年總統選舉（含）之後所有合約作為測試樣本（即表 4 中之 D 組樣本切分方式）—共十二個樣本外測試績效值，以期能更全面地比較單一與合併模型的鑑別績效。[17] 圖 3 為各種鑑別模型在十二個樣本外測試結果的 ARCI 盒鬚圖，圖 4 則為 ARII 表現的分配情況。[18]

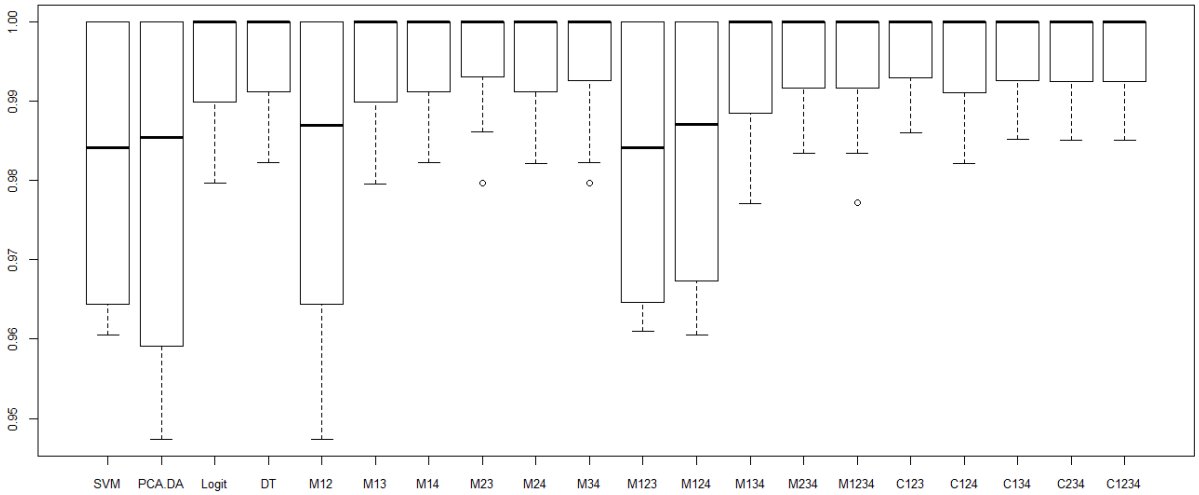


圖 3 各模型在十二個樣本外測試的 ARCI 分佈圖

由圖 3 可以觀察到不論是單一或合併鑑別模型在 ARCI 上雖然都有相當不錯的表現，但在變異程度上卻有所不同。因此若我們應該可以透過一套合理的評估方法來區分出各模型之優劣。另一方面我們可以由圖 4 輕易觀察到，若以 ARII 作為評估鑑別模型的標準，各鑑別模型之間其實存在相當的差異。若我們希望同時考量 ARCI 與 ARII，到底哪一個鑑別模型才是最好的呢？

[17] 針對四個樣本外測試，我們皆採取了三種加權方法進行嘗試，因此每個鑑別模型都會有十二個預測績效值。
[18] 在圖 3 及圖 4 中盒子的上下緣各代表第三及第一四分位數，盒中實線為第二四分位數（即中位數）；圖中的虛線延伸至 1.5 倍四分差 (interquartile range) 範圍內最遠的資料點。在圖 4 中，若某一合併模型在所有的樣本外測試中都沒有將任何一個合約鑑定為未正確，其 ARII 便無法計算，也就無法繪製其 ARII 盒鬚圖。

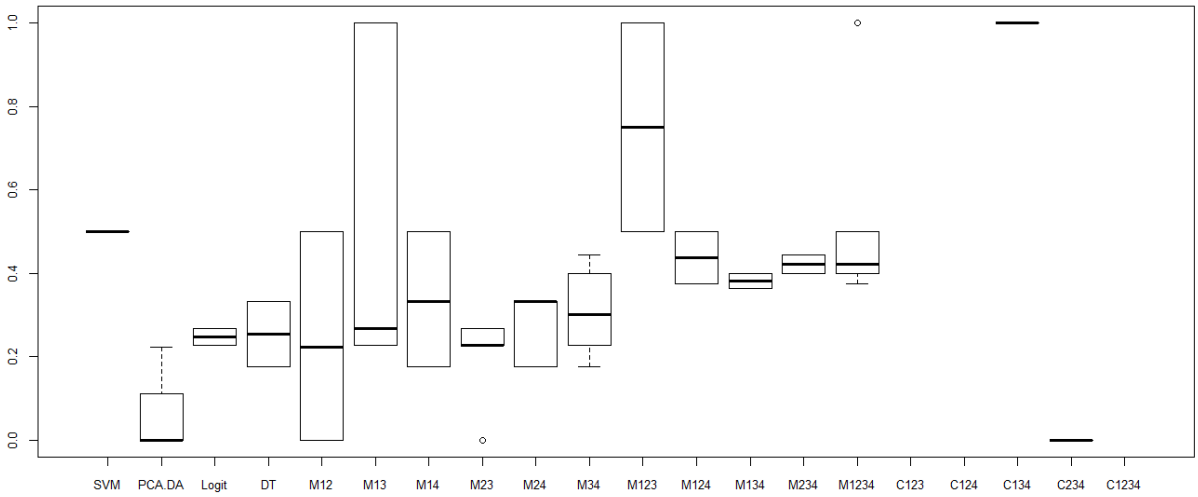


圖 4 各模型在十二個樣本外測試的 ARII 分佈圖

為了全面且明確地評估鑑別模型之績效，本文了提出三種挑選準則：正確性、穩定性、以及辨別度。「正確性」指的是一個合併鑑別模型應該要具有卓越的判別準確率(以 ARCI 或 ARII 衡量之，值愈高愈好)，才能幫助研究者在事前成功判定預測市場的預測是否會實現；「穩定性」指的是合併預測模型的判別準確率不應有過大的起伏（以 ARCI 或 ARII 的變異程度衡量之，變異程度愈低愈好），才能降低判斷失準的風險；「辨別度」指的是一個成功的合併鑑別模型應該要盡可能地對每一個預測市場事件進行判斷，而不是保守地只有在情況非常明顯時才能作出判讀（以未鑑別合約之比例衡量之，比例愈低愈好），如此才能真正發揮事前解讀市場預測的功能。

本文首先計算十二項樣本外測試鑑別準確率之平均值及變異程度（標準差），並予以排序以便比較鑑別模型之良窳。[19] 此外，我們也統計各鑑別模型的平均「未鑑別比例」作為第三個參考的準則。[20] 各合併模型的統計值及其排名結果如表 13 所示。

在確立了正確性、穩定性、及辨別度三種基本參考指標後，要如何同時參考這三種準則來選出最佳合併鑑別模型呢？本文建議在不同應用情況下，決策者可先依實際需求定義目標函數，將這三個指標的值納入函數計算中。目標函數有可能為正向的「效用函數」(utility function)、負向的「損失函數」(loss function)、或兩者的混和。接著便可透過追求目標函數的極值（如效用值的極大化或損失值的極小化），以便找出最適的合併鑑別模型。

舉例而言，若研究者希望能在事前儘量正確地挑出預測正確的合約，則其目標函數便可納入「正確性」，並在評估的過程中找出「正確性」最高的合併鑑別模型，再利用該模型來對市場合約的預測準確與否進行事前鑑別。在另一種可能的情況中，若研究者認為錯誤的判

[19] 無法計算 ARII 的空值部份，在此處計算時假設為 1。
[20] 未判別比例指的是該合併模型無法提出判斷的次數除以該模型所經手的預測市場合約數目。

斷將付出高昂的代價，則研究者在事前挑選的過程中便有可能會將合併鑑別模型的鑑別「穩定性」作為主要的考量。例如在總統大選這樣重大的事件中，研究者所選取的事前鑑別模型若在先前的樣本測試中表現時好時壞，便有可能剛好在這次的總統大選中作出錯誤的鑑別。在此情況下，錯誤的鑑別可能會為研究者甚至社會帶來極大的成本，此時就算該合併模型的鑑別正確性再高也抵不過一次鑑別失敗所帶來的損失。換言之，一個鑑別模型的表現愈不穩定，研究者所面對的風險就愈大，此時研究者的目標函數便應該設定為負向的損失函數，並追求預期損失的極小化。

在另一個類似的情況中，針對一些重要且呈現拉鋸的選舉事件上，正是人們需要預測市場的時候。此時若所採用的合併鑑別模型正好無法給出判斷（如四種工具的意見正好平手或未達共識），則研究者將無法在事前判斷預測市場的準確性，在此情況下即使該合併模型的正確性再高也無濟於事。此時研究者便可考慮將合併模型的未鑑別比例納入目標函數中，以盡量減少研究者對預測市場準確性「不置可否」的情況。

事實上，挑選最佳鑑別模型的準則應該取決於研究者在每一次事前判斷工作中的動機。研究者在不同情況下可選擇不同的準則，而每一種挑選準則或準則間的優先排序都有可能會挑出不同的合併鑑別模型。本文建議四種可能的決策目標層次，考量的因素依序由最簡單的正確性到最複雜的多目標決策，以作為挑選出最適合併鑑別模型之準則。

### *目標一：僅考量正確性*

若在挑選鑑別模型時「正確性」是唯一的考量，換句話說若決策者是風險中立的，那我們可將目標函數簡單設計為一個僅考量正向效用的函數如下：

$$\text{效用} = \text{平均準確率} \tag{14}$$

在實際的計算上，我們可將各種鑑別模型的 ARCI 與 ARII 之平均值直接相加便可得到最後的目標效用值。接著再比較目標值的大小後所得到排名第一的合併鑑別模型，是由 SVM、PCA-DA、與 Logit 三個模型透過共識決的方式合併組成的鑑別模型 C123（見表 14）。

### *目標二：考量正確性與穩定性*

若我們同時強調「正確性」與「穩定性」，那我們將目標函數設為類似資產選擇理論中的「均值—變異偏好」(mean-variance preferences) 的型式：[21]

$$\text{效用} = \text{平均準確率} - \text{準確率標準差} \tag{15}$$

[21] Mean-variance preference 是由 Markowitz (1952) 及 Tobin (1958) 所提出，目前已是資產組合選擇理論中用來描述決策者偏好的常見作法，例如 Kandel and Stambaugh (1987), Britten-Jones (1999), Bodie, Kane, and Marcus (2002), 及 Maccheroni et al. (2009)。在本文中由於標準差均小於 1，故未採用變異數以免變異數過小而看不出結果。

表 13 合併鑑別模型之績效排名

| | 平均值 | | | | 標準差 | | | | 平均排名 | 未鑑別比例 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | ARCI | 排名 | ARII | 排名 | ARCI | 排名 | ARII | 排名 | | 平均數 | 標準差 |
| C123 | 0.99650 | 1 | 1.00000 | 1 | 0.00633 | 1 | 0.00000 | 1 | 1.0 | 0.10067 | 0.08112 |
| C134 | 0.99630 | 2 | 1.00000 | 1 | 0.00669 | 2 | 0.00000 | 1 | 1.5 | 0.08998 | 0.09390 |
| C1234 | 0.99628 | 3 | 1.00000 | 1 | 0.00674 | 3 | 0.00000 | 1 | 2.0 | 0.12128 | 0.09334 |
| C124 | 0.99553 | 10 | 1.00000 | 1 | 0.00810 | 10 | 0.00000 | 1 | 5.5 | 0.09258 | 0.06244 |
| M234 | 0.99585 | 5 | 0.71110 | 11 | 0.00751 | 5 | 0.30219 | 10 | 7.8 | 0.00000 | 0.00000 |
| C234 | 0.99628 | 3 | 0.75000 | 7 | 0.00674 | 3 | 0.45227 | 20 | 8.3 | 0.12049 | 0.09307 |
| M1234 | 0.99533 | 12 | 0.75995 | 6 | 0.00858 | 12 | 0.29815 | 9 | 9.8 | 0.01330 | 0.03309 |
| M14 | 0.99558 | 7 | 0.66830 | 14 | 0.00801 | 6 | 0.35992 | 12 | 9.8 | 0.01910 | 0.04306 |
| M24 | 0.99556 | 9 | 0.69608 | 12 | 0.00804 | 8 | 0.37915 | 15 | 11.0 | 0.02928 | 0.05441 |
| M123 | 0.98233 | 17 | 0.87500 | 5 | 0.01866 | 17 | 0.22613 | 6 | 11.3 | 0.00000 | 0.00000 |
| M34 | 0.99560 | 6 | 0.65402 | 17 | 0.00805 | 9 | 0.36800 | 13 | 11.3 | 0.01908 | 0.04266 |
| (4) DT | 0.99558 | 7 | 0.62745 | 18 | 0.00801 | 6 | 0.39340 | 16 | 11.8 | 0.00000 | 0.00000 |
| (1) SVM | 0.98220 | 18 | 0.75000 | 7 | 0.01881 | 18 | 0.26112 | 7 | 12.5 | 0.00000 | 0.00000 |
| M124 | 0.98365 | 16 | 0.71875 | 10 | 0.01780 | 16 | 0.29736 | 8 | 12.5 | 0.00000 | 0.00000 |
| M13 | 0.99492 | 14 | 0.74900 | 9 | 0.00920 | 14 | 0.37094 | 14 | 12.8 | 0.02181 | 0.05484 |
| M134 | 0.99428 | 15 | 0.69090 | 13 | 0.01036 | 15 | 0.32312 | 11 | 13.5 | 0.00000 | 0.00000 |
| M23 | 0.99546 | 11 | 0.66567 | 15 | 0.00837 | 11 | 0.41868 | 18 | 13.8 | 0.03196 | 0.06302 |
| (3) Logit | 0.99493 | 13 | 0.62350 | 19 | 0.00918 | 13 | 0.39351 | 17 | 15.5 | 0.00000 | 0.00000 |
| (2) PCA-DA | 0.97955 | 20 | 0.05555 | 20 | 0.02304 | 20 | 0.10049 | 5 | 16.3 | 0.00000 | 0.00000 |
| M12 | 0.98170 | 19 | 0.65656 | 16 | 0.02020 | 19 | 0.42515 | 19 | 18.3 | 0.01334 | 0.02187 |

在實際計算上，我們分別將各鑑別模型之 ARCI 與 ARII 平均數相加後，再減去該模型在 ARCI 與 ARII 上的標準差，便可得到最後的目標效用值。接著我們比較目標效用值的大小，發現 C123 是最佳的合併鑑別模型（見表 14）。從表中我們可以得知模型 C123 在 ARCI 與 ARII 具有較高的平均準確率，且其 ARCI 與 ARII 在四個樣本外測試的變異數比其它 17 個鑑別模型低，因此是同時考量「正確性」與「穩定性」下最佳的合併鑑別模型。

*目標三：考量正確性、穩定性、及辨別度*

若除了「正確性」與「穩定性」外，我們也希望模型的「辨別度」不要太差，則我們將目標函數設為：

$$\text{效用} = (\text{平均準確率} - \text{準確率標準差}) - \text{未鑑別比例} \tag{16}$$

在實際的計算上，我們先分別將各鑑別模型之 ARCI 與 ARII 平均數相加，然後減去該模型在 ARCI 與 ARII 上的標準差，最後再減去該模型的未鑑別比例，便可得到最後的目標效用值。經過比較後，我們發現模型 C134 可使效用達到最大，故為最佳模型（見表 14）。C134 的正確性僅次於 C123，但它具備更高的辨別度，代表它可鑑別較多的合約。

*目標四：優先考量辨別度的多目標問題*

在目標三中，我們雖然同時考量了「正確性」、「穩定性」、與「辨別度」，但 C134 仍有接近一成的「未鑑別比例」，代表此模型有時會無法提出鑑別。因此若我們改以「辨別度」作為絕對優先考量，接著才考慮模型的「正確性」，並假設偏好順序為「辨別度 >> ARCI >> ARII」，則此目標便是經濟學中所謂的辭典偏好 (lexicographic preference) 或決策理論中的辭典策略 (lexicographic strategy) (Payne, Bettman, and Johnson, 1993)。在實際計算上，我們先挑出「未鑑別比例」最低的所有模型，第二步則從這些模型中挑出 ARCI 最高者；若有超過一個以上的模型滿足前述條件，最後再比較 ARII 的大小。經比較後，我們發現 M234 為此目標下的最佳合併鑑別模型（見表 14）。M234 的鑑別比例為 100%，代表它對每一個預測市場事件都能提出判斷，而且平均鑑別正確準確率與 C123、C134 同樣大於 99%以上。

本文所討論的一個重點問題就是：到底採用單一鑑別工具就足夠，還是合併鑑別模型的表現較好？藉由本節的分析，我們明顯看到不論目標函數的設計為何，最後所挑選出來的最適鑑別模型前幾名幾乎都由合併模型所包辦。因此，吾人可以得知在適當的合併方法下，合併鑑別模型的表現會比單一鑑別模型更好。

表 14 各種目標函數下合併鑑別模型之綜合表現

| | 平均值 | | 標準差 | | 未鑑別比例 | 目標一 | 目標二 | 目標三 | 目標四 |
|---|---|---|---|---|---|---|---|---|---|
| 模型 | ARCI | ARII | ARCI | ARII | | 效用值 | 效用值 | 效用值 | 辭典偏好順位 |
| C123 | 0.9965 | 1.0000 | 0.0063 | 0.0000 | 0.1007 | **1.9965** | **1.9902** | 1.8895 | 18 |
| C134 | 0.9963 | 1.0000 | 0.0067 | 0.0000 | 0.0900 | 1.9963 | 1.9896 | **1.8996** | 16 |
| C1234 | 0.9963 | 1.0000 | 0.0067 | 0.0000 | 0.1213 | 1.9963 | 1.9895 | 1.8683 | 20 |
| C124 | 0.9955 | 1.0000 | 0.0081 | 0.0000 | 0.0926 | 1.9955 | 1.9874 | 1.8948 | 17 |
| M234 | 0.9959 | 0.7111 | 0.0075 | 0.3022 | 0.0000 | 1.7070 | 1.3973 | 1.3973 | **1** |
| C234 | 0.9963 | 0.7500 | 0.0067 | 0.4523 | 0.1205 | 1.7463 | 1.2873 | 1.1668 | 19 |
| M1234 | 0.9953 | 0.7600 | 0.0086 | 0.2982 | 0.0133 | 1.7553 | 1.4486 | 1.4353 | 9 |
| M14 | 0.9956 | 0.6683 | 0.0080 | 0.3599 | 0.0191 | 1.6639 | 1.2959 | 1.2769 | 12 |
| M24 | 0.9956 | 0.6961 | 0.0080 | 0.3792 | 0.0293 | 1.6916 | 1.3044 | 1.2752 | 14 |
| M123 | 0.9823 | 0.8750 | 0.0187 | 0.2261 | 0.0000 | 1.8573 | 1.6125 | 1.6125 | 6 |
| M34 | 0.9956 | 0.6540 | 0.0081 | 0.3680 | 0.0191 | 1.6496 | 1.2736 | 1.2545 | 11 |
| (4) DT | 0.9956 | 0.6275 | 0.0080 | 0.3934 | 0.0000 | 1.6230 | 1.2216 | 1.2216 | 2 |
| (1) SVM | 0.9822 | 0.7500 | 0.0188 | 0.2611 | 0.0000 | 1.7322 | 1.4523 | 1.4523 | 7 |
| M124 | 0.9837 | 0.7188 | 0.0178 | 0.2974 | 0.0000 | 1.7024 | 1.3872 | 1.3872 | 5 |
| M13 | 0.9949 | 0.7490 | 0.0092 | 0.3709 | 0.0218 | 1.7439 | 1.3638 | 1.3420 | 13 |
| M134 | 0.9943 | 0.6909 | 0.0104 | 0.3231 | 0.0000 | 1.6852 | 1.3517 | 1.3517 | 4 |
| M23 | 0.9955 | 0.6657 | 0.0084 | 0.4187 | 0.0320 | 1.6611 | 1.2341 | 1.2021 | 15 |
| (3) Logit | 0.9949 | 0.6235 | 0.0092 | 0.3935 | 0.0000 | 1.6184 | 1.2157 | 1.2157 | 3 |
| (2) PCA-DA | 0.9796 | 0.0556 | 0.0230 | 0.1005 | 0.0000 | 1.0351 | 0.9116 | 0.9116 | 8 |
| M12 | 0.9817 | 0.6566 | 0.0202 | 0.4252 | 0.0133 | 1.6383 | 1.1929 | 1.1796 | 10 |

## 5.3 合併鑑別模型之規模與權重法之比較

本文提出合併鑑別模型來整合各個單一模型之特點，理論上合併鑑別模型的優勢，在於每個單一模型皆透過不同的方式來自資料中獲取資訊，因此合併模型便可以較全面地捕捉潛藏在資料中所有可能的規律。按此邏輯，合併的模型數量越多所捕捉到的規律理應越完整，那麼合併鑑別模型的表現是否真的比單一模型好呢？合併模型的鑑別績效是否與其規模呈正相關？圖 5 整理了規模由小到大的合併鑑別模型平均鑑別績效。由最小的規模 1（即四個單一鑑別模型）開始到最大的 4（即把四個單一模型全數納入），我們可以觀察到合併模型的鑑別正確率似乎會隨著規模加大而提升。[22]

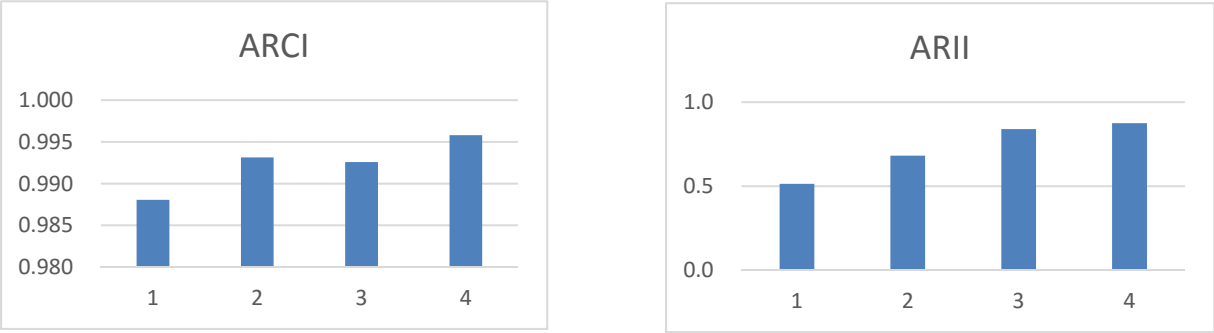


圖 5 合併鑑別模型規模與鑑別準確度

為了更嚴謹地檢驗各模型規模在鑑別績效上的差異是否顯著，我們首先檢視不同規模合併模型績效的分配是否為常態。表 15 為各規模下的模型鑑別績效進行 Shapiro-Wilk 檢定的結果，由表中可知所有績效結果（ARCI 與 ARII）皆非常態分配。在此前提下，我們對不同規模的鑑別模型績效進行無母數統計檢定。

表 16 為各規模之鑑別模型績效兩兩進行 Wilcoxon 等級和檢定 (Wilcoxon rank sum test) 之結果。先就 ARCI 而言，由表中可以得知規模為四的合併模型在 10% 的顯著水準下優於單一鑑別模型。若以 ARII 作為比較標準，則當合併的模型數量達三種以上時，其鑑別績效皆顯著地比單一模型或僅合併兩種鑑別工具的模型為佳。簡言之，在納入足夠的模型後，合併鑑別模型的結果就愈正確，特別是在「事實未正確準組」的事前篩選上，採用多種工具進行合併鑑別將可有效提升準確性。[23] 但在另一方面，無法鑑別的合約數量卻也會隨著合併模型的規模提升而增加。具體而言，由單一模型到最大由四個模型組成的合併模型，其無法鑑別 (un-consensus)合約的平均數量分別為 0、3.36、

[22] 模型規模為 1 代表四種單一鑑別模型；規模為 2 代表合併了兩種鑑別工具的模型，例如 M12 及 M24；規模為 3 代表合併了三種鑑別工具的模型，如 C123 及 M234；規模為 4 則為 C1234 及 M1234。
[23] 整體而言，若對各種模型規模同時進行無母數 Kruskal-Wallis 檢定，則 ARCI 的結果為 0.2975（p 值），而 ARII 的結果為 3.90e-06，意謂不同規模的合併模型在 ARII 上的績效皆顯著不同。

7.56、及 10.08。因此若要尋找一個辨別度高且兼具高準確性的鑑別模型，仍建議採用 5.2 節所介紹的作法。

表 15 各種規模之合併模型鑑別績效 Shapiro-Wilk 檢定結果

| | ARCI | | ARII | |
|---|---|---|---|---|
| 合併模型規模 | W statistic | *p*-value | W statistic | *p*-value |
| 1 | 0.7132 | 2.29E-08 | 0.8682 | 0.0015 |
| 2 | 0.6321 | 3.81e-12 | 0.8150 | 6.32e-05 |
| 3 | 0.6501 | 8.56e-14 | 0.8160 | 6.63e-05 |
| 4 | 0.5866 | 4.4e-07 | 0.6573 | 0.0021 |

表 16 各種規模下合併模型鑑別績效之 Wilcoxon 等級和檢定結果

| ARCI | | | | |
|---|---|---|---|---|
| 模型規模 | 1 | 2 | 3 | 4 |
| 1 | | | | |
| 2 | 0.1509 | | | |
| 3 | 0.1881 | 0.8331 | | |
| 4 | **0.0887** | 0.4327 | 0.4043 | |

| ARII | | | | |
|---|---|---|---|---|
| 模型規模 | 1 | 2 | 3 | 4 |
| 1 | | | | |
| 2 | 0.1099 | | | |
| 3 | **1.67e-05** | **8.75e-05** | | |
| 4 | **0.0072** | **0.0122** | 0.9684 | |

表格中數字為 *p*-value

本文在提出合併鑑別模型時曾參考合併預測的文獻提出了三種加權合併的方法：均等權重法、ARCI 加權法、以及 ARII 加權法，這三種加權法何者能產生較好的鑑別績效呢？表 17 整理了透過三種加權法以及採用共識決所得到的合併鑑別模型之鑑別正確率及未鑑別合約數。由表中可以觀察到使用簡單的均等權重法或共識決來合併模型效果似乎會比依賴過去表現來加權合併還要好。經 Wilcoxon 等級和檢定後我們發現各種合併法在 ARCI 的表現上並無顯著差異，但在 ARII 的表現上均等權重法顯著優於 ARCI 及 ARII 加權法（*p*-value 分別為 0.002169 和 0.01351），而共識決合併法則顯著優於 ARCI 加權法（*p*-value 為 0.04373）。由以上分析可知平均而言，均等權重法以及共識決皆為較準確的合併方法；若考量無法鑑別的合約多寡，則使用均等權重法時無法鑑別的合約數量顯著低於共識決下的數量（*p*-value 為 0.001791）。簡言之，使用均等權

重法並採多數決來合併模型平均而言是較佳的合併方式。

表 17 各種加權方法之鑑別準確度與未鑑別比例

| 合併方法 | | ARCI | ARII | 未鑑別合約數 |
|---|---|---|---|---|
| 多數決 | 均等權重法 | 0.9922 | 0.5367 | 6.0227 |
| | ARCI 加權法 | 0.9914 | 0.3283 | 0 |
| | ARII 加權法 | 0.9919 | 0.3762 | 0 |
| 共識決 | | 0.9948 | 0.5616 | 12.6591 |

# 6 結論

為了對預測市場中的合約預測於事前提出正確性的評估，目前文獻上大部份的作法是以預測市場的歷史準確率作為決策時的參考。然而，歷史準確度只能反應過去預測市場的整體表現，只適合用來評估「預測市場」這個方法的可靠性。以歷史準確率作為決策的參考忽略了每個市場合約間的差異，也完全忽略了市場交易的動態內容。因此，在預測市場作為一個具有實務價值方法的發展過程中，便必須克服這個問題，而非只是在事件結果揭曉後作事後的準確率計算。

本文有別於現今預測市場研究之處，便是提出一個基於市場交易特徵的**合併鑑別法**來評估市場的預測準確度。合併鑑別模型整合了迴歸模型、多變量分析、決策樹、以及支持向量機等四種模型，全面性地探索隱藏在高維度資料中的結構以擷取與市場預測正確率有關的潛在資訊。此合併鑑別法不但能夠在事件發生之前就對任一合約預測正確與否提出判斷，更可以隨著交易過程動態調整其鑑別結果。而採用合併鑑別方法的好處，在於此方法能依照決策者對「正確性」、「穩定性」、及「辨別度」三種不同層次的目標進行調整與挑選，以極大化決策者之目標函數。

更具體而言，本文中我們試圖探討以下三個問題：

1. 對預測市場的預測正確性事前提出鑑別的可行性有多高？
2. 合併鑑別模型是否表現的比單一鑑別模型來得好？
3. 最適的鑑別模型為何？

本文利用未來事件交易所自 2006 至 2011 年共 650 個選舉預測合約，並以「2008 年總統大選」、「2009 年縣市長選舉」、及「2010 年五都選舉」三個選舉事件中的合約對各種鑑別模型進行樣本外測試。結果發現不論是單一模型或合併模型都有能力在事前對預測市場的預測是否會準確提出準確率相當不錯的判別。

本文嘗試的合併方法計有不同規模（從僅使用單一鑑別工具到四種鑑別工具全部使用）及不同加權方式（均等權重法、ARCI 加權法、ARII 加權法）。在比較過各種合併鑑別模型後，本文發現將四種鑑別工具的預測進行組合可以顯著地提升鑑別的準確率（特別是在 ARII 的表現上），但同時無法鑑別合約的情況卻也會隨著合併模型的規模增大而增加。在加權方法上，本文發現簡單的均等權重法是較佳的合併方法。

就特定的鑑別模型而言，本文發現沒有任何一個合併鑑別模型全面性地優於其他模型。研究者應該要依其動機設立不同的目標函數，因為在不同的挑選準則下，所謂最適鑑別模型可能也會不盡相同。本文也進一步提出了以「正確性」、「穩定性」、「辨別度」三個基礎準則所建構的挑選流程，並在不同的目標函數下找出三個合併鑑別模型以滿足不同的研究需求。而挑選出的最適合併鑑別模型分別為：

1. 在只考量「正確性」的目標下，模型 C123 為最佳鑑別模型。
2. 在同時考量「正確性」及「穩定性」的目標下，模型 C123 仍為最佳鑑別模型。
3. 在同時考量「正確性」、「穩定性」、與「辨別度」的目標下，模型 C134 為最佳模型。
4. 若挑選準則的絕對優先順序為「辨別度 > ARCI > ARII」，模型 M234 為最佳鑑別模型。

最後，本文所使用的研究方法有一些可以進行延伸研究的地方如下：

1. 本文使用了相當多變數來鑑別選舉合約的預測正確性，那麼到底哪些變數在鑑別過程中被實際使用了、他們的邊際貢獻又是如何呢？就本文而言，每一次樣本測試中各個鑑別工具都會學習到不同的模型，而變數的邊際貢獻則依附於所學到的模型結構。在此情況下無法直接觀察變數的邊際貢獻，而需要思考以下基本問題：

   (1) 定義邊際貢獻的意義，並釐清其與模型結構間的關係：有些模型可看到變數的重要性，如 Logit 模型中的係數；有些模型可間接觀察，如 PCA-DA 中的主成份因素負載量與判別函數係數，以及決策樹中的變數使用順序及變數重要性；有些模型則是完全的黑箱，如 SVM。不同的模型之間要如何用一套統一的標準來衡量變數的邊際貢獻是相當具挑戰性的工作。

   (2) 發展一套系統化的方法由多個樣本測試中統整出邊際貢獻值：不同的樣本測試可能會訓練出不同的模型或學到不同的係數，要如何由一系列的樣本測試結果中歸納出一般化的變數邊際貢獻，要使用簡單的加權平均？還是依樣本測試鑑別績效作加權？

   (3) 對於未直接使用原始變數的鑑別工具 (PCA-DA, SVM) 測量其變數貢獻：另一種衡量變數邊際貢獻的作法是仿傚逐步回歸 (stepwise regression) 的

作法，比較某一變數加入前及加入後模型整體績效的變化。使用此法的優點是可以在完全不同的鑑別模型間進行簡單的比較，缺點是無法掌握變數間的非線性關係。

2. 「預測市場對某選舉事件的預測是否準確」本身也是一個可預測的事件，在本研究中是利用結合四種分類模型來達到事先鑑別的目的。由於預測市場本身便是一種有效的事前預測工具，且可用在二元預測 (binary prediction) 上。若可以延攬對預測市場運作具備相關知識或經驗的人士參與市場交易，也許將本文的研究對象—即「選舉合約的市場價格預測是否可信」—交由預測市場本身進行預測也是一個未來可能的研究方向。

3. 本文所使用的目標函數皆源自於既有且廣為人知的重要函數或偏好型式，至於哪種目標函數才是最佳的端視決策者的研究或實務目標而定，研究者可以想像有許多實務上常用的指標可以作為挑選時的目標函數。此外，不同的預測問題（如體育賽事或其他性質的預測市場合約）也許適合不同的目標函數，有待後續研究採用更多元、更大量的資料方能下定論。

## 致謝

感謝國科會專題研究計畫編號 NSC 101-2410-H-029-010-MY2、NSC 99-2628-H-004-001-MY3 與編號 NSC 98-2410-H-032-044-MY3 之經費支持，使本研究得以順利完成。

## 附錄一

在一般的資料探勘研究中常用的績效指標有 precision、recall (sensitivity)、及 f1-score，而本文採用的主要指標為 ARCI (precision) 及 ARII (NPV)。未採用 recall 及 f1-score 的原因在於預測市場一直以來皆有相當高的正確率，因此真正困難的問題在於如何事前挑出少數預測失準的合約。就本研究所使用的 650 個合約而言，其中 94.61%（615 個合約）的市場預測最後證明是正確的。在此情況下，若某一個鑑別模型完全無鑑別能力而把 650 個合約直接全歸為正確，其 sensitivity/recall 也會高達 94.61%。為了驗證這點，我們想知道若在最後挑選最佳鑑別模型時改用 recall 及 f1-score，會不會得到不同的結果呢？我們分別使用 recall 及 f1-score 代入目標函數一及目標函數二中，結果呈報於表 18。接著我們再由所有合併模型中挑選出分數最高者為最佳模型。雖然許多模型改以 recall 及 f1-score 計算後排名有所變動，但分數最高的模型仍然為 C123，與採用 ARCI 進行分析的結論並無二致。

表 18 比較採用 ARII, recall, 與 f1-score 評估之目標函數值

| | 平均值 | | | | 標準差 | | | | ARII | | | | Recall | | | | F1 score | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 模型 | ARCI | ARII | Recall | f1-score | ARCI | ARII | Recall | f1-score | U1 | U2 | R1 | R2 | U1 | U2 | R1 | R2 | U1 | U2 | R1 | R2 |
| (1) SVM | 0.98 | 0.75 | 1.00 | 0.99 | 0.02 | 0.26 | 0.01 | 0.01 | 1.73 | 1.45 | 9 | 6 | 1.98 | 1.95 | 8 | 7 | 1.97 | 1.94 | 16 | 13 |
| (2) PCA-DA | 0.98 | 0.06 | 0.97 | 0.97 | 0.02 | 0.10 | 0.02 | 0.01 | 1.04 | 0.91 | 20 | 20 | 1.95 | 1.91 | 18 | 13 | 1.95 | 1.92 | 20 | 20 |
| (3) Logit | 0.99 | 0.62 | 0.95 | 0.97 | 0.01 | 0.39 | 0.06 | 0.03 | 1.62 | 1.22 | 19 | 18 | 1.94 | 1.87 | 20 | 20 | 1.97 | 1.92 | 19 | 19 |
| (4) DT | 1.00 | 0.63 | 0.95 | 0.97 | 0.01 | 0.39 | 0.05 | 0.03 | 1.62 | 1.22 | 18 | 17 | 1.95 | 1.89 | 19 | 19 | 1.97 | 1.93 | 17 | 18 |
| M12 | 0.98 | 0.66 | 0.99 | 0.99 | 0.02 | 0.43 | 0.02 | 0.01 | 1.64 | 1.19 | 17 | 19 | 1.97 | 1.93 | 11 | 11 | 1.97 | 1.93 | 18 | 17 |
| M13 | 0.99 | 0.75 | 0.97 | 0.98 | 0.01 | 0.37 | 0.06 | 0.03 | 1.74 | 1.36 | 8 | 10 | 1.96 | 1.90 | 15 | 17 | 1.97 | 1.94 | 12 | 16 |
| M14 | 1.00 | 0.67 | 0.97 | 0.98 | 0.01 | 0.36 | 0.05 | 0.03 | 1.66 | 1.30 | 14 | 13 | 1.96 | 1.91 | 16 | 15 | 1.98 | 1.94 | 10 | 11 |
| M23 | 1.00 | 0.67 | 0.97 | 0.98 | 0.01 | 0.42 | 0.05 | 0.03 | 1.66 | 1.23 | 15 | 16 | 1.96 | 1.90 | 13 | 16 | 1.97 | 1.94 | 11 | 14 |
| M24 | 1.00 | 0.70 | 0.97 | 0.98 | 0.01 | 0.38 | 0.05 | 0.03 | 1.69 | 1.30 | 12 | 12 | 1.96 | 1.91 | 14 | 14 | 1.98 | 1.94 | 9 | 12 |
| M34 | 1.00 | 0.65 | 0.96 | 0.98 | 0.01 | 0.37 | 0.05 | 0.03 | 1.65 | 1.27 | 16 | 15 | 1.95 | 1.90 | 17 | 18 | 1.97 | 1.94 | 15 | 15 |
| M123 | 0.98 | 0.88 | 1.00 | 0.99 | 0.02 | 0.23 | 0.00 | 0.01 | 1.86 | 1.61 | 5 | 5 | 1.98 | 1.96 | 6 | 6 | 1.97 | 1.94 | 13 | 9 |
| M124 | 0.98 | 0.72 | 0.99 | 0.99 | 0.02 | 0.30 | 0.01 | 0.01 | 1.70 | 1.39 | 11 | 9 | 1.98 | 1.95 | 9 | 8 | 1.97 | 1.94 | 14 | 10 |
| M134 | 0.99 | 0.69 | 0.97 | 0.98 | 0.01 | 0.32 | 0.03 | 0.02 | 1.69 | 1.35 | 13 | 11 | 1.97 | 1.92 | 12 | 12 | 1.98 | 1.95 | 8 | 8 |
| M234 | 1.00 | 0.71 | 0.98 | 0.99 | 0.01 | 0.30 | 0.03 | 0.02 | 1.71 | 1.40 | 10 | 8 | 1.97 | 1.94 | 10 | 10 | 1.98 | 1.96 | 7 | 7 |
| M1234 | 1.00 | 0.76 | 0.98 | 0.99 | 0.01 | 0.30 | 0.03 | 0.01 | 1.76 | 1.45 | 6 | 7 | 1.98 | 1.94 | 7 | 9 | 1.98 | 1.96 | 6 | 6 |
| C123 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | 2.00 | 1.99 | 1 | 1 | 2.00 | 1.99 | 1 | 1 | 1.99 | 1.99 | 1 | 1 |
| C124 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | 2.00 | 1.99 | 4 | 4 | 2.00 | 1.99 | 4 | 4 | 1.99 | 1.98 | 5 | 5 |
| C134 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | 2.00 | 1.99 | 2 | 2 | 2.00 | 1.99 | 2 | 2 | 1.99 | 1.98 | 2 | 2 |
| C234 | 1.00 | 0.75 | 1.00 | 1.00 | 0.01 | 0.45 | 0.00 | 0.00 | 1.75 | 1.29 | 7 | 14 | 2.00 | 1.99 | 5 | 5 | 1.99 | 1.98 | 4 | 4 |
| C1234 | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.00 | 0.00 | 0.00 | 2.00 | 1.99 | 3 | 3 | 2.00 | 1.99 | 3 | 3 | 1.99 | 1.98 | 3 | 3 |

說明：U1 代表採用該指標搭配 ARCI 計算目標函數一之效用值，U2 則為目標函數二之效用值；R1 代表在目標函數一的排名，R2 則為目標函數二的排名。

# 附錄二

為了清楚展現各解釋變數之邊際貢獻度，本文針對 SVM, DT, Logit 三個單一鑑別模型分別進行測試。測試的方法是在 10 次交叉驗證過程的每組訓練及測試樣本中，分別計算採用全部變數及剔除某一特定變數後的表現差異（以 ARCI 及 ARII 衡量之），再將十組驗證之表現差異予以平均。最後，找出剔除變數後表現落差最大者，推論其邊際貢獻較為顯著。

上述的分析流程可以下列的虛擬程式碼 (pseudocode) 作為說明：

for each learning method $L$
    for $i$ = 1 to 40
        for $j$ = 1 to 10
            Train $L$ on $D_{-j}$ with all explanatory variables, compute performance $P_1$ on $D_j$
            Train $L$ on $D_{-j}$ with explanatory variables $X$-$i$, compute performance $P_2$ on $D_j$
            $M_j = P_2 - P_1$
        end for
        $M$ = Average($M_j$, $j$ = 1, …, 10)
    end for
end for

其中 $L$ 可為 SVM, DT, 或 Logit；($D_1$, $D_2$, …, $D_{10}$) 代表 10 次交叉驗證過程中所作的樣本切割；$X_i$, $i$ = 1, 2, …, 40 為本文所用的 40 個解釋變數；$P$ 可以為 ARCI 或 ARII。依此法所求得各變數在各種鑑別方法下的表現差異如表 19 所示，其中數字若為負值便代表剔除該變數後，模型鑑別的能力（以 ARCI 或 ARII 評估）有下降的情況。

表 19 剔除各變數後鑑別模型之績效差異

| | SVM | | DT | | Logit | |
|---|---|---|---|---|---|---|
| | ARCI | ARII | ARCI | ARII | ARCI | ARII |
| *GP_share_lyc_100* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00011 | 0.00500 |
| *GP_share_lyc_200* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00446 | -0.07579 |
| *GP_share_lyc_300* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | -0.00281 | -0.03333 |
| *GP_share_lyc_1* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | -0.00168 | -0.12000 |
| *GP_share_lyc_5* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | -0.00119 | 0.04135 |
| *GP_share_lyc_10* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00452 | -0.04944 |
| *GP_share_365d_100* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00459 | -0.06222 |
| *GP_share_365d_200* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00285 | -0.07500 |
| *GP_share_365d_300* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00014 | -0.08667 |
| *GP_share_365d_1* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00324 | -0.01722 |

| | | | | | | |
|---|---|---|---|---|---|---|
| *GP_share_365d_5* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | -0.00139 | -0.16556 |
| *GP_share_365d_10* | -0.00149 | 0.00000 | -0.00002 | -0.03333 | -0.00064 | -0.18571 |
| *GP_share_30d_100* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00147 | -0.08222 |
| *GP_share_30d_200* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00012 | 0.00635 |
| *GP_share_30d_300* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00180 | 0.03357 |
| *GP_share_30d_1* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00308 | -0.03143 |
| *GP_share_30d_5* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00116 | -0.18722 |
| *GP_share_30d_10* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00001 | 0.01444 |
| *Limit_ratio_volume* | -0.00149 | 0.00000 | 0.00000 | 0.00000 | 0.00488 | -0.00304 |
| *WBAS2_all* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00444 | -0.05000 |
| *Buy_sell* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00163 | -0.07500 |
| *Trades* | 0.00000 | 0.00000 | 0.00156 | 0.00667 | 0.00181 | 0.02802 |
| *Traders* | 0.00000 | 0.00000 | 0.00144 | 0.00000 | 0.00628 | 0.00135 |
| *Days* | 0.00000 | 0.00000 | 0.00000 | 0.01667 | 0.00287 | -0.08889 |
| *Volume* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | -0.00225 | 0.06611 |
| *Two_way* | 0.00588 | -0.10000 | 0.00000 | 0.00000 | 0.00178 | -0.06167 |
| *IP_share* | 0.00146 | -0.05000 | 0.00000 | 0.00000 | 0.00316 | -0.02000 |
| *Traded_order_ratio* | 0.00596 | -0.05000 | 0.00159 | 0.00167 | 0.00163 | 0.01944 |
| *Highest-price* | 0.00000 | 0.00000 | 0.00720 | -0.14333 | 0.00014 | 0.01944 |
| *NC* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00032 | -0.09722 |
| *Price_gap* | 0.00418 | -0.20000 | 0.00340 | 0.13500 | 0.00618 | -0.05167 |
| *Avatar_ratio_3* | 0.00594 | -0.05000 | 0.00000 | 0.00000 | -0.00148 | -0.01056 |
| *Avatar_15d_ratio_3* | 0.00286 | -0.15000 | 0.00000 | 0.00000 | 0.00104 | -0.06758 |
| *Avatar_30d_ratio_3* | 0.00000 | 0.00000 | 0.00000 | 0.00000 | 0.00117 | -0.08325 |
| *Avatar_365d_ratio_3* | 0.00291 | -0.15000 | 0.00000 | 0.00000 | 0.00160 | -0.17310 |
| *Avatar_volume_ratio_3* | -0.00005 | -0.10000 | 0.00000 | 0.00000 | 0.00134 | -0.07500 |
| *Avatar_volume_15d_ratio_3* | 0.00296 | -0.05000 | 0.00000 | 0.00000 | 0.00139 | -0.07000 |
| *Avatar_volume_30d_ratio_3* | -0.00005 | -0.10000 | 0.00000 | 0.00000 | -0.00279 | -0.21556 |
| *Avatar_volume_365d_ratio_3* | -0.00005 | -0.10000 | 0.00000 | 0.00000 | 0.00128 | -0.08000 |
| $P^w$ | 0.00894 | -0.06667 | 0.00306 | 0.15000 | 0.00002 | 0.01000 |

## 參考文獻

林鴻文、童振源、與葉家興 (2014), “選舉預測市場之選前鑑別模型：以最高價準則為門檻,” 《東吳政治學報》, 32(2), 117-171。(Lin, Hung-Wen, Chen-Yuan Tung, and Jason Yeh (2014), “The Discrimination Models of Accuracy for Election Prediction Markets

Prior to the Elections: Based on the Highest-price Criterion," *Soochow Journal of Political Science*, 32(2), 117-171.)

童振源、林馨怡、林繼文、黃光雄、周子全、劉嘉凱、與趙文志 (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), 131-166。(Tung, Chen-Yuan, Hsin-Yi Lin, Jih-Wen Lin, G. George Hwang, Tzu-Chuan Chou, Chia-Kai Liu, and Wen-Chih Chao (2009), "Prediction on Taiwan's Elections: Application and Empirical Studies of Prediction Markets," *Journal of Electoral Studies*, 16(2), 131-166.)

葉家興、童振源、林鴻文、與周子全 (2012), "判定預測市場準確度的模型," 工作論文稿。(Yeh, Jason, Chen-Yuan Tung, Hung-Wen Lin, and Tzu-Chuan Chou (2012), "A Model Assessing the Accuracy of Prediction Markets," working paper.)

Aizerman, Mark A., Emmanuel M. Braverman, and Lev I. Rozonoer (1964), "Theoretical Foundations of the Potential Function Method in Pattern Recognition Learning," *Automation and Remote Control*, 25, 821-837.

Allen, Kim, Kevin Daniels, Darby Kopp, and Brian Murdock (2004), "Analysis of 2004 Political Futures Markets," URL: http://www.intrade.com/news/images/Dartmouth_Election_Paper_11_9_04%20(2).doc.

Anzanello, Michel J. and Flavio S. Fogliatto (2011), "Selecting the Best Clustering Variables for Grouping Mass-Customized Products Involving Workers' Learning," *International Journal of Production Economics*, 130(2), 268-276.

Armstrong, J. Scott (2001), "Combining Forecasts," in J. Scott Armstrong (ed.), *Principles of Forecasting: A Handbook*, Springer.

Bates, James M. and Clive W. J. Granger (1969), "The Combination of Forecasts," *Journal of the Operational Research Society*, 20(4), 451-468.

Berg, Joyce E., Robert Forsythe, Forrest Nelson, and Thomas A. Rietz (2001), "Results from a Dozen Years of Election Futures Markets Research," In Charles Plott and Vernon Smith (eds.), *Handbook of Experimental Economics Results*, North Holland.

Berg, Joyce E., Robert Forsythe, and Thomas A. Rietz (1997), "What Makes Markets Predict Well? Evidence from the Iowa Electronic Markets," in Wulf Albers, Werner Güth, Peter Hammerstein, Benny Moldovanu, and Eric van Damme (eds.), *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*, New York: Springer.

Berg, Joyce E., Forrest Nelson, and Thomas A. Rietz (2008), "Prediction Market Accuracy in the Long Run," *International Journal of Forecasting*, 24(2), 285-300.

Berg, Joyce E. and Thomas A. Rietz (2003), "Prediction Markets as Decision Support Systems," *Information Systems Frontiers*, 5(1), 79–93.

Bergfjord, Ole J. (2008), "Prediction Markets as a Tool for Management of Political Risk," *The Journal of Prediction Markets*, 2(2), 1–12.

Blaskowitz, Oliver and Helmut Herwartz (2011), "On Economic Evaluation of Directional Forecasts," *International Journal of Forecasting*, 27(4), 1058-1065.

Blin, Jean-Marie and Andrew B. Whinston (1975), "Discriminant Functions and Majority Voting," *Management Science*, 21(5), 557-566.

Bodie, Zvi, Alex Kane, and Alan J. Marcus (2002), *Investments*. Boston: McGraw Hill.

Boser, Bernhard E., Isabelle M Guyon, and Vladimir N. Vapnik (1992), "A Training Algorithm for Optimal Margin Classifiers," *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh: ACM, 144-152.

Breiman, Leo (2001), "Decision Tree Forests," *Machine Learning*, 45(1), 5-32.

Britten-Jones, Mark (1999), "The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights," *Journal of Finance*, 45, 655-671.

Brüggelambert, Gregor (2004), "Information and Efficiency in Political Stock Markets: Using Computerized Markets to Predict Election Results," *Applied Economics*, 36(7), 742-768.

Chang, Chih-Chung and Chih-Jen Lin (2005), LIBSVM-A Library for Support Vector Machines, URL: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Chen, Shu-Heng, Chen-Yuan Tung, Chung-Ching Tai, Bin-Tzong Chie, Tzu-Chuan Chou, and Shu G. Wang (2011), "Prediction Markets: A Study on the Taiwan Experience," in Leighton V. Williams (ed.), *Prediction Markets: Theory and Applications*, Routledge.

Clemen, Robert T. (1989), "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 5(4), 559-583.

Cortes, Corinna and Vladimir Vapnik (1995), "Support-Vector Networks," *Machine Learning*, 20(3), 273-297.

Crane, Dwight B. and James R. Crotty (1967), "A Two-Stage Forecasting Model: Exponential Smoothing and Multiple Regression," *Management Science*, 13(8), B501-B507.

Crask, Melvin R. and William D. Perreault, Jr. (1977), "Validation of Discriminant Analysis in Marketing Research," *Journal of Marketing Research*, 14(1), 60-68.

Deck, Cary, Shengle Lin, and David Porter (2013), "Affecting Policy by Manipulating Prediction Markets: Experimental Evidence," *Journal of Economic Behavior and Organization*, 85(C), 48-62.

Deng, Shigan and Tsung-Han Yeh (2011), "Using Least Squares Support Vector Machines for the Airframe Structures Manufacturing Cost Estimation," *International Journal of Production Economics*, 131(2), 701-708.

Doumpos, Michael and Constantin Zopounidis (2009), "Monotonic Support Vector Machines for Credit Risk Rating," *New Mathematics and Natural Computation*, 5(3), 557-570.

Erikson, Robert S. and Christopher Wlezien (2008), "Are Political Markets Really Superior to Polls as Election Predictors?" *Public Opinion Quarterly*, 72(2), 190-215.

Ferri, Cèsar, Jose Hernández-Orallo, and Elena Ramona Modroiu (2009), "An Experimental Comparison of Performance Measures for Classification," *Pattern Recognition Letters* 30(1), 27-38.

Forsythe, Robert, Murray Frank, Vasu Krishnamurthy, and Thomas W. Ross (1995), "Using Market Prices to Predict Election Results: The 1993 UBC Election Stock Market," *Canadian Journal of Economics*, 28(4a), 770-793.

Forsythe, Robert, Forrest Nelson, George R. Neumann, and Jack Wright (1992), "Anatomy of an Experimental Political Stock Market," *American Economic Review*, 82(5), 1142–1161.

Forsythe, Robert, Thomas A. Rietz, and Thomas W. Ross (1999), "Wishes, Expectations and Actions: A Survey on Price Formation in Election Stock Markets," *Journal of Economic Behavior and Organization*, 39, 83-110.

Gao, Ciwei, Ettore Bompard, Roberto Napoli, Qiulan Wan, and Jian Zhou (2008), "Bidding Strategy with Forecast Technology Based on Support Vector Machine in the Electricity Market," *Physica A: Statistical Mechanics and its Applications*, 387(15), 3874-3881.

Gigerenzer, Gerd (2002), *Calculated Risks: How to Know When Numbers Deceive You*, Simon and Schuster.

Gruca, Thomas S., Joyce E. Berg, and Michael Cipriano (2005), "Consensus and Differences of Opinion in Electronic Prediction Markets," *Electronic Markets*, 15(1), 13-22.

Hahn, Robert W. and Paul C. Tetlock (2006), "Introduction to Information Markets," in Robert W. Hahn and Paul C. Tetlock (eds.), *Information Markets: A New Way of Making Decisions*, AEI Press.

Hartigan, John A. (1975), *Clustering Algorithms*. New York: John Wiley & Sons.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001), *The Elements of Statistical Learning*, Springer.

Hayek, Friedrich (1945), "The Uses of Knowledge in Society," *American Economic Review*, 35, 519–530.

Ho, Teck-Hua and Kay-Yut Chen (2007), "New Product Blockbusters: The Magic and Science of Prediction Markets," *California Management Review*, 50(1), 144-158.

Jacobsen, Ben, Jan Potters, Arthur Schram, Frans van Winden, and Jorgen Wit (2000), "(In)Accuracy of a European Political Stock Market: The Influence of Common Value Structures," *European Economic Review*, 44(2), 205-230.

Kambil, Ajit and Eric van Heck (2002), *Making Markets: How Firms Can Design and Profit from Online Auctions and Exchanges*, Cambridge, MA: Harvard Business School Press.

Kandel, Shmuel and Robert F. Stambaugh (1987), "A Mean-Variance Framework for Tests of Asset Pricing Models," *Review of Financial Studies*, 2(2), 125-156.

Kessler, Stephan and Bernd Scherer (2011), "Hedge Fund Return Sensitivity to Global Liquidity," *Journal of Financial Markets*, 14(2), 301-322.

Ledyard, John O. (2006), "Designing Information Markets for Policy Analysis," in Robert W. Hahn and Paul C. Tetlock (eds.), *Information Markets: A New Way of Making Decisions*, AEI-Brookings Joint Center.

Leigh, Andrew and Justin Wolfers (2006), "Competing Approaches to Forecasting Elections: Economic Models, Opinion Polling and Prediction Markets," *Economic Record*, 82(258), 325-340.

Leigh, Andrew, Justin Wolfers, and Eric Zitzewitz (2003), "What Do Financial Markets Think of War in Iraq," NBER Working Paper No. 9587, URL: http://www.nber.org/papers/w9587.

Liaw, Andy and Matthew Wiener (2002), "Classification and Regression by Random Forest," *R News*, 2(3), 18-22.

Loh, Wei-Yin and Nunta Vanichsetakul (1988), "Tree-Structured Classification via Generalized Discriminant Analysis," *Journal American Statistical Association*, 83(403), 715-725.

Luckner, Stefan, Christof Weinhardt, and Rudi Studer (2006), "Predictive Power of Markets: A Comparison of Two Sports Forecasting Exchanges," in Thomas Dreier, Rudi Studer, and Christof Weinhardt (eds.), *Information Management and Market Engineering*, Universitätsverlag Karlsruhe.

Maccheroni, Fabio, Massimo Marinacci, Aldo Rustichini, and Marco Taboga (2009), "Portfolio Selection with Monotone Mean-Variance Preferences," *Mathematical Finance*, 19(3), 487-521.

Mahdi, Salehi and Rostami Neda (2013), "Bankruptcy Prediction by Using Support Vector Machines and Genetic Algorithms," *Studies in Business and Economics*, 8(1), 104-114.

Manski, Charles F. (2006), "Interpreting the Predictions of Prediction Markets," *Economics Letters*, 91, 425-429.

Marček, Dušan and Milan Marček (2006), "Application of Dynamic Models and an Support Vector Machine to Inflation Modelling," *Bulletin of the Czech Econometric Society*, 13(23), 167-172.

Markowitz, Harry (1952), "Portfolio Selection," *Journal of Finance*, 7(1), 77-91.

Mitchneck, Beth (1995), "An Assessment of the Growing Local Economic Development Function of Local Authorities in Russia," *Economic Geography*, 71(2), 150-170.

Mitra, Devashish, Dimitrios D. Thomakos, and Mehmet A. Ulubasoglu (2002), "Protection for Sale in a Developing Country: Democracy Versus Dictatorship," *Review of Economics and Statistics*, 84(3), 497-508.

Morgan, James N. and John A. Sonquist (1963), "Problems in the Analysis of Survey Data and a Proposal," *Journal of the American Statistical Association*, 58(302), 415-434.

Newman, Bruce I. and Jagdish N. Sheth (1985), "A Model of Primary Voter Behavior," *Journal Consumer Research*, 12(2), 178-187.

Oliven, Kenneth and Thomas A. Rietz (2004), "Suckers are Born but Markets Are Made: Individual Rationality, Arbitrage, and Market Efficiency on an Electronic Futures Market," *Management Science*, 50(3), 336-351.

Pai, Ping-Feng and Chih-Sheng Lin (2005), "A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting," *Omega*, 33(6), 497-505.

Pal, Mahesh (2005), "Random Forest Classifier for Remote Sensing Classification," *International Journal of Remote Sensing*, 26(1), 217-222.

Payne, John W., James R. Bettman, and Eric J. Johnson (1993), *The Adaptive Decision Maker*. Cambridge University Press.

Rhode, Paul W. and Koleman S. Strumpf (2003), "Historical Presidential Betting Markets," *Journal of Economic Perspectives*, 18(2), 127-141.

Robertson, Thomas S. and James N. Kennedy (1968), "Prediction of Consumer Innovator: Application of Multiple Discriminant Analysis," *Journal of Marketing Research*, 5(1), 64-69.

Rosen, Dennis L. and Donald H. Granbois (1983), "Determinants of Role Structure in Family Financial Management," *Journal of Consumer Research*, 10(2), 253-258.

Segovia-Vargas, María Jesús, Sancho Salcedo-Sanz, and Carlos Bousoño-Calzón (2004), "Prediction of Insolvency in Non-Life Insurance Companies Using Support Vector Machines, Genetic Algorithms and Simulated Annealing," *Fuzzy Economic Review*, 9(1), 79-94.

Smith, Vernon (1982), "Markets as Economizers of Information: Experimental Examination of the 'Hayek Hypothesis'," *Economic Inquiry*, 20(2), 165–179.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz (2005), "Information (In)Efficiency in Prediction Markets," in Leighton V. Williams (ed.), *Information Efficiency in Financial and Betting Markets*, Cambridge, UK: Cambridge University Press.

Sokolova, Marina and Guy Lapalme (2009), "A Systematic Analysis of Performance Measures for Classification Tasks," *Information Processing and Management*, 45(4), 427-437.

State, Luminita, Catalina Cocianu, and Doina Fusaru (2010), "A Survey on Potential of the Support Vector Machines in Solving Classification and Regression Problems," *Informatica Economica*, 14(3), 128-139.

Stock, James H. and Mark W. Watson (2004), "Combination Forecasts of Output Growth in a Seven-Country Data Set," *Journal of Forecasting*, 23(6), 405-430.

Takieddine, Samer and Francis Kofi Andoh-Baidoo (2014), "An Exploratory Analysis of Internet Banking Adoption Using Decision Tree Induction," *International Journal of Electronic Finance*, 8(1), 1-20.

Tam, Kar Yan and Melody Y. Kiang (1992), "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," *Management Science*, 38(7), 926-947.

Tay, Francis E.H. and Lijuan Cao (2001), "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, 29(4), 309-317.

Tobin, James (1958), "Liquidity Preference as Behavior Towards Risk," *Review of Economic*

*Studies*, 25, 65-86.

Tso, Geoffrey K.F. and Kelvin K.W. Yau (2007), "Predicting Electricity Energy Consumption: A Comparison of Regression Analysis, Decision Tree and Neural Networks," *Energy*, 32(9), 1761-1768.

Vapnik, Vladimir and Aleksandr Lerner (1963), "Pattern Recognition Using Generalized Portrait Method," *Automation and Remote Control*, 24, 774-780.

Walker, David A. (2006), "Predicting Presidential Election Results," *Applied Economics*, 38, 483-490.

Wilton, Peter C. and Edgar A. Pessemier (1981), "Forecasting the Ultimate Acceptance of an Innovation: The Effects of Information," *Journal of Consumer Research*, 8(2), 162-171.

Wolfers, Justin and Andrew Leigh (2002), "Three Tools for Forecasting Federal Elections: Lessons from 2001," *Australian Journal of Political Science*, 37(2), 223-240.

Wolfers, Justin and Eric Zitzewitz (2004), "Prediction Markets," *Journal of Economic Perspectives*, 18(2), 107-126.

Wolfers, Justin and Eric Zitzewitz (2006), "Prediction Markets in Theory and Practice," NBER Working Paper No. 12083, URL: http://www.nber.org/papers/w12083.

Zarnowitz, Victor (1967), *An Appraisal of Short-Term Economic Forecasts*, New York: Columbia University Press.

# Assessing the Accuracy of Prediction Markets: Single Versus Combined Identification Models

Chung-Ching Tai

*Department of Economics, Tunghai University*

Ping-Tsung Chih

*Department of Industrial Economics, Tamkang University*

Hung-Wen Lin

*Department of Economics and Business Management, Nanfang College of Sun Yat-Sen* University

Chen-Yuan Tung

*Graduate Institute of Development Studies, National Chengchi University*

As prediction markets (PM) being used widely in many fields, contemporary researchers and practitioners have to rely on historical accuracy to evaluate the plausibility of current events. Based on the empirical and theoretic findings on the accuracy of prediction markets, this paper proposes a combined identification method which can evaluate the accuracy of PM events in advance. The proposed method not only takes a variety of market features into account, but also combines the forecasts of different statistical and machine learning techniques to fully capture the patterns underneath. We test the proposed method with transaction data from 2006 to 2011. This study proves that it is possible to evaluate the accuracy of the any PM event in advance with high accuracy. We also show that the combined modeling is a superior method in the sense that it not only can provide higher identification accuracy, but is also flexible enough to incorporate decision makers' goals and preferences into the identification process.

Keywords: Prediction markets, Identification in advance, Combined forecast, Logistic regressions, Principal component analysis, Discriminant analysis, Decision trees, Support vector machine

JEL classification: C38, C53, G13, G14