行政院國家科學委員會專題研究計畫 成果報告

自覺選號對樂透彩券價格需求彈性之影響
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 96-2415-H-263-003-
執 行 期 間 ：96年08月01日至97年07月31日
執 行 單 位 ：致理技術學院財務金融系(科)

計畫主持人：李世欽

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫可公開查詢

中 華 民 國 97年10月28日

行政院國家科學委員會補助專題研究計畫 ■ 成果報告
□期中進度報告

自覺選號對樂透彩券價格需求彈性之影響

計畫類別：■ 個別型計畫　□ 整合型計畫
計畫編號：NSC 96 － 2415 － H － 263 － 003 -
執行期間： 96 年 08 月 01 日至 97 年 07 月 31 日

計畫主持人：李世欽

共同主持人：

計畫參與人員：

成果報告類型(依經費核定清單規定繳交)：■精簡報告
□完整報告

本成果報告包括以下應繳交之附件：
□赴國外出差或研習心得報告一份
□赴大陸地區出差或研習心得報告一份
■出席國際學術會議心得報告及發表之論文各一份
□國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

執行單位：致理技術學院

中 華 民 國 97 年 10 月 28 日

中文摘要

傳統文獻關於自覺選號之研究或是彩券價格需求彈性之估計，只能利用公開市場資訊來做分析，然而此種估計的分析方法可能有偏差，失去若干相當程度的準確性。本文主要是利用台北銀行所提供的 2002 年到 2003 年國內樂透彩券民眾實際下注資料，去探討自覺選號的行為，並利用彩券報酬去估計需求彈性。本研究發現自覺選號投注者之選號不為隨機，每一期號碼的偏好也不固定。總銷售量與自覺選號投注的彩券價格需求彈性分別為 1.4與 1.8。

關鍵字：自覺選號，台灣樂透彩，需求彈性

英文摘要

Recent work has attempted to investigate the behavior of conscious selection and estimate the price elasticity of demand for lotto tickets by using only public available data. However, the results obtained from this type of data may be biased or inaccurate.The purpose of this paper is to identify patterns in the conscious selection behavior of players based on an actual time series database maintained by the only lottery operator in Taiwan. Moreover, I employ the expected value of holding lottery ticket so as to estimate the demand elasticity for lotto tickets. Using the combinations of combinations by lotto players from 2002 to 2003, I find that conscious selection players do not bet randomly and consistently over time. Furthermore, the price elasticity for the total demand and proportion of conscious selection are 1.4 and 1.8, respectively.

Key words: Conscious selection, Taiwan Lotto, price elasticity of demand

一、前言及研究目的

本文的主要目的在探討國內樂透彩券投注者行為，利用民眾實際下注資料去計算有效價格及建立彩券價格需求之模型，推估彩券之彈性。因彩券發行或管理機構並不揭露投注的分配，所以文獻關於自覺選號之研究或是彩券價格需求彈性之估計，只能利用公開市場資訊：各期中獎人數及銷售量來做分析。然而此種估計的分析方法，可能對於彩券投注者行為之反推或需求彈性之估計，失去若干相當程度的合理性與準確性。了解彩券購買者的行為，或是分析影響彩券銷售量的變數是重要的，尤其是當台灣民眾對彩券的新鮮感漸漸消退，銷售量隨著時間的增加而快速下降時，研究彩券之需求彈性，提供給政府或新的彩券發行商，做為制定票面價格的參考，是有其必要性。

本文主要研究的對象為台北銀行首次於民國 91 年所發行之電腦型樂透彩券(6/42)，在台北銀行經營彩券銷售業務的第一年內，根據台北銀行經理楊瑞東(民 92) 指出，其九十一年度「彩券元年」公益彩券銷售總金額就達到一千零七十三億台幣，超過國內生產毛額的 1%，而與世界各國比較發現，在全世界的國家銷售量排名中為第九名，在亞洲僅次於日本，而國內的電腦彩券銷售量則為世界排名的第四名，亞洲排名的第一名。在購買者的調查上顯示，有超過六成的民眾的買過彩券，其中七成教育程度為高中以上學歷，購買者以中產階級為主。此現象顯示彩券市場之重要性與普遍性是不容忽視的。

本計劃利用台北銀行所提供的國內樂透彩券民眾實際下注資料，去研究自覺選號的行為，及利用有效價格去建立彩券價格需求之模型，去了解民眾投注行為對彩券價格需求彈性之影響程度。因此，所擬用的資料分為二部分，一為台北銀行官方網站所 公告的資料：每週之週二與週五開獎 之6 個中獎號碼及1個特別號，各期之銷售總額，頭獎、貳獎、參獎、肆獎、及普獎的金額與中獎人數。另一個資料庫為台北銀行所提供的每期每注民眾自己實際下注的號碼，此資料不包括投注者以電腦選號下注的方式，而只有以所謂的「自覺選號」投注方式的號碼。有別於傳統文獻作法，對彩券需求的探討只看整體投注者的購買量，本文嘗試分離出自覺選號銷售量的部分作研究，並以事後實現報酬的觀點，計算出每一張彩券的報酬，估計出整體及自覺選號的需求彈性。

二、文獻探討

國外對於彩券相關議題的研究發展已經相當成熟 ，因國內引進彩券時間較晚，以樂透彩券為研究對象的文獻相對上也比較少。一些對台灣彩券購買者的行為的分析，或是推估彩券之需求彈性的研究，多半是以問卷調查方式，用個人所得、年紀、職業、性別、教育程度等來了解彩券購買者的動機與需求，或是用公

益彩券的財貨特性以及與六合彩的替代關係，來推估彩券之彈性(陳定國、婁國仁與陳定國，民91；劉代洋與陳慧琪，民90）。另一方向之研究則以中獎人數及銷售量等時間序列資料，來推估購買者的行為及計算彩券之需求彈性。然因缺乏真實資料下，在估計彩券之期望價格時，為簡化分析，不是考慮隨機選號下的模型，就是只考量頭獎的預期金額下做研究，而忽略二獎、三獎、四獎及普獎的報酬（林靖中、康信鴻和詹司如（民93）、林玫吟與王智賢（民94）、王智賢與林玫吟（民95）及 Tzeng, Wang與 Tain (2007)）。如此估計出來的彩券之預期價格或是彩券需求彈性，可能會有所誤差。

國外文獻較常用彩券的有效價格（effective price），來估計彩券之價格需求彈性（Cook and Clotfelter, 1993; Gulley and Scott, 1993; Scott and Gulley, 1995; Walker, 1998; Farrell and Walker, 1999; Purfield and Waldron, 1999; Farrell, Lanot, Hartley and Walker, 2000）。其中以 Farrell, Lanot, Hartley and Walker (2000)的研究方法較經典也較詳細，作者研究英國樂透彩券需求彈性，因彩券發行或管理機構並不揭露民眾投注的分配，無法得其知真實投注行為，所以假設民眾之投注號碼分配是每期都固定不變，及各組選取的號碼彼此是獨立的下，利用各期之中獎人數及銷售金額之時間序列資料去推估其投注之分配。該研究發現投注者選擇他們的號碼不是隨機選，其投注之分配不是均等分配（Uniform），其特質表現出來的是無人中頭獎的次數（roll over）比隨機選模型還要高出許多，他們稱此行為為自覺選號（conscious selection）。若以購買彩券的報酬率來看，在一般期間或是有二次無人中頭獎時，隨機選號的平均報酬率為 0.45 至 0.67 ，但若是自覺選號策略下購買彩券的人，卻只有 0.44 至 0.65 。接著用三種方法方法估計：一般迴歸方式、代理變數及非線性之 FIML (Full information maximum likelihood)，得到的是彩券需求彈性接近一，顯示是彩券經營機構是有達到極大化利潤的目標。雖然該研究有考慮頭獎及其他獎項之報酬率，用自覺選號所估計的分配來做一些預測之效果也比均等分配模型較好，但不過用中獎人數及銷售金額之資料來反推民眾之真實投注行為，仍有一些極大的落差，顯示之前的假設所做的推估，無法完全捕捉市場真實的行為。同樣的問題也出現於 Cox, Daniell, 及 Nicole (1998) 所做的研究，用極大化熵的方式(maximum entropy) 來推估投注之真實分配，仍出現許多極端值。

三、研究方法

為了解國內樂透彩彩券投注者選號碼的偏好，及電腦選號與自覺選號行為之差異比較分析，本研究由北銀所提供的二年每期人工投注的選號明細資料，配合北銀官方網站公告的每期開獎號碼，可計算出每期民眾由自覺選號策略投注且中獎的人數。採電腦選號策略的每期中獎人數及銷售量，則由北銀官方網站公告的總中獎人數及總銷量平減掉採自覺選號策略投注之中獎的人數及銷售

量可求得，中獎人數可分為頭獎、貳獎、參獎、肆獎及普獎，中獎率則為總中獎人數除以銷售量。

本文首先利用各獎項之中獎率變數之相關性來觀察自覺選號的行為，探討其隨機程度是否異於電腦選號。若是隨機選號模型下，則此中獎率之理論值應為 2.906。其次，並檢定是否此種自覺選號非隨機的現象是否表示民眾對 42 個號碼的偏好是不一樣的，某些號碼很熱門，某些則為冷門球號，且每一期民眾之投注號碼分配之偏好是固定不變的假設。利用選出三個球號來，分別代表在二年的銷售期間，平均、最熱門及最冷門的號碼，探討其之投注者下注之時間頻率是否每一期皆變動不大，是否符合 Farrell, Lanot, Hartley and Walker (2000) 的假設，選號的頻率每期固定不變的假設。

若自覺選號的投注者其選號方式，其隨機程度遠小於電腦選號，則用隨機選號之二項分配計算每一張彩券的期望報酬，其結果可能較有很大的偏差。且若投注者每一期的之投注號碼分配不是固定不變，則無法用事前估計出每一個球號的選號機率，再求出每一張彩券的期望報酬。因此本研究採用事後的實現報酬來當成事前期望報酬的一個替代變數，即為每一期每一個獎項的獎金乘以其中獎人數，再除以銷售量乘以彩券面額。

(1) $EV = \frac{1}{Q} \times [Rollover + Superdraw + (1-t)k(50Q - 200W_5) + \mathrm{Re}alize\mathrm{Re}turn_{small}]$

EV是每一張彩券事後實現的報酬，Q是銷售量，Rollover 是上一期頭獎獎項之累積獎金，Superdraw為台北銀行促銷之加碼獎金，t為政府徵收率，k 為頭獎獎金分配比率，W_5 為普獎中獎人數，$\mathrm{Re}alize\mathrm{Re}turn_{small}$ 是事後實現的小獎報酬。本文將上述各變數分整體投注量及自覺選號二部分各自處理，分別計算出二者之彩券報酬。在Rollover 及Superdraw 為零下，因政府徵收率為 44%，彩券面額為50元，所以彩券事後實現的報酬值應為28。

因解釋變數銷售量與彩券報酬之間有內生性，故本研究用兩階段最小平方法去處理此問題。第一階段：先以最小平方法(ordinary least squares, OLS) 去對彩券報酬與銷售量前一期之外生變數的線性關係進行估計。第二階段：再以第一階段所估計的彩券報酬配適值，計算出彩券有效價格，為將彩券之票面價格減去彩券之期望報酬。再代入原為內生變數的解釋變數，以OLS 法估計銷售量與彩券有效價格配適值間的線性迴歸關係。銷售量與彩券有效價格均採對數形式，求得迴歸係數即是彩券需求之彈性。

(2) Stage 1: EV =f(Constant, lag_q, Trend, Rollover, Superdraw)

(3) Stage 2: q =f(Constant, lag_q, Trend, Ep,Wc)

其中 q 表銷售量之自然對數，lag_q表前一期之銷售量，Ep為彩券之有效價格，Wc表前期中獎人數比例，Trend表時間變數。

四、 結果與討論

表 1 為樂透彩券每期中獎人數及銷售量的敘述統計，Panel A與Panel B 分別為由電腦選號及自覺選號策略投注下的中獎計算結果。由此表可發現，以銷售量平均值來說，採自覺選號策略投注的銷售量為電腦選號的 1.6 倍，約占總銷售量的六成，相對於世界大部分的樂透，其快速選號（Lucky Dip）的服務，亦即提供玩家一組隨機的號碼組合，每一期採用此選號的比例大約是占總銷售量的百分之十，其資料顯示台灣民眾購買樂透彩彩券的方式多以自覺選號居多，若非用電腦選號策略去投注，在研究上，選號的隨機性便是一個非常重要的因素，因為它反應在計算彩券的報酬上。

其次，由表 1 可看出，因自覺選號之銷售量較高，因此自覺選號策略投注下的每一個獎項的中獎人數平均值自然也高於電腦選號，將其相除以銷售量下，約為 1.5 至 1.6 倍，二者之中獎率也符合理論之預期。採電腦選號的中獎率幾乎在 2.7 與 3.2 中間波動，但自覺選號之值波動幅度較大，各獎項之標準差均高於電腦選號，如自覺選號最高時，可一次很多人來領獎，中獎率高達 5.13%，最差之狀況為 1.34%，買的人多半損龜，其可能反應出採自覺選號與電腦選號的投注者行為明顯不同。

表 2 為樂透彩券每期中獎人數比率相關係數表，中獎人數比率為每個獎項之中獎人數除以當期銷售量，Panel A 與 Panel B 分別為由電腦選號及自覺選號策略投注下的中獎計算結果。由此表可發現，採自覺選號策略投注的各獎比率相關性高於電腦選號，其普獎中獎率 (M3)分別與頭獎 (M6)、貳獎 (M5+B)、參獎 (M5)及肆獎 (M4)之相關係數為 0.46、0.52、0.77、0.95。此可推論出，採自覺選號策略的投注者其選號模式非常接近，以致於一旦中獎時，便有很多人一起分享中獎獎金，由上述的分析結果可知，自覺選號的投注者其選號方式，其隨機程度遠小於電腦選號。

然而，此種自覺選號非隨機的現象是否表示，如 Farrell, Lanot, Hartley and Walker,(2000) 所推論，民眾對 42 個號碼的偏好是不一樣的，某些號碼很熱門，某些則為冷門球號，且每一期民眾之投注號碼分配之偏好是固定不變的假設。本文選出三個球號來，分別代表在二年的銷售期間，平均、最熱門及

最冷門的號碼，分別為 c5 ， c12 ， c42 ，將其之投注者下注頻率劃於圖 1 。由圖 1 可知，此三者的平均被民眾所選到的機率確實有有熱門及冷門球號的差別，但與(Farrell, Lanot, Hartley and Walker,2000）的假設不一致，三者選號的頻率每期不是固定。台灣民眾的選號方式雖然是自覺選號，但不是每一期皆固定不變，而是有可能跟著中獎號碼的資訊才決定下注的號碼（何淮中、李世欽與林修葳，民 95）。

由上述選號行為之分析可知，自覺選號投注者不是隨機選號方式，且每一期的民眾之投注號碼分配之偏好也不是固定不變的。因此國外傳統文獻上關於每張彩券的期望報酬的計算方法，較不能適用於台灣的彩券上，國內文獻關於需求彈性的研究也須加以修正。表 3 為全體彩券投注者及自覺選號的銷售量與期望報酬，此銷售量是採自然對數方式，以方便計算彩券之需求彈性。若每一期都有人中頭獎，即無上期頭彩彩金累積獎金至本期，則每一張彩券的期望報酬就會等於獎金支出比率乘以彩券的購買價格（林靖中、康信鴻與詹司如，民 93)。從表 3 可知，以總銷量所導出之彩券的期望報酬(EV_Total)，平均之彩券的期望報酬平均數為 28.02，接近於理論值 28，自覺選號策略所計算出的報酬(EV_Conscious)值略低，為 26.15，標準差也較大，為 8.59。因為相對於電腦選號，採自覺選號的投注者，因選號彼此相關性高，且樂透彩是採獎金分配制，所以投注者雖然享受中獎喜悅，但一旦中獎後可分配之金額卻變少，這點類似於財務上所謂的贏家詛咒（Thaler, R.H. ,1992)。

表4為兩階段最小平方法所估計出的彩券銷售量與彩券有效價格之結果，由此表可發現，Trend 的係數為皆為負值，顯示當隨著時間的增加,台灣民眾對彩券的新鮮感漸漸消退，同時也導致銷售量隨著時間的增加而快速下降 ，這點與(Mikesell,1987 ; Clotfelter and Cook,1989 ; Gulley and Scott,1993 ; Mikesell,1994；Miers,1996；林靖中、康信鴻與詹司如,民93；林玫吟與王智賢,民94；王智賢與林玫吟,民96）結果一致。不過自覺選號銷售量下降的程度較總銷售量少，符合控制幻覺（Illusion of control）的預期，Langer (1975)發現號碼是自己親自選取的，相對於號碼是隨機選取的彩券，他們願意付出較高的價格來才會放棄此張彩券。擲骰子賭局的實驗中，雖然無論擲骰子的人是誰，輸贏的機率應一樣，但受試者在遊戲裡若是自己親自擲骰子下，則會投注較高金額的賭注且自認為自己的贏的機率較高，Strickland et all.,(1966)。

用前期中獎人數比例 （WC_{t-1}）檢測是否有私房錢的現象 （House money effect），結果其迴歸係數皆不顯著。前一期之銷售量的係數皆為正數，顯示有若干程度的上癮性，尤其是以自覺選號程度較大。迴歸係數 EP即為需求彈性之值，總銷售量之需求彈性約為1.4，自覺選號值較大約為1.7至1.8，彩券價格有降價的空間，可透過提高獎金支付比率或是調整獎金分配率提高報酬著手，且如

為吸引自覺選號者，則降價空間應可更大。當隨著時間的增加，台灣民眾對彩券的新鮮感漸漸消退，同時也導致銷售量隨著時間的增加而快速下降，造成政府及彩券發行商有財政及經營的壓力，本研究用台北銀行所提供的國內樂透彩券民眾實際下注資料，去研究自覺選號的行為，了解民眾投注行為對彩券價格需求彈性之影響程度，較能精確的估計彩券之需求彈性，可提供做為未來制定票面價格的參考。

參考文獻

林靖中、康信鴻與詹司如，(民 93)，以時間數列觀點談樂透銷售量之行銷策略意涵，《台大管理論叢》，15，99-122 頁。

林玫吟與王智賢，(民 94)，台灣樂透彩券之價格需求彈性：隨機選號與自覺選號下之估計，《人文及社會科學集刊》，16，487-503 頁。

王智賢與林玫吟，(民 95)，Number selection strategy of lottery players-an empirical study of the Taiwan lottery，《台灣經濟預測與政策》，37， 49-67 頁。

何淮中、李世欽與林修葳，(民95)，Gambler's Fallacy in the Taiwan Lotto Market，台大經濟論文叢刊，forthcoming。

台北銀行，(民 92)，中華民國公益彩券網站。

陳定國、婁國仁與陳定國，(民 91)，台北市地區彩券購買行為之研究，淡江大學，碩士論文。

陳櫻琴與徐長芳，(民 93)，我國公益彩券發行之研究，中原大學，碩士論文。

楊瑞東，(民 92)，「彩券有情，公益用心」，國立中央大學經濟系，當代經濟學專題演講。

劉代洋與楊志唯，(民 90)，我國公益彩券獎金結構之分析，《管理與系統》，8，335-348 頁。

Clotfelter, C. and P. Cook (1989), *Selling hope: State lotteries in America*. Cambridge: Harvard University Press.

Cook, p. and C. Clotfelter (1993), The peculiar scale economies of lotto, *American Economics Review,* 83(3), 634-643.

Cox, S.J., G.J. Daniell, and D.A. Nicole (1998), Using maximum entropy to double one's expected winnings in the UK National Lottery, *The Statistician*, 47, 629-641.

Farrell, L., and I. Walker (1999), The welfare effects of lotto: evidence from the U.K., *Journal of Public Economics*, 72(1), 99-120.

Farrell, L., G. Lanot, R. Hartley, and I. Walker (2000), The demand for lotto: the role of conscious selection, *Journal of Business and Economic Statistics*, 18, 228-241.

Gulley, O. and F. Scott (1993) The demand for wagering on state-operated lotto games. *National Tax Journal*. 46(1), 13-22.

Haigh, J. (1997), The Statistics of the National Lottery*Journal of the Royal Statistical Society A*. 160 (2) pg.187-206。

Miers, D. (1996), The Implementation and effects of Great Britain's National Lottery, *Journal of Gambling Studies*, 12(4), 343-373.

Mikesell, J. (1987), The effect of maturity and competition on state lottery markets, *Journal of Policy Analysis and Management*, 6, 251-253.

Mikesell, J. (1994), State lottery sales and economic activity. *National Tax Journal*. 47(1), 165-171.

Purfield, C. and P. Waldron (1999), Gambling on lotto numbers: testing for substitutability or complementarity using semi-weekly turnover data, *Oxford Bulletin of Economics and Statistics*, 61(4), 527-544.

Strickland, L.H., R.J. Lewicki, and A.M. Katz (1966), Temporal orientation and perceived control as determinants of risk-taking, *Journal of Experimental and Social Psychology*, 2, 143-151.

Thaler, R.H. (1992), *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, New York: The Free Press.

Walker, I. (1998), The economic analysis of lotteries, *Economic Policy*, 13(27), 359-392.

表 1 敘述統計

Panel A 電腦選號

PC	頭獎人數	貳獎人數	參獎人數	肆獎人數	普獎人數	銷售量	中獎率
Min:	0	0	82	4369	66438	2.45E+06	2.701212
Mean:	0.926108	6.093596	209.468	9344.271	141167.7	5.18E+06	2.908391
Median:	1	5	171	7744	117901	4.32E+06	2.907314
Max:	6	28	944	43345	659623	2.45E+07	3.239853
Std Dev.:	1.116686	4.573504	134.268	5901.055	89093.61	3.28E+06	0.055977
Skewness:	1.525676	2.000788	2.898544	2.984212	3.008998	3.03E+00	1.331322
Kurtosis:	2.892141	5.074884	10.53068	11.16758	11.3938	1.16E+01	9.831911

Panel B 自覺選號

Concious	頭獎人數	貳獎人數	參獎人數	肆獎人數	普獎人數	銷售量	中獎率
Min:	0	0	61	4185	83097	4.44E+06	1.340349
Mean:	1.394089	8.837438	322.2414	14670.48	223178.8	8.27E+06	2.890547
Median:	1	7	256	12794	206888	7.64E+06	2.818574
Max:	11	77	1814	60130	811928	3.07E+07	5.13212
Std Dev.:	1.901632	8.271876	244.4248	7907.978	101932.2	3.32E+06	0.726077
Skewness:	2.483605	3.650956	2.734662	2.003151	2.126934	2.64E+00	0.566956
Kurtosis:	7.716703	23.64532	11.09496	7.25265	8.637089	1.17E+01	0.336954

表 2 相關係數表

Panel A 電腦選號

	M6	M5+B	M5	M4	M3
M6	1.00	-0.06	0.02	0.01	0.03
M5+B	-0.06	1.00	0.05	0.12	0.14
M5	0.02	0.05	1.00	0.51	0.47
M4	0.01	0.12	0.51	1.00	0.88
M3	0.03	0.14	0.47	0.88	1.00

Panel B 自覺選號

	M6	M5+B	M5	M4	M3
M6	1.00	0.38	0.66	0.54	0.46
M5+B	0.38	1.00	0.51	0.56	0.52
M5	0.66	0.51	1.00	0.87	0.77
M4	0.54	0.56	0.87	1.00	0.95
M3	0.46	0.52	0.77	0.95	1.00

圖 1 三個球號之下注頻率之時間序列圖

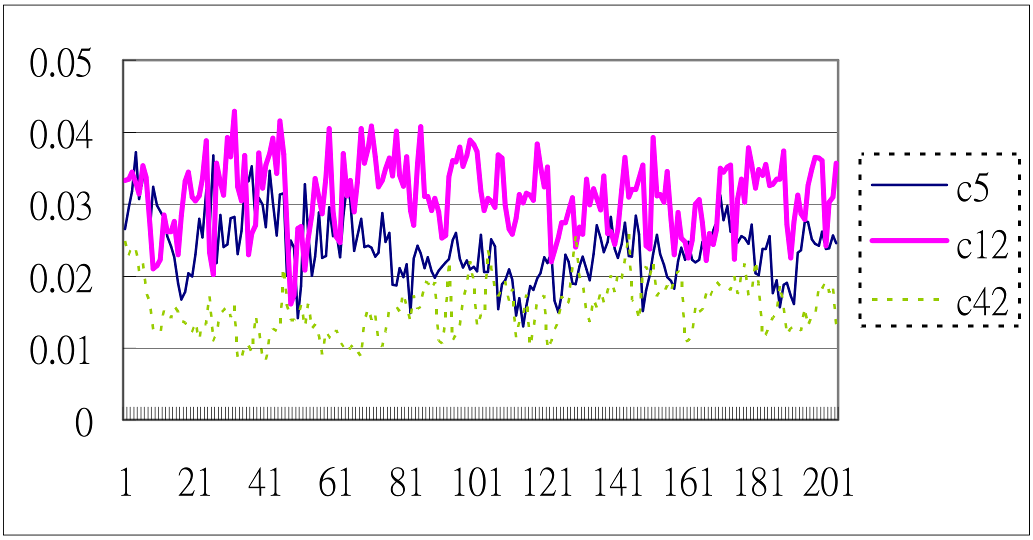


表 3 銷售量與期望報酬

	Sales_Total	Sales_Conscious	EV_Total	EV_Conscious
Min:	6.84261	6.84261	18.84413	10.43544
Mean:	7.09390	7.09486	28.02402	26.14198
Median:	7.07632	7.07783	27.99939	26.01362
Max:	7.74200	7.74200	40.35895	48.10515
Std Dev.:	0.16273	0.16273	4.81015	8.59056
Skewness:	1.04397	1.02826	-0.31825	0.30793
Kurtosis:	1.50283	1.47813	0.16155	-0.74905
Total N:	203	203	203	203

表 4 彩券銷售量之估計

2TSTL	**Sales_Total**				**Sales_Conscious**			
	Model 1		**Model 2**		**Model 3**		**Model 4**	
Independent variable	Coefficient	T-Value	Coefficient	T-Value	Coefficient	T-Value	Coefficient	T-Value
Intercept	6.3845	17.13	6.4001	17.14	5.2349	13.64	5.2640	13.64
Trend	-0.0012	-9.47	-0.0012	-9.49	-0.0006	-4.76	-0.0006	-4.81
LAG Q	0.3874	8.36	0.3855	8.31	0.6089	11.01	0.6049	10.89
EP	-1.4294	-13.64	-1.4153	-13.33	-1.8003	-10.28	-1.7865	-10.14
WC			-0.0072	-0.87			-0.0067	-0.80
R-squared	0.7376		0.7386		0.6493		0.6504	

本研究的內容與原計畫完全相符。研究結果可供彩券主管單位與學術界參考。

出席國際學術會議心得報告

計畫編號	NSC 96-2415-H-263 -003 -
計畫名稱	自覺選號對樂透彩券價格需求彈性之影響
出國人員姓名 服務機關及職稱	李世欽
會議時間地點	2008/5/18~2008/5/20 杭州, 大陸
會議名稱	第十五屆全球金融年會暨第五屆中國（杭州）WTO與金融工程國際會議
發表論文題目	Selection behavior of Taiwan lotto player-Dynamic analyses of number selection

一、參加會議經過

（一） 5 月 17 日：搭乘國內航空公司，經香港到中國大陸杭州蕭山機場，完成研討會註冊與報到手續。

（二） 5 月 18日：研討會正式開始，會議地點為浙江省人民大會堂國際會議廳。早上為會議簽到及會議開幕式，由浙江省民營企業國際合作促進會沈毅俊秘書長主持，大會主席程惠芳教授作專題演講：跨國投資與全球金融服務相關議題。

（三） 5 月 19日：所有與會的專家學者於浙江省人民大會堂門口拍照。早上的會議是美國全球金融學會主席 Manuchehr Shahrokhi 教授與程惠芳教授 主持，並由大會 keynote speaker Mike Moore 作專題演講 : Globalization, challenges, threats. 下午聆聽全球资本市場發展與風險管理相關

議題。

（四） 5 月 20日：最後一天的研討會地點改為浙江人民工業大學屏峰校區，早上聆聽各分組報告，了解專家學者最近熱門的實務現況與學術研究。下午由本人於博易樓 C204 研究室進行學術論文口頭發表 ，發表題目為 Selection behavior of Taiwan lotto players-Dynamic analyses selection，會議主席與與會的教授們也提供 相當重要與寶貴的意見，作為本研究論文改進的參考。

（三） 5 月 21 日：搭機返國，經香港返國

二、與會心得

15th Annual Global Finance Association Conference 是由中華人民共合國的浙江工業大學舉辦的，我的研究主題– Selection behavior of Taiwan lotto players-Dynamic analyses selection, 台灣樂透彩券投注者選號行為-動態選號模式分析，很榮幸的被大會邀請公開報告。本人在學術論文口頭報告發表時，獲得與會專家學者一致好評外，藉由他們的問題與意見，更能了解本研究不足之處，將來作為修正的參考，並朝頂尖期刊方向邁進，增加自己的研究能力與學術貢獻度。

自從大陸加入世界貿易組織後，中國的開放經濟創造了亮眼的成績。第 15 屆全球財務學會年會和第 5 屆世界貿易組織與財務工程的中國國際會議在杭州舉行，將近 400 個國家與外國專家，政府官員、國內外學術界專家

教授與企業界金融家家參加此場會議，一同討論全球與中國的經濟發展、金融改革與創新、美國次級債風暴、通脹與宏觀調控、際化與政府服務創新 、上市公司監管與治理資本市場現況、財務工程、統計計量、財務管理、海外投資與風險管理、企業海外上市風險管理、匯率變動與外匯投資基金風險管理和 WTO 國際貿易爭端議題，同時也交換各界的學術研究成果。這次會議共收到超過 300 篇學術報告 ，超過 200 個主題將 由東方教育研討會 (Oriental Education Forum) 編輯出版 ，在此會議中，國際和地方的金融家與企業家一起聚集在杭州建立起西部國際金融專業委員會,大會主席與專家學者共同期望能創造最明亮的未來。

這一次之行，非常感謝國科會補助，讓本人能前往中國大陸參加國際學術研討會。除了學術論文與外語能力的提升 ，與國外專家學者加強國際交流，增進此領域的新知互動外，更能了解中國大陸金融業與學術的發展，也確定了中國大陸在全球资本化市場實務和學術發展中,扮演一個非常重要的角色。

Selection behavior of Taiwan lotto players

---Dynamic analyses of number selection

Hwai-Chung Ho [1], Shih-Chin Lee [2], Hsiou-wei Lin [3]

Jan. 8, 2008

[1] Institute of Statistical Science, Academia Sinica and Department of Finance, National Taiwan University
E-mail: hcho@stat.sinica.edu.tw

[2] Department of Finance, Chihlee Institute of Technology
E-mail: d89723008@ntu.edu.tw

[3] Department of International Business, National Taiwan University
E-mail: plin@management.ntu.edu.tw

Abstract

This paper develops various dynamic regression models to describe the betting behavior of the Taiwan lotto players. The data collected for our analysis are accurate and precise since we exhaust a large database of lotto players choices of the number combinations maintained by the only lottery operator in Taiwan. There are three main findings in this study. First, the gambler's fallacy temporarily influences players' selection of lotto numbers. The players tend to believe they can improve their chances of winning by avoiding very recent winning numbers. In addition, we find that after controlling for the mechanism of player strategy, the gambler's fallacy is still observed. Second, such negative influence can be partially offset by searching and picking the numbers with the highest observed winning frequencies in the past as this variable feature acts as a signal of above-average performance. Third, the players using the system bet strategy have stronger misconceptions about random processes than the players using the ordinary bet strategy. The first two findings are related to Rabin and Vayanos (2007) model, which states that people judge the performance of a signal depending not only on the luck with reversals, but also on the underlying state with persistence.

Key words: Gambler's Fallacy, Taiwan Lotto, Selection behavior
JEL classification code: D81, G14, L83

1 Introduction

In this paper we aim to model the misconceived belief of lotto players about a random process by studying a large database of Taiwan lotto players' choices of the number combinations. By giving people a low cost opportunity to fulfill their aspirations of becoming millionaires, lotto has come to enjoy worldwide popularity (Statman, 2002). As noted by Thaler (1992), lotto games which have attracted the most attention in wagering markets, are better suited for testing the concepts of rationality than stock markets. It is because they not only share many of the characteristics of traditional capital markets, but also possess the advantage that every bet has a terminal payoff and involves actual money payments (Durham, Hertzel, and Martin, 2005).

In recent years a great deal of attention has been given to the qualitative interpretation of the cognitive bias (Mullainathan, 2002; Rabin, 2002; Rabin and Vayanos, 2005), which is prevalently supported in experimental literature. Insufficient attention, however, has been given to quantitative analysis of wagering markets, especially for the study of lottery play (Rogers, 1998). This is possibly because analyzing the behavior of lottery participants would require exact frequencies of numbers chosen by the players and, unfortunately, lottery operators rarely release that information (Simon, 1999; Papachristou, 2004).

Recent work has attempted to investigate the betting behavior or to estimate the price elasticity of demand for lottery tickets by using only public and limited available data, such as the number of winners for each prize pool and the level of sales (Farrell, Lanot, Hartley, and Walker 2000; Papachristou, 2004). However, such analyses may be less reliable nor accurate and thus limit their acceptance and effectiveness. In the following our study is based on a large scale data set of the number combinations chosen by players. To our knowledge, this is the first conclusive demonstration of the existence of misconceptions about a random processes, based on an actual time series database.

The gambler's fallacy hypothesis has led us to infer that people would underestimate repetition of recent signals from a random binary series. However, people may also expect that such repetitive signals will reappear. This is the well-documented "hot-hand fallacy", in which people expect that the streak should be more likely to continue after a streak of wins than they would after a streak of losses in a basketball shooting game (Gilovich, Vallone, and Tversky, 1985). Rabin and Vayanos (2005) propose a model to reconcile the gambler's fallacy and the seemingly opposite inductions of hot-hand fallacy in the prediction of random sequences. In their model, an individual judges the performance of a signal depending not only on the luck which serves the gambler's fallacy, but also on the underlying state. Under the condition that the state is constant and that people know this, then those people will commit the gambler's fallacy. This is the pervasive finding in unbiased coin flipping experiments. However, under other circumstances, the hot hand could cancel or even overcome the reversal effect of the gambler's fallacy.

The first goal is to investigate whether the Taiwan lotto market shows evidence of the type of misconception as reported in the basketball shooting game. Specially, we aim to test whether there is perceived persistence for popular numbers, Such analyses form a good addition to the relevant literature since despite the fact that Clotfelter and Cook (1993) and Terrell (1994) document the gambler's fallacy in the three-digit numbers lottery game, there is currently no clear evidence provided on the positive bias of hot hand fallacy. Furthermore, we aim to investigate whether such an effect dominates the reversal effect of the gambler's fallacy.

The lotto market in Taiwan was first introduced in 2002, and the enthusiasm for purchasing immediately spread throughout the country. Tickets sales were the tenth largest globally in the first year of operation. One of the explanations to reasons for the lotto's rapid growth and extensive public attention is that it provided the opportunity to allow players to pick their own numbers (Thaler, 1992). By allowing such selections on their own, most people would likely be tricked into believing that they had a better chance to win. Such phenomenon may be explained by the "illusion of control" whereby people exert control over random events (Langer, 1975). In an experiment, Strickland, Lewicki and Katz (1966) find that the subjects in their experiments bet more money and played with more confidence than other people in their chance of winning if they threw the dice themselves. This was so even though the winning probability was equal in these two cases. Our data has the special advantage of allowing us to distinguish players' action by betting types (ordinary bet, system roll, and system bet) in the year 2003 sample period. This leads to the second goal of this paper of exploring on whether the system bet has stronger cognitive bias than the ordinary bet. The system betting type of players who place bets on all combinations of a pre-selected set of numbers are believed to likely take more chances and have more confidence in the numbers they choose.

An alternative explanation to the phenomenon of misconception observed in the results of this research is the wealth effect (Cook and Clotfelter, 1993), which leads those players who regularly choose their own lotto numbers stop playing the game after their selected numbers hit. Simon (1999) documents that there are numerous people who bet their ownership of a set of lucky numbers and around half of UK players select the same numbers in each draw. If players adopt the strategy of betting the same numbers regardless of the winning combinations, and winners stop buying lotto tickets for a certain time span, this behavior would be irrelevant for the misconceived belief. Since most of previous studies on this issue do not distinguish between misconceived belief and wealth effect, the present paper attempts to rule out wealth effect as an interpretation of the outcomes of this investigation.

In this study, we analyze a sample of 1,679,676,226 combinations of lottery ticket numbers consciously chosen by the players themselves in the period from 2002 to 2003. The results show that the players that have won previous lottery draws tend to consciously select number combinations in the following draws. However, in the case of roll-over draws, players not only raise their stakes but also give preference to the computerized random selection method over the conscious method when choosing the combination of ticket numbers.

The results show that number combinations of lottery tickets are not chosen randomly by the players. Moreover, individual numbers that have been drawn in previous lottery draws usually have the lowest picking frequency in the following lottery draws. More precisely, the picking frequency of individual numbers follows a random distribution prior to the announcement day of the winning numbers. Nevertheless, on the day after the announcement day, the picking frequency of the winning numbers announced temporarily decreases 19.5%. This evidence suggests that lotto players do suffer from the gambler's fallacy. Another concern is whether losing players' strategy of betting on the same combination of losing numbers and winning players' decision to withdraw themselves from the market before the next lottery draw will have negative effect on winning numbers selection. In addition, we find that after controlling for the mechanism of player strategy, the gambler's fallacy is still observed; however, we rule out this alternative, as it appears to be a product of misconception

In order to capture the effect of past information available to the players at the point of making their decisions, we regress the frequency of individual numbers on the past winning numbers to investigate whether the players will bet on the numbers with the highest winning frequencies in previous lotto draws. We provide evidence that most lotto participants will pick the numbers with the highest winning frequency in previous lotto draws, as long as these numbers have not been drawn in the most recent draw. Namely, lotto players believe that the numbers that came out most frequently in the past will continue to exhibit a pattern of above-average performance in the future, but underestimate the probability of repetition of the most recent winning numbers even if they have a high historical rate of recurrence. This evidence is consistent with Rabin and Vayanos (2005) model, which states that individuals judge the performance of a fund manager depending not only on luck from which the gambler's fallacy is generated, but also on the latent variable describing the ability of manager (usually call the variable of underlying state).

When we distinguish the combinations into three types of betting strategies, the result shows that the players using the system type of bet strategy have stronger misconceptions than the players using the ordinary bet strategy. The picking frequency declines sharply around the winning numbers on announcement day, especially among system betting types players. While the frequency drops approximately 26.5% on average on the draw immediately after the number hitting day, the frequency for the ordinary bet and system roll fall only 17.1% and 16.2% respectively. The higher degree of misconceptions exhibited by system type of players appear in similar findings in consecutive selections.

Finally, we also illustrate from a new perspective that the gambler's fallacy, which corresponds to the 'representativeness heuristic' described by Tversky and Kahneman (1974), exists among bettors in the short horizon. The people misconceive that there are very close to half heads and half tails in an unbiased coin flipping sequence, therefore, they expect that the next flip will be head if previous throws are a streak of tails (Rabin, 2002). However, such heuristic drive not only results in the misconception regarding random signals series, but also concerning combination selections. For example, most people avoid picking an all consecutive combination (e.g. (1,2,3,4,5,6)), because these combinations do not look random (Simon, 1999). In this paper, we report that the lotto players have a tendency to under-select consecutive numbers in comparison with random selection, because combinations of such preference may look more like random. The result closely resembles the model of Mullainathan (2002), who develops a human inference process in which people divide the posterior space (conditional on a given set of observations) and then choose a category which is most likely. In the similar misconception procedure, the lotto players partition total numbers into 6 categories and are more likely to choose less consecutive combinations.

The rest of this paper is organized as follows: In the next section, we examine the factors that determine of conscious selection. In Section 3, we conduct the tests for gambler's fallacy and extend the misconception about random process. In Section 4, we explore the relation between the betting type and the extent of misconception. Section 5 presents our conclusions.

2 Determinants of conscious selection

Many lotto players prefer to pick numbers by themselves rather than have the computer do it for them. Cook and Clotfelter (1993) referred to the behavior of picking numbers in a non-uniform way as "conscious selection." Such selection procedure is popular because it allows the selectors themselves to feel more in control of the random outcome and hence to have a belief in a higher probability of winning. This phenomenon is called "illusion of control", which is documented by the study of cognitive bias in gambling (Langer, 1975).

Our data consists of the first 203 draws of the Taiwan lotto game for the period of January 2002 to December 2003. There are total 1,679,676,226 combinations of lottery ticket numbers consciously chosen by the players themselves in the sample period. The Figure 1 presents the level of sales (in million tickets) and the proportion of conscious selection. The total demand for lotto tickets had significant sales jumps during the occurrence of either rollovers (double rollovers) or Chinese lunar new years. For example, ticket sales have reached 55 million tickets at the 56^{th} roll over draw, 40 million (37 million) tickets at the 6^{th} (110^{th}) draw during the time period of Chinese lunar new years.

[Insert Figure 1 about here]

The time series pattern of sales for the conscious selection part appears to coincide the total sales' trend. The average portion of conscious selection tickets is about 63%, in range from 45% to 69%. In other words, the average portion of the random combination tickets which are generated by the lottery terminal is 37%. This is slightly larger than the 10%-20% of tickets which are determined by Quick Pick in most lotto games around the world (Simon, 1999).

To examine when players prefer to choose the numbers by themselves or by computer, we start by estimating the following regression:

$$Conscious(t)=\alpha+ \beta_1 Conscious(t-1)+ \beta_2 WC(t-1)+ \beta_3 ROLL\_OVER(t-1)+ \sum_{i=1}^{4}\beta_{4,i}S_i(t)+\varepsilon(t) \quad (1)$$

The dependent variable *Conscious(t)* is the proportion of tickets for which numbers have been chosen by players at draw *t*. Explanatory variables include a lagged dependent variable, the proportion of winners for which numbers had been chosen by players, and a dummy variable for the super-draws, and roll-overs. During the sample period, there are four types of "super-draws" in which the Taipei Bank has guaranteed a minimum size of jackpot. S1 is 1 if the jackpot prize increases to one hundred million dollars. S2 is 1 if the jackpot prize increases sales by 16% as the bonus number is larger than the other winning numbers. S3 is 1 if the jackpot prize increases sales by 16% without any regulative term. S4 is 1 if the jackpot prize is guaranteed for a minimum of one hundred million dollars.

We aim to investigate whether gamblers select tickets by themselves rather than by computer in response to a previous win based on a conscious selection decision. To achieve this, we measure the proxy of the proportion of winners *WC(t-1)* by using three different variables.

(1) *WC_CS* = the proportion of winners from the conscious selection tickets.

(2) *WC_CS_RATIO*= the proportion of winners from the conscious selection tickets/ the proportion of winners from the total tickets

(3) *WC_PEOPLE* = the number of prize-winners from conscious selection tickets/ the number of prize-winners by total tickets

Table 1 reports the results of the determinants of conscious selection. The estimated results show that the coefficient of the lagged dependent variable *CONSCIOUS* is positive and significant at the 1% level. This evidence suggests that the consistency of the conscious selection has a pattern of habit that is persistent over time. The proportion of previous winners is positively significant at the 1% level regardless of the proxy variable we use. This result implies that the players will continue to pick the number by themselves when they have won in the previous draw by choosing the numbers manually. Such behavior is consistent with the Boynton (2003), who show that the subjects seem to apply a win-stay strategy to predict outcomes from a random binary series.

[Insert Table 1 about here]

Bersabe and Arias (2000) investigate the superstitious behavior of gamblers in a dice experiment and document that subjects who previously won more on throwing the dice themselves were more confident of winning when they threw again. Croson and Sundali (2005) examine the existence of heuristic biases by using videotapes of roulette gambling in a casino. The gamblers tended to keep playing and increasing the bets after they had won previously. Similar evidence also suggests that lottery players are more likely to place the bets than to cash the winning tickets (Clotfelter and Cook, 1989).

The negative and significant coefficients on *ROLL_OVER* indicate that sales for conscious selection are not raised as much as for random selection, though there is an impressive rise of sales due to a previous unwon jackpot. The four dummy variables of super-draw are not significant. The promotions have no effect on the portion of conscious selection. The result may be because that players are less susceptible to super-draw money than to jackpot prize by rollover (Forrest, Simmons, and Chesters, 2002)

3 Methodology

3.1 Average probability of individual numbers

Figure 2 presents the frequency of individual numbers picked by the Taiwan lotto players. The average mean of picking frequency are close to the theoretical average (1/42). By contrast, their maxima or minima fluctuate a great deal among the 42 numbers. For example, the maximum frequency for the 25th number is (0.0477) at least six-fold than the minimum frequency for the 41st number (0.007603). Furthermore, we find some relation between the past winning information and the betting frequency by investigating when the minimum frequency occurs. Except for the 40th number, all the minimum frequencies take place right after the corresponding numbers hit in the previous draw. On the other hand, the maximum frequencies are observed when their numbers have not been hit during the period of three consecutive previous draws in 41 of 42 cases. This result suggests that lotto players do not pick their numbers randomly and the players are negatively influenced by the recent winning umbers. In the next sections we investigate in more detail how players select their numbers relative to the number of hitting.

[Insert Figure 2 about here]

3.2 Reaction to hits of winning numbers

Figure 3 represents the reaction of the average picking frequency (ranking order) to the winning numbers hit. We plot the average picking frequency separately for the numbers won or lost, investigating whether or not there exists a significant impact on the picking frequency due to a winner announcement. Therefore, the frequency is partitioned by two groups depending on whether the balls win or lose. The picking ranking order frequency represents the ordered rank from 1 (least picking frequency) to 42 (most picking frequency).

[Insert Figure 3 about here]

If lotto players choose their numbers randomly, we would expect the average picking frequency (ranking order) to be 0.0238 (21.5 in ranking order), regardless of the winning numbers that have been drawn in the past. Notice in Figure 3 that the winner announcement leads to a significant frequency (ranking order) drop. The picking frequency on winner numbers drops immediately from 0.0237 (21.2 in ranking order) on the hitting draw to 0.0191 (10.4 ranking order) after the first day of draw, and recovers to its previous level over 4 draws. Thus, the frequency of winner numbers seems slightly higher than loser numbers over the long window of 7 days.

The temporal frequency drop of 0.00461 (19.5 percent of the previous draw) on the day after the numbers' hitting day provides a piece of evidence that lotto players do suffer from the gambler's fallacy. This is consistent with the picking three-digit numbers game studied by Clotfelter and Cook (1993) and Terrell (1994). There is a decrease in the amount of bets on a number after it is drawn. Then gradually the bets return back to a normal level over 2 months. Not surprisingly, the average lotto frequency drop is temporary but recovers more quickly than the three-digit numbers game. One explanation to this is that the rules of the lotto game are slightly different from that of picking a 3-digit number.

The probability that the same daily 3-digit number repeats is lower than that of an individual winning number for the lotto game. A player will expect that a of the repeat 3-digit winning number should be an average of 1000 days if they play a straight bet.[1] In contrast, they will believe that the individual number repeat will be only 6 draws for the lotto game.

It has been shown that lotto players would underestimate repetition of winner numbers. The gambler's fallacy may have negative influence on selection behavior. However, other anecdotal evidence regarding gamblers' betting behavior is based on the erroneous belief that machines are not necessarily perfectly calibrated. People

1 There are two principal types of bets, straight bets and box bets. The probability of a straight bet is 1 in 1000. The box type of bet wins a prize if any permutation of the number is drawn. The probability of a box bet is 6 in 1000

record the outcomes of the machines for some time and then bet on the favored numbers with the highest observed frequencies (Keren and Lewis, 1994). We would provide evidence that not only players' underestimating repetition of recent winner numbers, but also their overestimating of the numbers with the highest observed winning frequencies in the past.

To see how the betting behavior is influenced by the belief, we analyze the player's reaction to hits of winning numbers but sorting the average picking frequencies according to their observed winning frequencies in the past. The ranked frequencies are assigned to 6 quintiles group by their winning frequencies within the last 99 draws. The highest quintile group is called the "Hot winner" group and the lowest quintile group is called the "Cold winner" group.

As figure 4 shows, after the winner numbers announcement draw, these frequencies group remain to have a significant drop. However, before the announcement draw, the average picking frequency for the Hot winner group is higher than the Cold winner group. For example, at the draw of announcement, the average picking frequency for the Cold winner group 0.021177 as compared with the Hot winner group 0.026772 is 0.005595 (26.4%). The players have the belief that the number has a greater chance to win if it appears at a higher frequently than other numbers.

[Insert Figure 4 about here]

3.3 Alternative interpretation of gambler's fallacy phenomenon:

One of the reasons why players might fall into the gambler's fallacy is that they continue to bet and choose the same combination in each draw until they win a prize. Clotfelter and Cook (1993) mention that since people play the lottery only with the purpose of achieving specific financial goals, they may be motivated to decrease their bet in winning numbers, and stop betting when their selected numbers win. For example in Table 2, suppose a lotto player regularly selects the same ticket combination which includes the individual number i. If results are announced at draw t-1 and number i is one of the winning numbers, then the player will be a winner if the ticket combination matches at least two more winning numbers. On the other hand, if the ticket holder does not match at least two more winning numbers in addition to number i, he will lose the bet. If the loser continues to bet on the same ticket and the winner retires from the market in the next draw, there will be a significant reduction of the picking frequency of individual number i at draw t. Under such a strategy, it can be proven that the phenomenon of gambler's fallacy is irrelevant to the consequences brought by misconception.

[Insert Table 2 about here]

To measure the impact of playing strategy on selection behavior, the probability distribution of the numbers chosen by players divide into two parts : winning tickets which match at least two winning numbers and losing tickets which do not earn any prize. The variable $Q(i,t)$ is the probability distribution of numbers chosen by lotto players for the individual numbers i at draw t. The variable *Q_winner* is used to compute the distribution of the winning tickets. The variable *Q_loser(i,t)* is used to compute the distribution of the tickets which do not earn any lotto prizes. If players choose distribution uniformly, we would expect the *Q_winner* (winning tickets) and *Q_loser* (losing tickets) to be 0.002474, 0.021335 respectively.

$Q(i,t)=Q\_winer(i,t)+Q\_loser(i,t)$

If lotto players adopt such a betting strategy and no. i is listed in the winning numbers at draw t-1, we would expect the difference between $Q(i,t\text{-}1)$ and $Q(i,t)$ to be positive. Gambler's fallacy can be induced by using this strategy as the winning ticket holders at draw *t*-1 stop betting on number combinations covering no. *i* at draw *t*. However, if we control for the strategy effect by deducting *Q_win(i,t-1)* from the difference, but still find that values are positive, the gambler's fallacy can be interpreted not as a consequence of the wining players strategy to withdraw themselves from the market, but as a product of the phenomenon of misconception.

We use a t-test to test whether the mean for *Q(i,t-1)-Q(i,t)-Q_win(i,t-1)* is significantly less than zero. The t-test for mean is 0.002705786 with a one-sided p-value of 0.0000 (t= 23.40). Because the statistic has a significantly positive value, this t-test suggests that when we control for mechanism of player strategy, the winner numbers announcement leads to a significant frequency drop, and therefore the tendency to under-predict repetition of recent winning numbers is still observed. This result suggests that the mechanism of player strategy cannot completely explain the behavior of gambler's fallacy.

3.5 A dynamic model

To investigate whether the past winning numbers have impact on the subsequent demand for these numbers, various regression models on the distribution of numbers picked at each draw are proposed are proposed below. We focus on the dynamic model for frequencies of numbers selected with lagged dependent variables, while taking into account of other static models for robustness check.

Model 1: $Q(i,t) = \alpha + \beta_1 HIT(i,t) + \beta_2 HOT(i,t) + \varepsilon(t)$ (2)

Model 2: $Q(i,t) = \alpha + \beta_1 Q(i,t-1) + \beta_2 HIT(i,t) + \beta_3 HOT(i,t) + \varepsilon(t)$ (3)

Model 3: $Q(i,t) = \alpha + \beta_1 Q(i,t-1) + \beta_2 HIT\_DOUBLE(i,t) + \beta_3 HOT(i,t) + \varepsilon(t)$ (4)

Model 4: $Q(i,t) = \alpha + \beta_1 Q(i,t-1) + \beta_2 HIT(i,t) + \beta_3 HOT(i,t) + \beta_4 DAY(i,t)$

$$+ \beta_5 BALL\ (i,t) + \varepsilon(t) \qquad (5)$$

The dependent variable *Q(i,t)* is the probability distribution of numbers chosen by lotto players for the individual numbers *i* at draw *t*. A set of explanatory variables are introduced as follows: *HIT* is -1 if the number hits at a previous draw, and 0 otherwise. *HOT* is 1 if the number listed in the top 7 winning numbers within the last 99 draws, and 0 otherwise.[2] Next, to know whether the participants' tendency of preferring winning numbers increases as the numbers hits twice in a row , we use the dummy variable *HIT_DOUBLE* and set it to be -1 if the number hits both at draw t-1 and at draw t-2, and 0 otherwise.

The variable *DAY* counts the number of drawings which have taken place since the number was last selected. This variable is designed to capture the effect of "overdue" where the players may be in favor of certain numbers that have been drawn less frequently. Many lotto participations are reluctant to select higher numbers and prefer some special significant numbers, such a s family birthdays (Haigh, 1995; Hill and Williamson, 1998). We use the variable *BALL* which is i to study people who pick numbers from their own low numbers set (for example birthday numbers).

We regress the probability distribution of individual numbers on the lagged dependent variable, dummies frequency of occurrence of the numbers in the past draws, and the characteristics of the numbers. From Model 1 in Table 3, which is Model 2 above, the coefficient *HIT* is positive and is significant at the 1% level, which suggests that lotto players prefer not to pick numbers that have occurred in the last draws. This kind of behavior in picking the numbers is consistent with representativeness heuristics that leads to the gambler's fallacy hypothesis (Tversky and Kahneman, 1974). Underlying the fallacy is the implicit assumption that winning

[2] Whether we define the dummy variable HOT as 1 if the number lists in the most popular top 7 winning numbers within the last 30, 50, or 150 draws, the results are similar.

numbers come from a random draw by a lotto machine, implying that each ball is expected to be drawn at an equal number of frequencies, even in the short segment (they follow uniform distribution and the probability for each ball equals 1/42). The result shows evidence of underestimated repetition of recent winning numbers for the lotto players.

[Insert Table 3 about here]

The dummy variable Hot is constructed to capture whether there is perceived persistence for popular numbers. The coefficient on HOT is all positive significantly at the 1% level. This shows that the players have the belief that the number has a greater chance to win if it appears at a higher frequently than other numbers. This evidence is consistent with the notion modeled in Rabin and Vayanos (2005) that people judge the performance of a signal depending on not only the luck which serves the gambler's fallacy, but also on the underlying state. Players in a lotto game believe not only in lucky winning numbers, but also that past favorable numbers will persist excessively for future draws under the assumption that the lotto machine is biased. Together the coefficients in Model 1 suggest that the information on the past history of winning numbers do indeed have a significant impact on subsequent demand for these numbers, though the adjusted R-square is lower and the residual is highly correlated.

Model 3 (i.e., the Model 2 in Table 3) is a lagged dependent variables model (LDV) or a dynamic regression model which includes a lagged dependent variable Q. Keele and Kelly (2006) suggest that the LDV model is often used as a means of capturing dynamic effects and that the inclusion of a lagged dependent variable is capable of eliminating residual serial correlation when autocorrelation is detected in the OLS regression. But the estimated coefficients will be biased or even inconsistent and it is no longer valid to use the Durbin Watson statistic in the presence of serially correlated disturbances in the regression model with a lagged dependent variable. Therefore, we use Breusch-Godfrey Lagrange multiplier (LM) to test for correlation in a LDV model. The result supports rejection of no first-autocorrelation at 5 percent level and the explanatory power of the regressions increase from below 20% to 67% in MODLE 2. The coefficient of the lagged depend variable *Q* is positively significant at the 1% level. This evidence indicates that popular numbers persist over time, which is consistent with previous findings reported in 3-digit lottery game (Clotfelter and Cook, 1993; Terrell, 1994) that some numbers are bet more than others.

The coefficient for *HIT* in MODEL 2 of Table 2 is 0.0055, significant at the 1% level. This estimated dummy variable implies that the probability of player's picking individual numbers shows a sudden drop by 23.3% of the uniform individual distribution when the number was hit in the previous draw. The estimated coefficient of *HOT* is 0.0019 and is less than *HIT*'s. This could suggest that the reversal effect induced by the gambler's fallacy dominates over the excessive persistence of the most frequent "hot" numbers. The difference in the estimated coefficient between *HIT* and *HOT* is 0.0036. This magnitude means that the probability of individual numbers chosen by players is decreased only by the 15.5% of the uniform individual distribution if the number won in a previous draw, and also has occurrence rates higher than other numbers over a long horizon.

While most of the estimated coefficients are significant and consistent with what we noted earlier, we need to be careful about making conclusions in light of the fact that the LM Test has rejected the residual serial correlation in MODEL 4 (Model 3 of Table 2) and MODEL 5 (Model 4 of Table 2). The coefficient *DAY* is significant and indicates that the probability of player's picking individual numbers increases by

0.01% per day after the number wins. Together the coefficients suggest that the gambler's fallacy plays a significant role in the betting behavior of gamblers.

3.4 Nonconsecutive combinations

The representativeness heuristic describes the biases of decision when people judge the probability of an event by how similar the event is to the population from which it is drawn. For example, which outcome is more likely when flipping an unbiased coin 6 times, HHHTTT or THTTHT ? Many players tend to pick the second outcome because it looks more "random", even though both outcomes are equally likely to happen. Another textbook example of the heuristic is which ticket combination is more likely to win the first prize, (1,2,3,4,5,6) or (3,16,17,29,34,37) ? Most people prefer the second ticket because winner numbers come from a random machine and 6 consecutive numbers do not appear to be sampled from a random process.

We now introduce for each lottery combination a random variable that measures the degree of continuity of the six numbers chosen from 1 to 42. Let X be the set of all possible combinations of Taiwan (6/42) lotto tickets, i.e.,

$$X=\{\ (t_1,\ldots,t_6):\ 1 \le t_1 < \ldots < t_6 \le 42\ \},$$

and define the variable *JUMP* on X by:

$$JUMP(t_1,\ldots,t_6) = 5-\sum_{i=1}^{5} I_{(t_{i+1}-t_i)}\ ;\ \text{where}\ \ I_{(t_{i+1}-t_i)}=1 \text{ if } t_{i+1}-t_i=1,\ \text{i}=1,\ldots,5, \text{ and 0, otherwise.} \tag{6}$$

For each combination the variable *JUMP* is the counts of incidences in which two neighboring integers (in monotone order) chosen are nonconsecutive. For example the *JUMP* is 0 for the ticket combination (1,2,3,4,5,6), all consecutive numbers, and is 5 for the combination (1,7,11,16,29,41), no consecutive numbers.

Table 4 lists the average of empirical means of *JUMP* for the winning numbers and combinations chosen by players in various sub-samples. We use a one-sample t-test to test whether the average mean for *JUMP* variable is different from the expected mean by random selection, 4.285714. Table 3 shows that the mean of jump for the first 203 winner numbers drawn are not significant, while the number combinations chosen are positive and significant (t=2287.64) in all sample periods. This result indicates that players tend to pick less consecutive numbers than random selection and further arguments may be required to explain such preference.

[Insert Table 4 about here]

Mullainathan (2002) develops a model of human inference in which people make predictions by using categorical thinking rather than Bayesian process. In his model, initially the categorical thinkers form partition of the posterior space and then choose the category which is most likely given the observations. The concept of "categorical thinker" is promising of explaining our finding that lotto players tend to pick numbers that are less consecutive. The players make such choice in an optimal way. The players make theirs choices of numbers for a category that has a larger probability to occur. In other words, combinations with not less than 4 jumps are more likely to be chosen since the popular mean of the variable JUMP is greater than 4.[3]

[3] The theoretical proportion for various *JUMP* categories are calculated as follows:

Jump	0	1	2	3	4	5	total
No. of bet combinations	37	3330	77700	660450	2179485	2324784	5245786

4 Misconception across betting types

4.1 Reaction of average picking frequency across betting types

Previous sections have for the most part analyzed the impact of past information on the betting decision. In this section, we attempt to explore the relation between betting type and the extent of misconceptions about random processes.

Starting at the 100th draw (the first draw of year 2003), players in the Taiwan lotto game were allowed to place three types of bets on the number drawn: ordinary bet, system roll, and system bet. Table 5 lists the three betting types and their betting amounts. For an ordinary bet the players simply select 6 numbers from 1 to 42. The system roll refers to a selection of only 5 numbers and the computer would assign each of the remaining 37 numbers to these 5 numbers to form 37 combinations. The system bet allows the players to pick 7 to 16 numbers from 1 to 42. For example, the participants would place 28 ordinary bet combinations if choosing a system 8.

[Insert Table 5 about here]

The descriptive statistics for the three types of bets presented Table 6 are based on 104 draws for the year 2003 sample period. Most of the players appeared to use ordinary bet to pick the numbers. The average proportions of the ordinary bets during the sample period is about 80%. The average proportions of the system bets decrease as their corresponding numbers of combinations increase.

[Insert Table 6 about here]

% bet	0.000705	0.063480	1.481189	12.590106	41.547349	44.317172	100

Those lotto players who place system bet picking more than six numbers tend to take chances and seem to have more confidence than players choosing ordinary bet for one draw. Strickland, Lewicki and Katz (1966) find that the subjects in their experiments bet more money and played with more confidence than other people in their chance of winning if they threw the dice themselves. This was so even though the winning probability was equal in these two cases. This phenomenon may be explained by the "illusion of control" whereby people exert control over random events (Langer, 1975). For example, allowing lotto players to choose their own numbers, appear to trick them into believing that they had a better chance to win. It is worth pointing out that there had not been explosive growth in popularity and sales of lotto tickets in North America until New Jersey provided the opportunity to allow players to pick their own numbers (Thaler, 1992).

There are 10 different scales of system bets (as shown in Table 4), ranging from sys 7 to sys 16. To summarize, we divided the 10 scales into 3 groups depending on the numbers of combination, 100 below, 101-500, and 501 above. The categories represent Small, Medium, and Large type system bet players.

Figure 5 represents a time series plot of the average picking frequency (ranking order) of the winning numbers for various betting types, and shows that the frequency (ranking order) declines sharply around the numbers draw day, especially among system betting type players. The frequency drops at least 26.5% in average on the day immediately after the number hitting day for the three types of system bet. In contrast, the frequency for the ordinary bet and system roll fall only 17.1% and 16.2% respectively. This indicates that the players who pick their numbers by system bet strategy may believe more in that their luck would reverse than those who bet single ticket per draw. The players who use the system roll strategy experience less frequency drop than the others. This may be because the computer assigns the remaining numbers when the players have selected their 5 numbers by system roll. As such, the operation could reduce the effect of gambler's fallacy on selection behavior.

[Insert Figure 5 about here]

4.2 Dynamic model across betting types

In order to study whether the players of system bet have stronger misconception than those who use the ordinary bet strategy, we use the following regression analysis to test for the significance of the differences.

$$Q(i,t)=\alpha+ \beta_1 Q(i,t-1)+ \beta_2 HIT(i,t) + \beta_3 HOT(i,t)+ \beta_4 HIT\_NOT\_P6\ (i,t)) + \beta_5 HOT\_NOT\_P6(i,t)+ \varepsilon(t) \quad (7)$$

We pool the frequency for ordinary bet data and other type of betting data and obtain the coefficient of independent variable in Table 7. We adopt the explanatory variables similar to the previous section, but include two more dummy variables. The dummy variable *HIT_NOT_P6* is -1 if the number hit at the previous draw and the observation does not come from an ordinary bet. The dummy variable *HOT_NOT_P6* is -1 if the number lists in the most popular top 7 winning numbers within the last 99 draws and the observation does not come from an ordinary bet.

[Please insert Table 7 about here]

The coefficients *HIT_NOT_P6* in Table 7 are all negative and significant at 1% for the three different sizes of system bet, but are not significant for the system roll.

This implies that players who use the system bets are more misconceived than those using the other types of betting strategies, and suggests that people who bet more money and have more confidence in the chance of winning tend to be more subject to the gambler's fallacy. For example, the OLS estimates of the coefficient implies that the frequency drops 0.0052 on the day after the number hitting day for the ordinary system bet, which is not as strong as for the small system bet. This indicates that, with other factors being equal, the expected frequency drop for the small system bet is about 46% (0.0024/0.0052) more than for the ordinary bet.

The coefficients *HOT_NOT_P6* in Table 6 are positive and significant at 1% for the small system bet, significant at 10% for the medium system bet, but not significant for the large system bet. However, the system bet for the fourth column is still significant at 1%. This implies that the expected frequency for the system bet is about 44% (0.0007/0.0016) more than for the ordinary bet. The two coefficients for the *HIT_NOT_P6* and *HOT_NOT_P6* are not as significant as we expected. Overall, the OLS results are consistent with the notion that players who bet more money or have more confidence in their numbers commit stronger degree of "illusion of control."

4.3 Nonconsecutive combinations for system bets

The previous sections document that the players who use system bet strategy tend to base their selection of numbers on past winner numbers and commit more cognitive bias than ordinary bet type of players. Is the difference in degree of cognitive bias also reflected with significance by using the *JUMP* variable? We use the two-sample t-test to test whether the mean of *JUMP* for system bet and system roll are different with ordinary bet. The result shows that the mean for system bet and roll is significant 0.23 percent (t=2.21) higher than those using ordinary bet. This is consistent with previous finding that players using system bet strategy have more cognitive bias than those who use ordinary bet.

5 Conclusions

The results established in this paper provide clear and broad-based evidence that Taiwan lotto players do not pick their numbers randomly and that the gambler's fallacy temporarily influences their betting behavior. However, such negative influence is less pronounced with respect to past winning numbers. While the frequency with which players pick the winning numbers drops by 23.3 percent immediately after the hitting day, concurrently the frequency of numbers with higher occurrence rates decreases by only 15.5 percent.

Players might under-predict the repetition of recent winning numbers not only because they suffer from a misconception regarding random processes, but also because they who might adopt winner stop-playing-strategy. In this paper, we attempt to analyze selection behavior by controlling the impact of winning players' stop-playing-game strategy effect. The results show that lotto players exhibit a significantly and temporarily decrease in betting winning numbers on the day after the announcement day. However, there is still a consistent reduction of the picking frequency of individual numbers, even when after control for player strategy. The evidence that players might fall into the gambler's fallacy may be a product of misconception instead of the winning players' stop-playing-game strategy.

From this, we can extend the interpretation of the concept of representativeness heuristic, which describes the misconceived beliefs of people regarding random processes, to address the erroneous perception on which lotto participants based the selection of their ticket numbers combinations. As the results show, most players avoid picking consecutive numbers because such number combinations would seem less likely randomly distributed, and therefore would not resemble the distribution of the numbers appearing in past lotto draws.

The results of this study are related to the model of Rabin and Vayanos (2005), who state that people judge the performance of a signal depending not only on a luck factor but also on the persistence of an underlying state. In their model, individuals fall into the gambler's fallacy while knowing that the conditions that affect the outcomes are constant.

However, under other circumstances, the excessive persistence could cancel or even overcome the reversal effect of the gambler’s fallacy. In our study, the players form their expectations by searching for winning numbers with the highest observed frequencies attempting to pick those numbers with above-average performance. This selection procedure captures the idea of a persistent underlying state.

Our analysis also shows a further relation between the betting strategy and the extent of misconceptions. Gamblers, who bet more money on lottery systems other than the ordinary systems, have a higher level of confidence on their chosen numbers. The results also indicate that cognitive bias has a more significant impact on the system bet participants, especially on the players of small system bets.

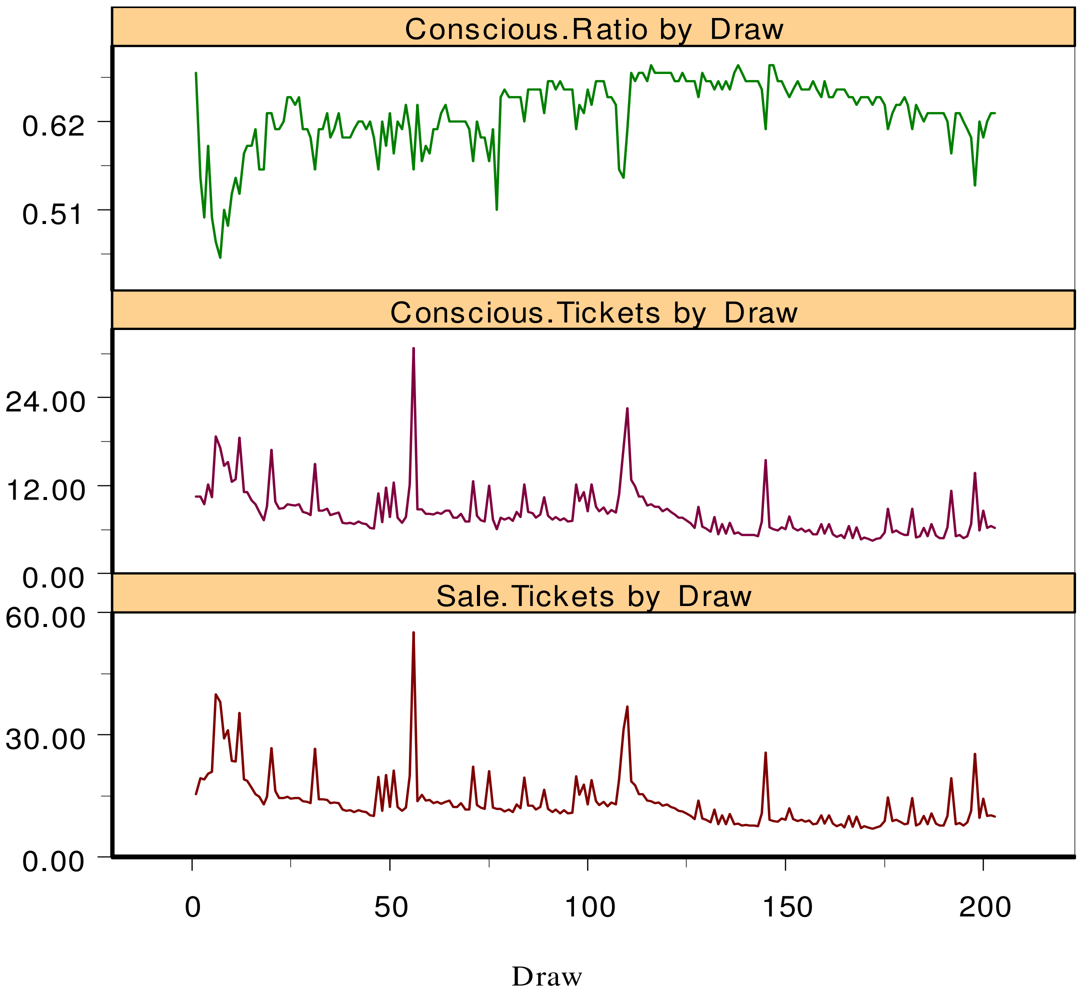


Figure 1 The time series pattern of the proportion of numbers consciously chosen by lotto players

Table 1 Determinants of the ratio of conscious number selection

$$Conscious(t)=\alpha+\beta_1 Conscious(t\text{-}1)+\beta_2 WC(t\text{-}1)+\beta_3 ROLL\_OVER(t\text{-}1)+\sum_{i=1}^{4}\beta_{4,i}S_i(t)+\varepsilon(t)$$

The dependent variable is the number of tickets chosen divided by the total number of tickets sold. Explanatory variables include lagged dependent variable, and dummies variable for the super-draws, and roll over. *ROLL_OVER* is 1 if the prize of jackpot is not won and 0 otherwise. During the sample period, there are four types of "super-draws" in which the Taipei Bank guaranteed a minimum jackpot prize. S1 is 1 if the prize of jackpot increased in one hundred million dollars. S2 is 1 if the prize of jackpot increased sales by 16% due to the bonus number being higher than the remaining winning numbers. S3 is 1 only if the prize of jackpot increased sales by 16%. S4 is 1 if the lotto draw guaranteed a minimum jackpot prize of one hundred million dollars. We measure the proxy of the proportion of previous winners *WC(t-1)* by using three different variables: (1) *WC_CS* = the proportion of winners from the conscious selection tickets (Model 1), (2) *WC_CS_RATIO*= the proportion of winners from the conscious selection tickets/ the proportion of winners from the total tickets (Model 2), and (3) *WC_PEOPLE* = the number of prize-winners from conscious selection tickets/ the number of prize-winners by total tickets.(Model 3) Significance levels of 0.01, 0.05 and 0.1 are denoted by ***, **, and *, respectively.

VARIABLE	MODEL 1		MODEL 2		MODEL 3	
Variable	Coefficient	t-Statistic	Coefficient	t-Statistic	Coefficient	t-Statistic
Intercept	0.1223***	4.48	0.0645**	1.99	-0.0387	-0.81
conscious (t-1)	0.7692***	18.48	0.7650***	18.63	0.9245***	17.49
WC_CS	0.0095***	3.76				
WC_CS_RATIO			0.0892***	4.47		
WC_PEOPLE					0.0576***	4.7
ROLL OVER	-0.0288***	-5.95	-0.0282***	-5.91	-0.0280***	-5.9
S1	0.0006	0.04	-0.0009	-0.05	-0.0005	-0.03
S2	0.0003	0.03	-0.0003	-0.03	-0.0003	-0.03
S3	0.0011	0.06	-0.0004	-0.02	-0.0004	-0.02
S4	0.0018	0.30	0.0016	0.28	0.0016	0.29
Adjusted R-squared	0.6644		0.6736		0.6768	
Breusch-Godfrey LM Test	1.3244		0.4694		0.1570	
Draws	202		202		202	

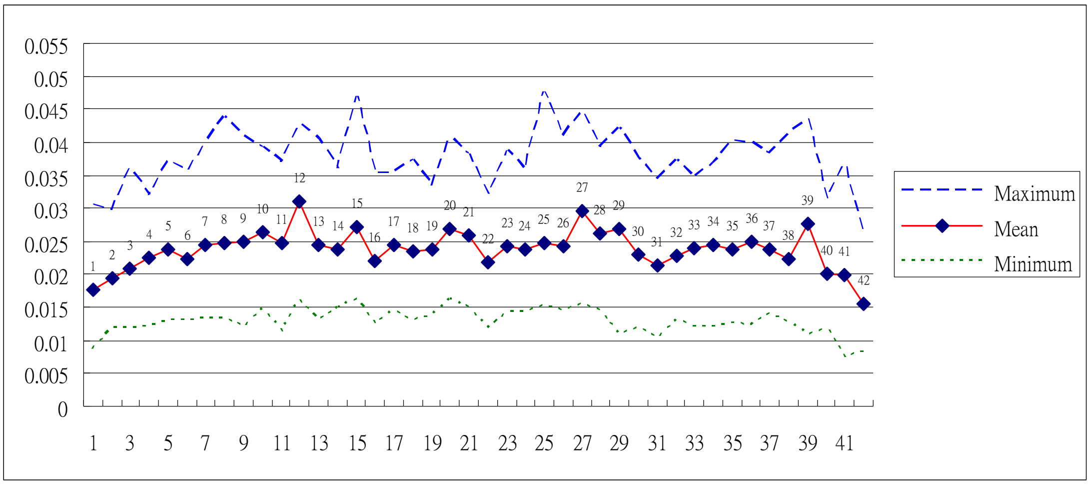
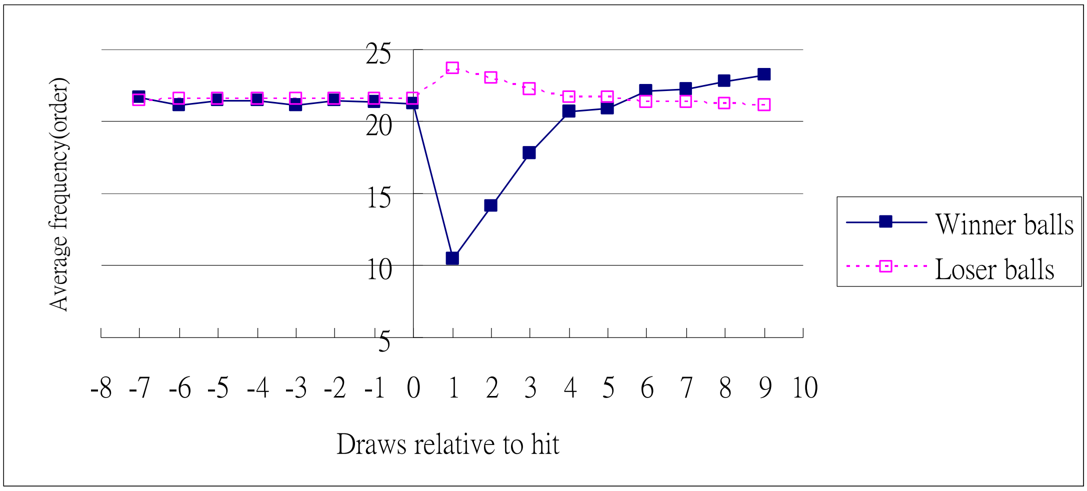


Figure 2 The average probability of picking individual numbers



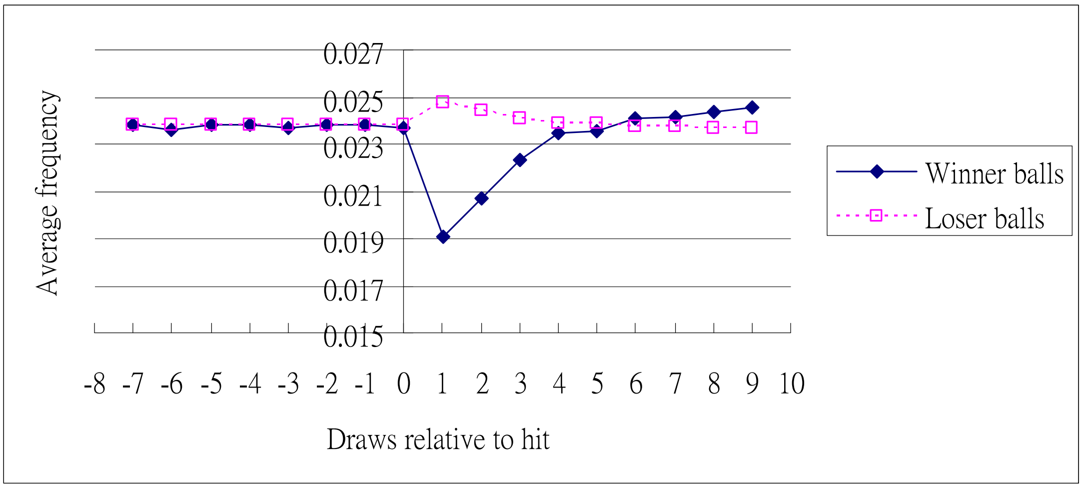
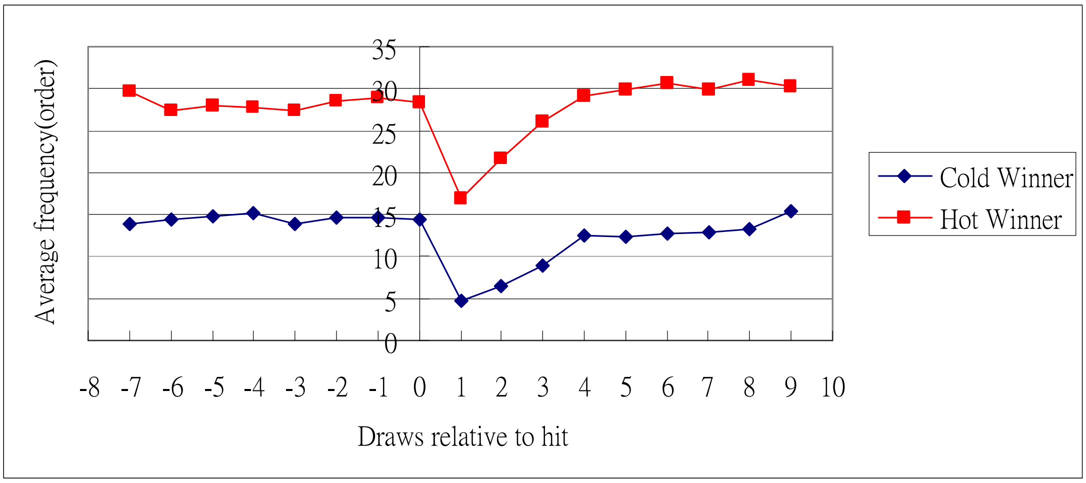


Figure 3 Picking frequencies for reaction to hit.

The figure presents the reaction of the average picking frequency (ranking order) to hit of winning. The picking frequencies are ranked from 1 (the lowest) to 42 (the highest). The frequencies are partitioned into two groups depending on whether the balls were drawn (winner) or not (loser). The winner group (solid line) contains 1309 numbers drawn in 2002-2003 sample period. Day 0 (the hitting day) is the day the number is drawn.



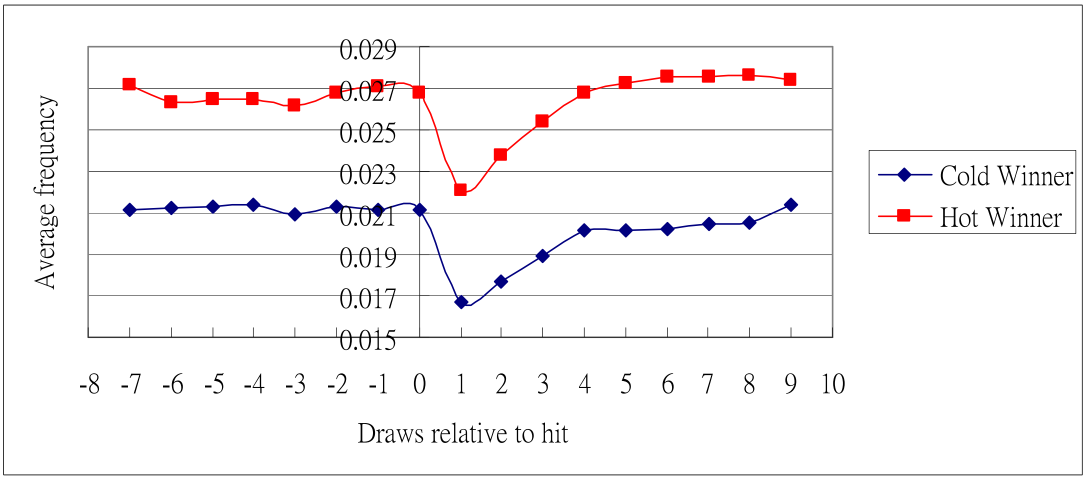


Figure 4 Picking frequencies for reaction to hit across winning frequencies.

The figure presents the reaction of the average picking frequency (ranking order) to hit of winning but sorting the average picking frequencies according to their observed winning frequencies in the past. The picking frequencies are ranked from 1 (the lowest) to 42 (the highest). The ranked frequencies are assigned to 6 quintiles group by their winning frequencies within the last 99 draws. The highest quintile group is called the "Hot winner" group and the lowest quintile group is called the "Cold winner" group. Day 0 (the hitting day) is the day the number is drawn.

Table 2. Playing strategy for lotto tickets covering no. *i*

Time	Player bet tickets combinations	
t-1 draw	Player bet tickets covering no. *i*	
t-1 draw	Lotto operator announces the draw result and no. *i* is one of the winning numbers	
t-1 draw	Winner tickets covering no. *i*	Loser tickets covering no. *i*
t draw	Out of Market	Loser still bet on the same tickets covering no. *i*

Table 3 Determinants of the probability of the numbers picked by the players

The dependent variable is the probability distribution of numbers chosen by lotto players, *Q(i,t)*, for the individual numbers i at draw t. Explanatory variables include lagged dependent variable *Q(i, t-1)* and dummies for the frequency of occurrence of numbers in the past draws. *HIT* is -1 if the number i hit at draw t-1, and 0 otherwise. *HIT_DOUBLE* is -1 if the number i hit at draw t-1 and draw t-2, and 0 otherwise. *HOT* is 1 if the number i is amongst the top 7 winning numbers within the last 99 draws, and 0 otherwise. The variable *DAY* counts the number of drawings taken place since the last time when the number was selected. *BALL* is i if the number is i. Significance levels of 0.01, 0.05 and 0.1 are denoted by ***, **, and *, respectively.

	Model 1		**Model 2**		**Model 3**		**Model 4**	
Coefficient	Coefficient	T-Value	Coefficient	T-Value	Coefficient	T-Value	Coefficient	T-Value
Intercept	0.0240***	326.94	0.0092***	41.74	0.0086***	31.54	0.0094***	39.99
Q(i,t-1)			0.6390***	69.2	0.6295***	54.68	0.6140***	59.5
HIT	0.0058***	34.97	0.0055***	48.32			0.0053***	41.42
HIT_DOUBLE					0.0034***	11.17		
HOT	0.0043***	27.46	0.0019***	16.18	0.0016***	11.16	0.0021***	16.85
DAY							0.0001***	5.37
BALL							0.0000	-0.18
Adjusted R-squared	0.2982		0.6670		0.5015		0.6691	
Breusch-Godfrey LM Test	122.7774		3.6178		55.4586		5.2050	
Draws	103		103		103		103	

Table 4 Descriptive statistics of *JUMP*

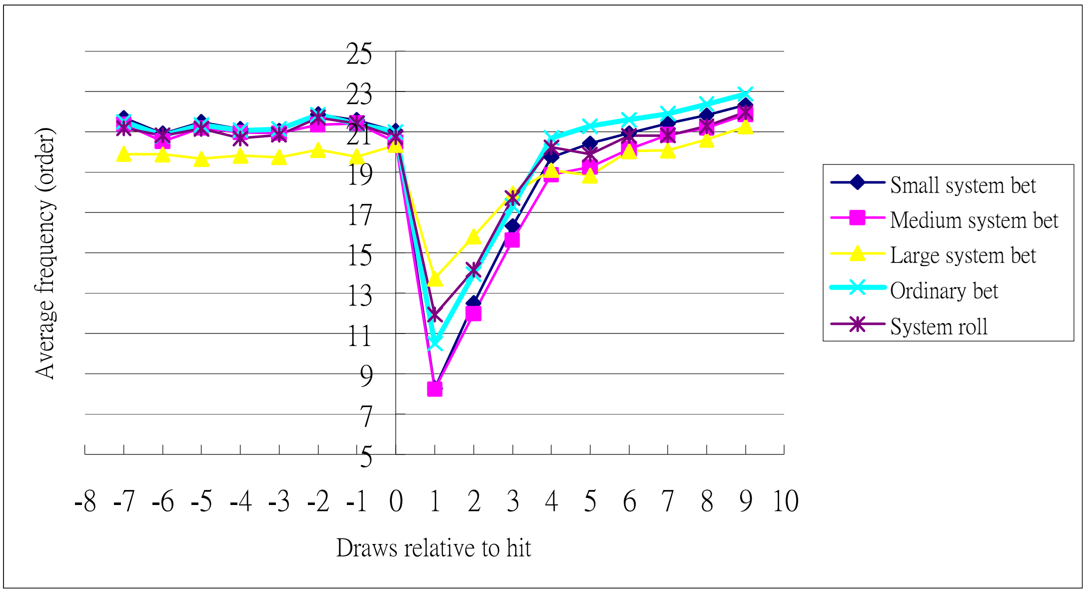
	MIN	MAX	MEAN	T-STATISTICS	OBERSTATIONS
Winning numbers	2	5	4.3300	0.85	203
91 JUMP	4.1463	4.5134	4.3081***	930.97	99
92 JUMP	4.2015	4.4042	4.3514 ***	2287.77	104
ALL JUMP	4.1463	4.5134	4.3303 ***	2287.64	203

Table 5 Three types of betting and their amounts

Bet type	No. of Ordinary Bet combinations	Amount (in NT$)	Bet type	No. of Ordinary Bet combinations	Amount (in NT$)
Ordinary	1	50	Sys 12	924	46,200
Sys 7	7	350	Sys 13	1716	85,800
Sys 8	28	1,400	Sys 14	3003	150,150
Sys 9	84	4,200	Sys 15	5005	250,250
Sys 10	210	10,500	Sys 16	8008	400,400
Sys 11	462	23,100	Sys Roll	37	1,850

Table 6 Descriptive statistics for betting types

Bet type	Min	Mean	Max	Std Dev.	Skewness	Kurtosis
Sys Roll	0.363	0.471	0.808	0.096	2.010	3.834
Ordinary	72.519	79.695	86.173	1.813	0.434	4.339
Sys 7	5.891	11.837	15.968	1.742	-1.763	4.177
Sys 8	4.547	5.099	6.720	0.396	1.548	3.122
Sys 9	1.078	1.399	2.096	0.214	1.532	2.409
Sys 10	0.703	1.018	1.901	0.238	1.676	2.997
Sys 11	0.077	0.194	0.392	0.056	0.705	0.759
Sys 12	0.018	0.155	0.354	0.065	0.483	0.243
Sys 13	0.000	0.052	0.233	0.039	1.562	4.845
Sys 14	0.000	0.037	0.157	0.041	0.904	0.083
Sys 15	0.000	0.020	0.207	0.040	2.095	4.588
Sys 16	0.000	0.022	0.374	0.071	3.496	11.764



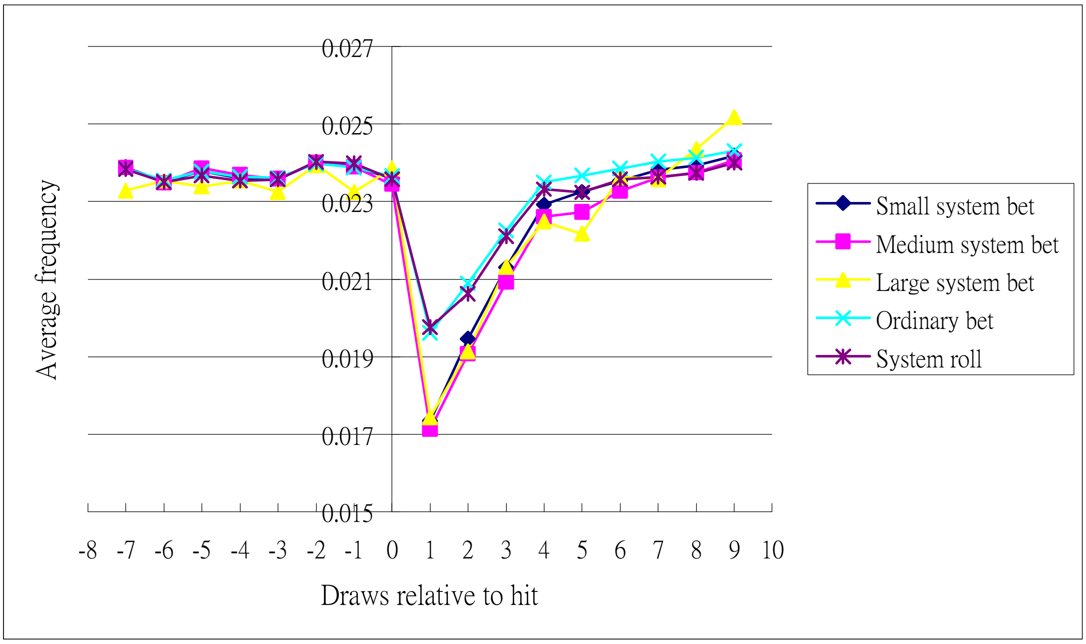


Figure 5 Picking frequencies for reaction to hit across betting types.

The figure presents the reaction of the average picking frequency (ranking order) to hit of winning. The picking frequencies are ranked from 1 (the lowest) to 42 (the highest). To summarize, we divided the 10 different scales of system bets into 3 groups depending on the number of combination, 100 below, 101-500, and 501 above. The categories represent Small, Medium, and Large type system bet players. Day 0 (the hitting day) is the day the number is drawn.

Table 7 Determinants of the probability of the numbers picked across betting types

$$Q(i,t) = \alpha + \beta_1 Q(i,t\text{-}1) + \beta_2 HIT(i,t) + \beta_3 HOT(i,t) + \beta_4 HIT\_NOT\_P6(i,t)) + \beta_5 HOT\_NOT\_P6(i,t) + \varepsilon(t)$$

The dependent variable *Q(i,t)* is the probability distribution of numbers chosen by lotto players for the individual numbers i at draw t. Explanatory variables include lagged dependent variable *Q* and dummies for the frequencies of occurrence of the numbers in the past draws. *HIT* is -1 if the number i hit at draw t-1, and 0 otherwise. *HOT* is 1 if the number i is amongst the top 7 winning numbers within the last 99 draws, and 0 otherwise. *HIT_NOT_P6* equals to *HIT* if the probability distribution does not come from an ordinary bet. *HOT_NOT_P6* equal to *HOT* if the probability distribution does not come from an ordinary bet. Significance levels of 0.01, 0.05 and 0.1 are denoted by ***, **, and *, respectively.

	Small system bet		**Medium system bet**		**Large system bet**		**All System bet**		**System roll**	
	Coefficient	T Value	Coefficient	T Value	Coefficient	T Value	Coefficient	T Value	Coefficient	T Value
Intercept	0.0093***	62.03	0.0101***	61.78	0.0168***	65.83	0.0093***	61.88	0.0106***	57.67
Q(i,t-1)	0.6362***	100.87	0.6066***	89.41	0.3168***	32.23	0.6383***	101.5	0.5726***	74.35
HIT	0.0052***	43.64	0.0052***	39.47	0.0053***	14.9	0.0052***	43.98	0.0050***	35.83
HOT	0.0016***	14.35	0.0017***	13.41	0.0028***	8.2	0.0016***	14.41	0.0021***	15.32
HIT_NOT_P6	0.0024***	14.95	0.0024***	13.33	0.0025***	5.04	0.0024***	14.67	-0.0001	-0.71
HOT_NOT_P6	0.0007***	4.74	0.0003 *	1.76	0.0003	0.54	0.0007***	4.5	0.0002	1.12
Adjusted R-squared	0.6864		0.6263		0.1813		0.6880		0.5353	
Breusch-Godfrey LM Test	2.7280		19.2770		0.9973		3.2939		0.7411	
Draws	103		103		103		103		103	

References

Barber, B.M. and T. Odean (2003), All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Working Paper.

Barber, B.M., T. Odean, and L. Zheng (2005), Out of sight, out of mind: The effects of expenses on mutual fund flows, *Journal of Business*, 78: 2095-2119.

Bersabe, R. and R.M. Arias (2000), Superstition in gambling, *Psychology in Spain*, 4, 28-34.

Boynton, D.M. (2003), Superstitious responding and frequency matching in the positive bias and gambler's fallacy effects, *Organizational Behavior and Human Decision Processes*, 91, 119-127.

Burns, B.D. and B. Corpus (2004), Randomness and inductions from streaks: “Gambler's fallacy” versus “Hot hand,” *Psychonomic Bulletin & Review*, 11, 179-184.

Clotfelter, C. and P. Cook (1989), *Selling hope: State lotteries in America*. Cambridge: Harvard University Press.

Clotfelter, C. and P. Cook (1993), The ‘Gambler’s Fallacy’ in lottery play, *Management Science*, 39, 1521-1525.

Coleman, L. (2004), New light on the longshot bias, *Applied Economics*, 36, 315-326.

Cook, P. and C. Clotfelter (1993), The Peculiar Scale Economies of Lotto, *American Economic Review*, 83, 634-643.

Cox, S.J., G.J. Daniell, and D.A. Nicole (1998), Using maximum entropy to double one's expected winnings in the UK National

Lottery, *The Statistician*, 47, 629-641.

Croson, R. and J. Sundali (2005), The Gambler's fallacy and the hot hand: Empirical data from casinos, *Journal of Risk and Uncertainty*, 30, 195-209.

Durham, G.R., M.G. Hertzel and J.S. Martin (2005), The market impact of trends and sequences in performance: New evidence, *Journal of Finance*, 60, 2551-2569

Farrell, L., G. Lanot, R. Hartley, and I. Walker (2000), The demand for lotto: the role of conscious selection, *Journal of Business and Economic Statistics*, 18, 228-241.

Farrell, L. and I. Walker (1999), The welfare effects of lotto: evidence from the UK, *Journal of Public Economics*, 72, 99–120.

Forrest, D., R. Simmons, and N. Chesters (2002), Buying a dream: alternative models of demand for lotto, *ECONOMIC INQUIRY*, 485-496.

Gilovich, T., R. Vallone and A. Tversky (1985), The hot hand in basketball: On the misperception of random sequences, *Cognitive Psychology*, 17, 295-314.

Haigh, J. (1995), *Inferring gamblers' choice of combinations in the National Lottery*, Bull. IMA, 31, 132-136. Oxford: Butterworth-Heinemann.

Haigh, J. (1997), The statistics of the national lottery, *Journal of Royal Statistical Society A*, 160, 187-206.

Henze, N. (1997), A statistical and probabilistic analysis of popular lottery tickets, *Statistica Neerlandica*, 51, 155-163.

Hill, E. and J. Williamson (1998), Choose six numbers, any numbers, *The Psychologist*, 11, 17–21.

Jain, P.C. and J.S. Wu (2000), Truth in mutual fund advertising: Evidence on future performance and fund flows, *Journal of Finance*, 55:937–58.

Jarque, C.M. and A.K. Bera (1987), A test for normality of observations and regression residuals, *International Statistical Review*, 55, 163-172.

Keele, L. and N.J. Kelly (2005), Dynamic models for dynamic theories: the ins and outs of lagged dependent variables, *Political Analysis*, 14, 186-205.

Keren, G. and C. Lewis (1994), The 2 fallacies of gamblers: type I and type II, *Organizational Behavior and Human Decision Processes*, 60, 75-89.

Langer, E.J. (1975), The illusion of control, *Journal of Personality and Social Psychology*, 32, 311-328.

Lee, Y. K and C. T. Chang (2005), The social impacts of the public welfare lottery: An empirical study in Taiwan, *Modern Asian studies*, 39, 133-153.

Montier, J. (2002), *Behavioural Finance: Insights into irrational minds and markets*: Wiley.

Moore, P.G. (1997), The development of the UK national lottery: 1992-96, *Journal of the Royal Statistical Society Series A-Statistics in Society*, 160,169-185.

Mullainathan, S. (2000), Thinking through categories, Working paper, Department of Economics, Massachusetts Institute of Technology.

Papachristou, G. (2004), The British gambler's fallacy, *Applied Economics*, 36, 2073-2077.

Price, D. and E.S. Novak (2000), The income redistribution effects of Texas lottery games, *Public Finance Review*, 28, 82–92.

Rabin, M. (2002), Inference by believers in the law of small numbers, *Quarterly Journal of Economics*, 117, 775-816.

Rabin, M. and Vayanos, D. (2007). The gambler's and hot-hand fallacies: theory and applications, Working Paper.

Riniolo, T.C. and L.A. Schmidt (1999), Demonstrating the gambler's fallacy in an introductory statistics class. *Teaching of Psychology*, 26, 198-200.

Rogers, P. (1998), The cognitive psychology of lottery gambling: A theoretical review, *Journal of Gambling Studies*, 14, 111-134.

Roney, C.J.R. and L.M. Trick (2003), Grouping and gambling: a gestalt approach to understanding the gambler's fallacy, *Canadian Journal of Experimental Psychology*, 57, 69-75.

Simon, J. (1999), An analysis of the distribution of combinations chosen by UK National Lottery players, *Journal of Risk and Uncertainty*, 17, 243-276.

Statman, M. (2002), Lottery players/stock traders, *Financial Analysts Journal*, 58, 14-21.

Stranahan, H. and M.O. Borg (1998a), Separating the decisions of lottery expenditures and participation: A truncated Tobit approach, *Public Finance Review*, 26, 99–117.

Stranahan, H. and M.O. Borg (1998b), Horizontal equity implications of the lottery tax, *National Tax Journal*, 51, 71–82.

Strickland, L.H., R.J. Lewicki, and A.M. Katz (1966), Temporal orientation and perceived control as determinants of risk-taking, *Journal of Experimental and Social Psychology*, 2, 143-151.

Terrell, D. (1994), A test of the gambler's fallacy – Evidence from pari-mutuel games, *Journal of Risk and Uncertainty*, 8, 309-317.

Terrell, D. (1998), Biases in assessments of probabilities: New evidence from greyhound races, *Journal of Risk and Uncertainty*, 17, 151-166.

Thaler, R.H. (1992), *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, New York: The Free Press.

Tversky, A. and D. Kahneman (1971), Belief in the law of small numbers, *Psychological Bulletin*, 76, 105-110.

Tversky, A. and D. Kahneman (1974), Judgment under uncertainty: heuristics and biases, *Science*, 185, 1124-1131.