台灣樂透市場投注者選號行為之研究
Selection Behavior of Taiwan Lotto Players
-Dynamic analyses of number selection

Hwai-Chung Ho, Shih-Chin Lee, Hsiou-wei Lin

Outline

- (I) Importance of studying lotto markets
- (II) Motivation and purpose
- (III) Hypotheses
- (IV) Methodologies
- (V) Empirical results
- (VI) Conclusions

(I) Importance of studying lotto markets

- **(1)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.**

- **(2) Durham, Hertzl, Martin (2005) : Betting markets have several advantages over traditional capital markets and experimental laboratory.**

(II) Motivation and purpose

- **Indirect analyses (publicly available data)**

  - **(Farrell, Lanot, Hartley, and Walker 2000; Papachristou, 2004) Investigate the betting behavior or to estimate the elasticity of demand for lottery by using only public, limited available data.**

  - **Propose a more efficient method**

Direct analyses

- **(1) Dynamic models**

  (Rabin, 2002; Rabin and Vayanos, 2007) : Develop cognitive models to explain gambler's fallacy and hot-hand biases in people's decision-making.

- **(2) Thinking through category**

  (Mullainathan, 2002) : Present a model of human inference in which people use coarse categories to make inferences.

  The first two models provide some more insights into financial anomalies .

- **(3) Illusion of control**

  Individuals believe that they exert control over events that are in fact randomly determined (Langer, 1975).

(III) Hypotheses

Gambler's fallacy

- **Expecting outcomes in random sequences to exhibit systematic reversals**
- **In the fairness of coin-flipping experiments, subjects seem to believe that heads and tails should balance even in small samples (Tversky and Kahneman,1971)**
- **Pick-three lottery game:**
  1. **Clotfelter and cook (1993):Maryland lottery**
  2. **Terrell (1994) :New Jersey lottery**
- **Lotto(6/49): Papachristou (2004) documents that history information only marginally affected in UK.**

Hot-hand fallacy

- **Basketball fans expect that players have significant hot hands, being more likely to make a shot following a successful streak (Gilovich, Vallone, and Tversky, 1985).**
- **Rabin and Vayanos (2007) propose a model to reconcile the gambler's fallacy and hot-hand fallacy in the prediction of random sequences.**
- **In their model, 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 the manager .**

Thinking through categories

- **The set of categories forms a partition of the posterior space and people choose the category which is most likely given by the data.**

- **Which lotto ticket is more likely to win the jackpot prize ?**

  1.) 3 16 17 29 34 37 2.) 1 2 3 4 5 6

Most people prefer the first ticket because winner numbers come from a random machine and the event of **6 consecutive numbers** is less likely than that of non-consecutive numbers and thus creates an impression that the latter is more random.

Illusion of control

- **Lotteries in North America did not become popular until New Jersey introduced a game which allowed players to select their own numbers (Thaler, 1992)**
- **Subjects bet more money and played with more confidence than other people in their chance of winning if they threw the dice themselves. Strickland, Lewicki and Katz (1966)**
- **System bet players tend to take chances and seem to have more confidence than ordinary bet type player.**

(IV) Methodologies

- **Data**
- **Conscious selection**
- **Average picking frequency**
  - **Winner ball and loser ball groups**
- **Dynamic models**
  - **Variable HIT : Luck**
  - **variable HOT : Ability**
- **Non-consecutive combinations**
  - **Variable JUMP**
- **Three types of bets**
  - **Ordinary bet, System roll, System bet**

(V) Empirical results

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

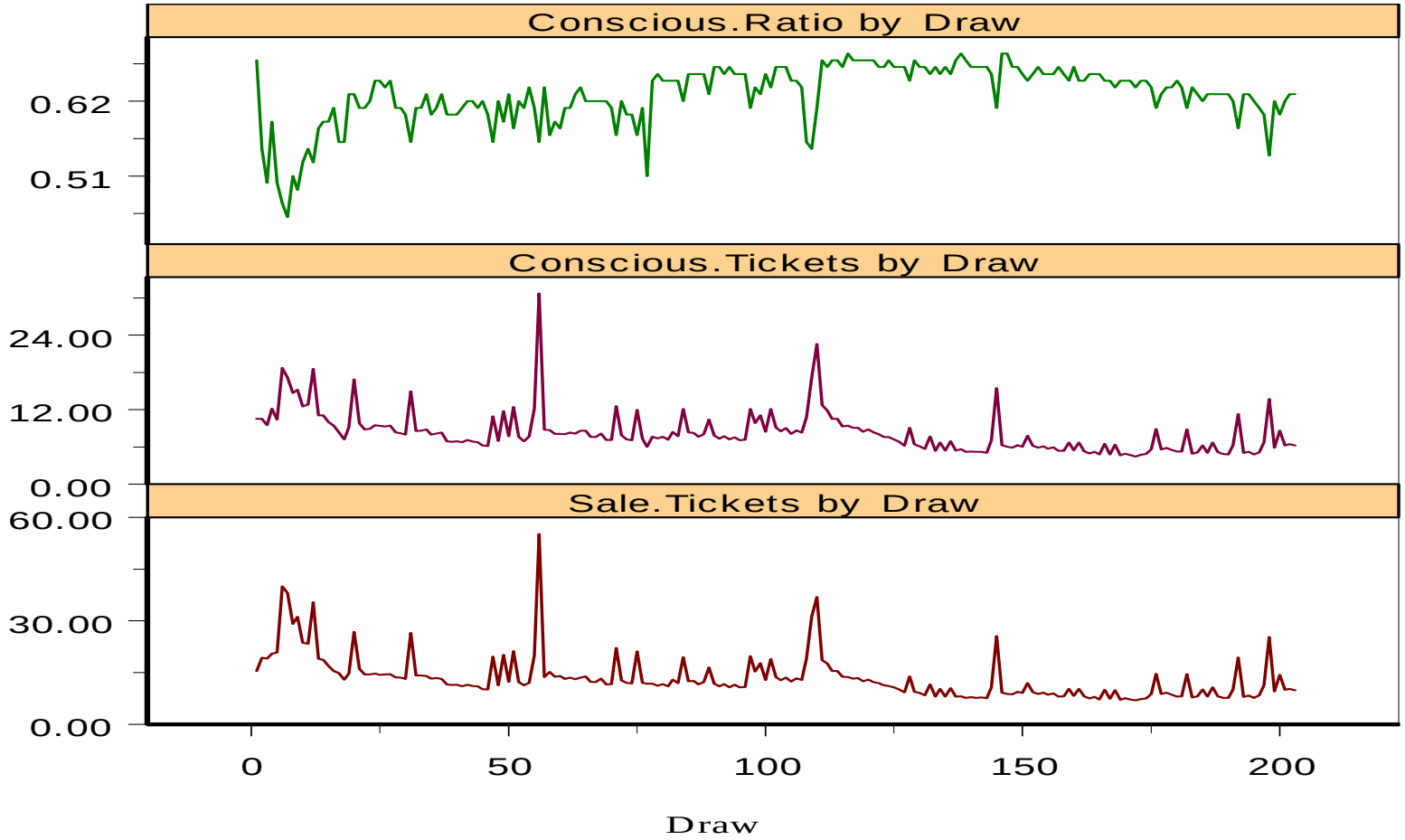


Table 1 Determinants of the ratio of conscious number selection

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

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

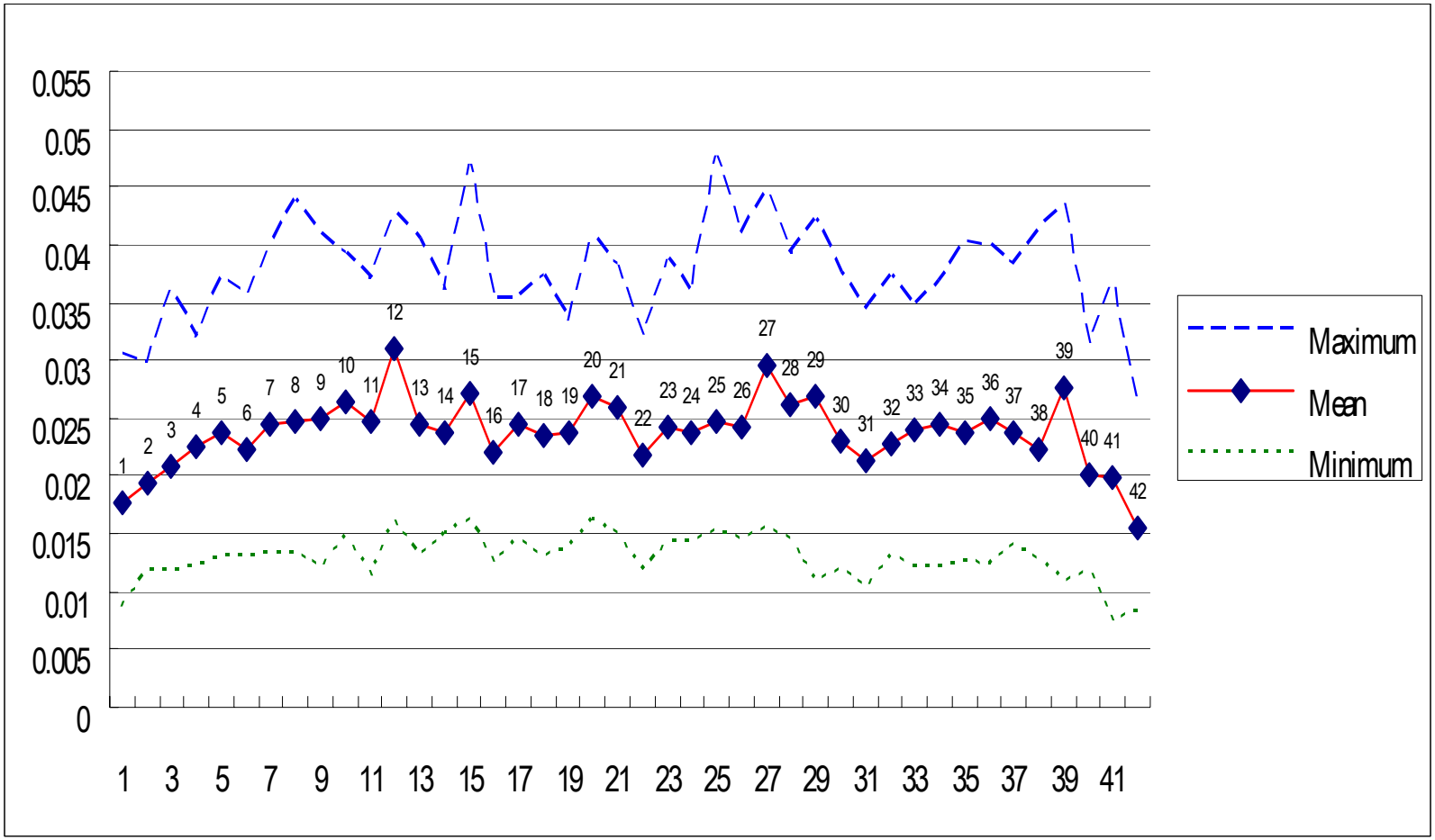
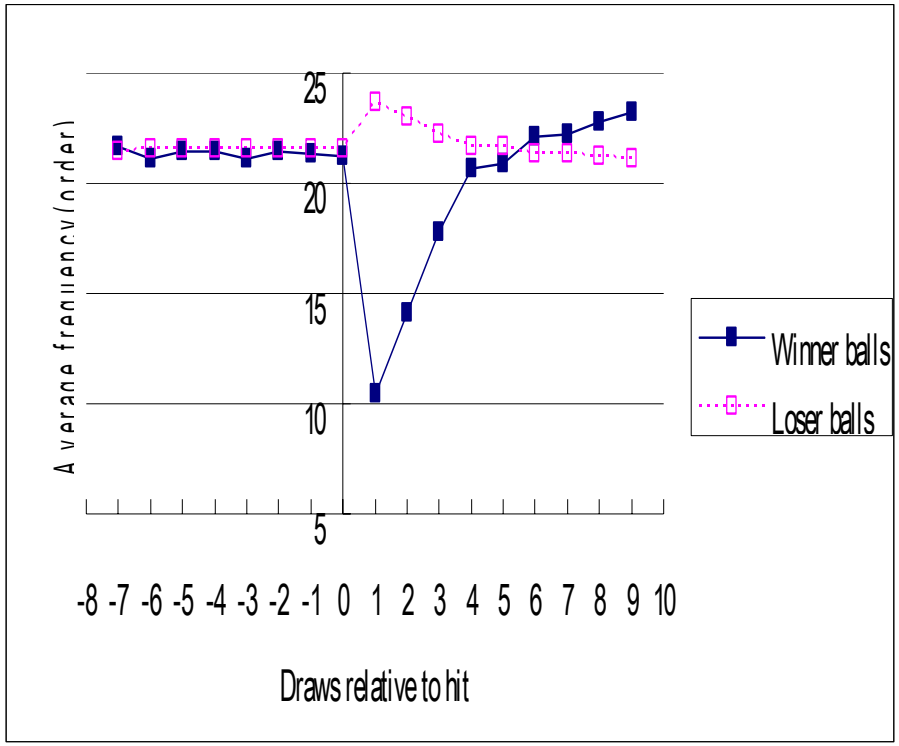


Figure 3 Picking frequencies for reaction to hit.



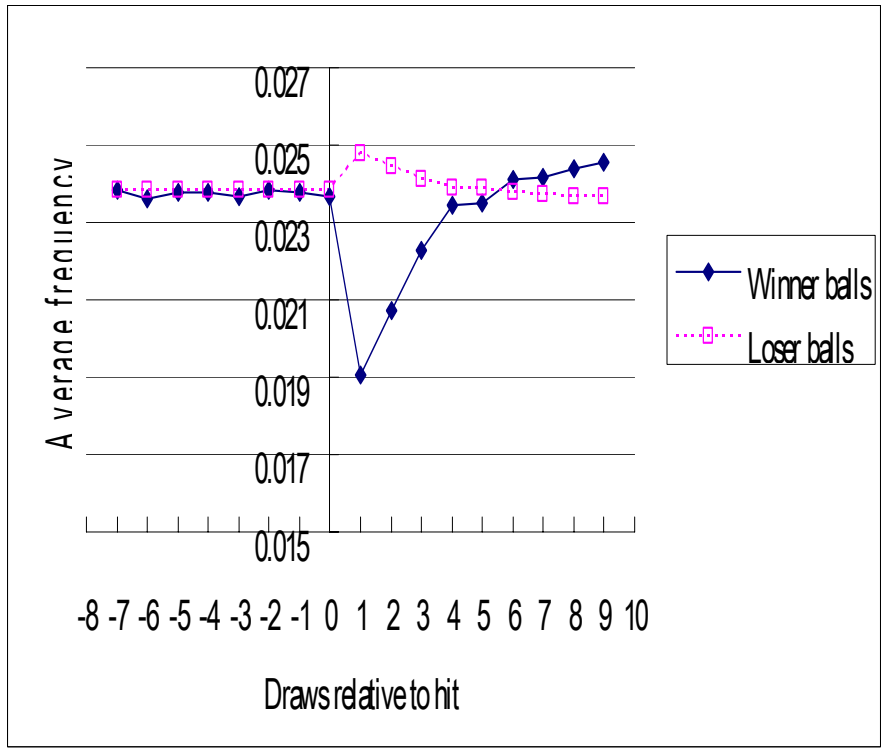
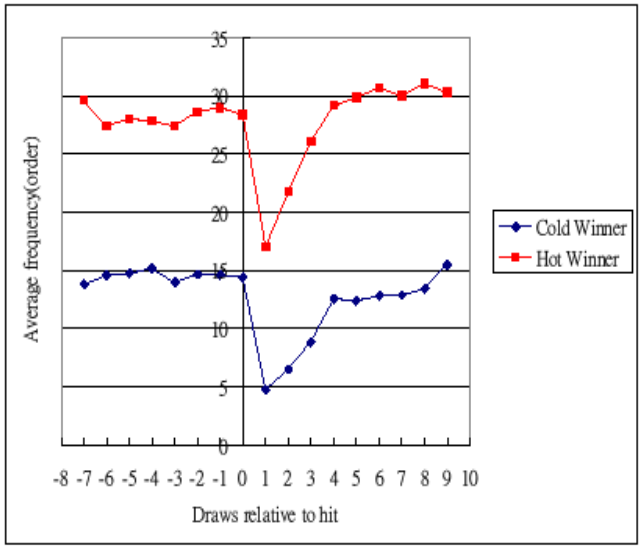


Figure 4 Picking frequencies for reaction to hit across winning frequencies.



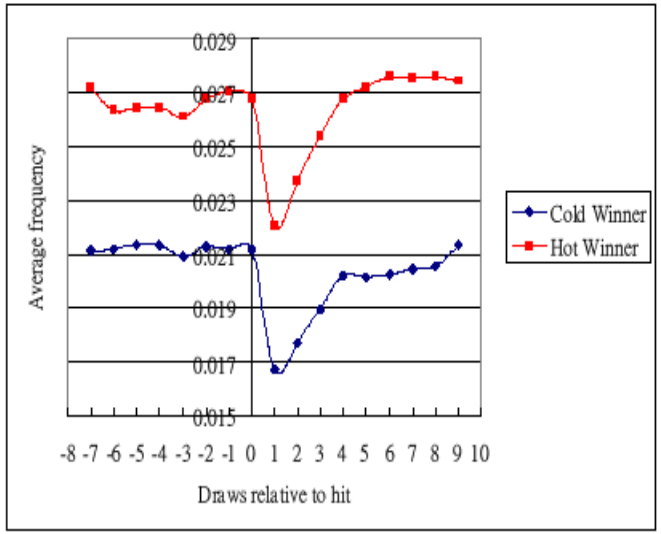


Table 2. Playing strategy for lotto tickets covering no. i

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

$$\text{Model 1: } Q(i,t) = \alpha + \beta_1 HIT(i,t) + \beta_2 HOT(i,t) + \varepsilon(t) \qquad (3.2)$$

$$\text{Model 2: } Q(i,t) = \alpha + \beta_1 Q(i,t\text{-}1) + \beta_2 HIT(i,t) + \beta_3 HOT(i,t) + \varepsilon(t) \qquad (3.3)$$

$$\text{Model 3: } Q(i,t) = \alpha + \beta_1 Q(i,t\text{-}1) + \beta_2 HIT\_DOUBLE(i,t) + \beta_3 HOT(i,t) + \varepsilon(t) \qquad (3.4)$$

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

$$+ \beta_5 BALL(i,t) + \varepsilon(t) \qquad (3.5)$$

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

	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

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

and define the variable JUMP on X by:

$$JUMP(t_1,\ldots,t_6)=5-\sum_{i=1}^{5} I_{(t_{i+1}-t_i)} \quad ;\ \text{where}\quad I_{(t_{i+1}-t_i)}=1\ \text{if}\ \ t_{i+1}-t_i=1,\ \text{i=1,…,5,} \qquad (3.6)$$

and 0, otherwise.

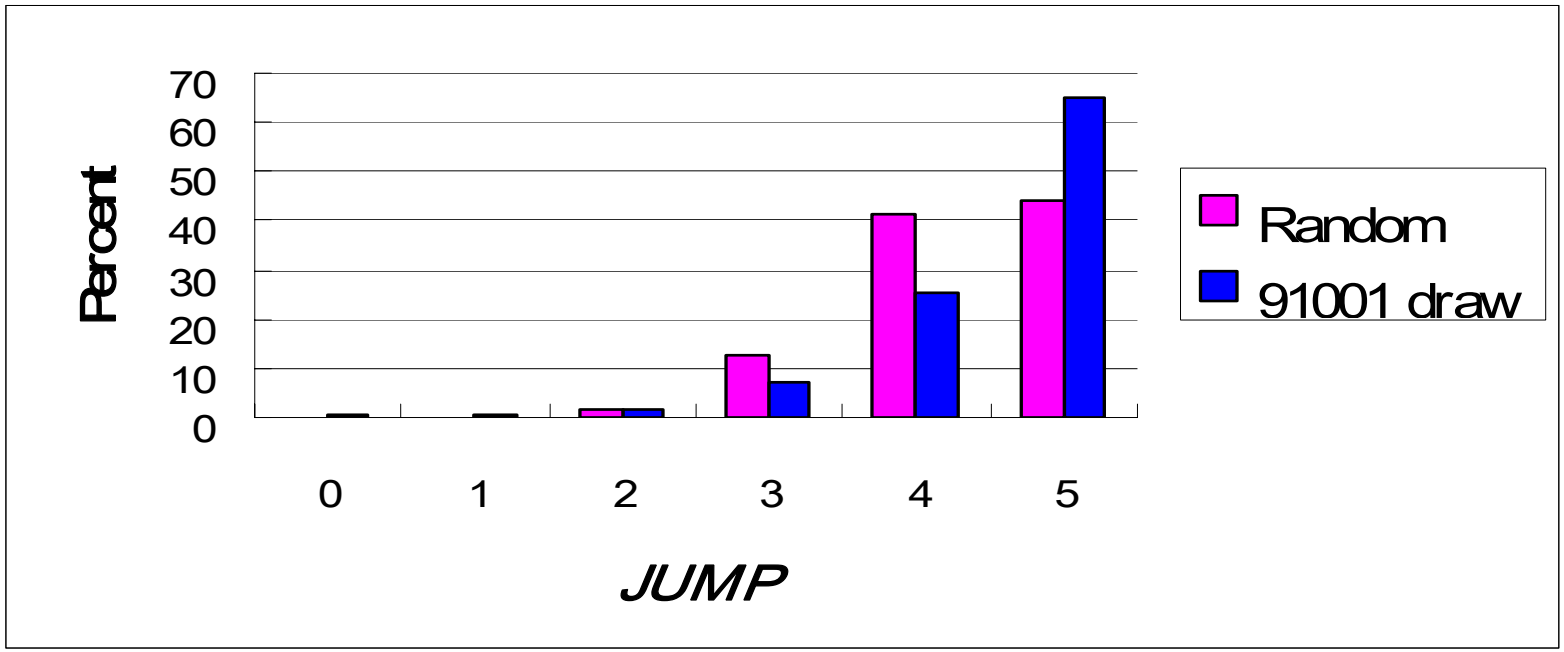


Table 4 Descriptive statistics of JUMP

$$t_i = \sqrt{S_i}(\mu_i - \mu)/\sigma_\mu \quad \text{i=1,.., n;} \qquad \text{T=}\sum_{i=1}^{n} t_i/\sqrt{n}$$

	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

Three types of betting starting at the 100th draw

- **Ordinary bet : select 6 numbers**
- **System roll : select 5 numbers, the computer assign the remaining 37 number to these 5 numbers.**
- **System bet : selection 7 to 16 numbers**

Table 3.4 Three types of betting and their amounts

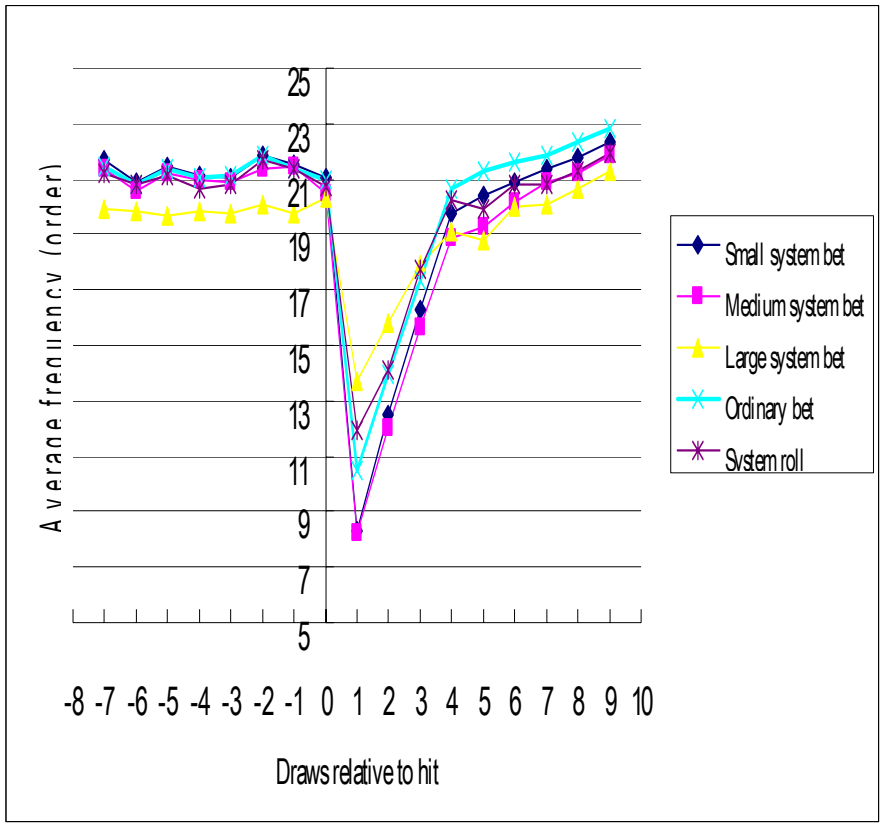
Bet type	No. of Ordinary Bet combinations	Amount (in NT$)	Bet type	No. of Ordinary Bet combinations	Amount (in NT$)
Ordinary	1	50	Sys 12	924	46,200
Sys 7	7	350	Sys 13	1716	85,800
Sys 8	28	1,400	Sys 14	3003	150,150
Sys 9	84	4,200	Sys 15	5005	250,250
Sys 10	210	10,500	Sys 16	8008	400,400
Sys 11	462	23,100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

- **Small system bet : Sys7-Sys9**
- **Medium system bet : Sys10-Sys11**
- **Large system bet : Sys12-Sys16**

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



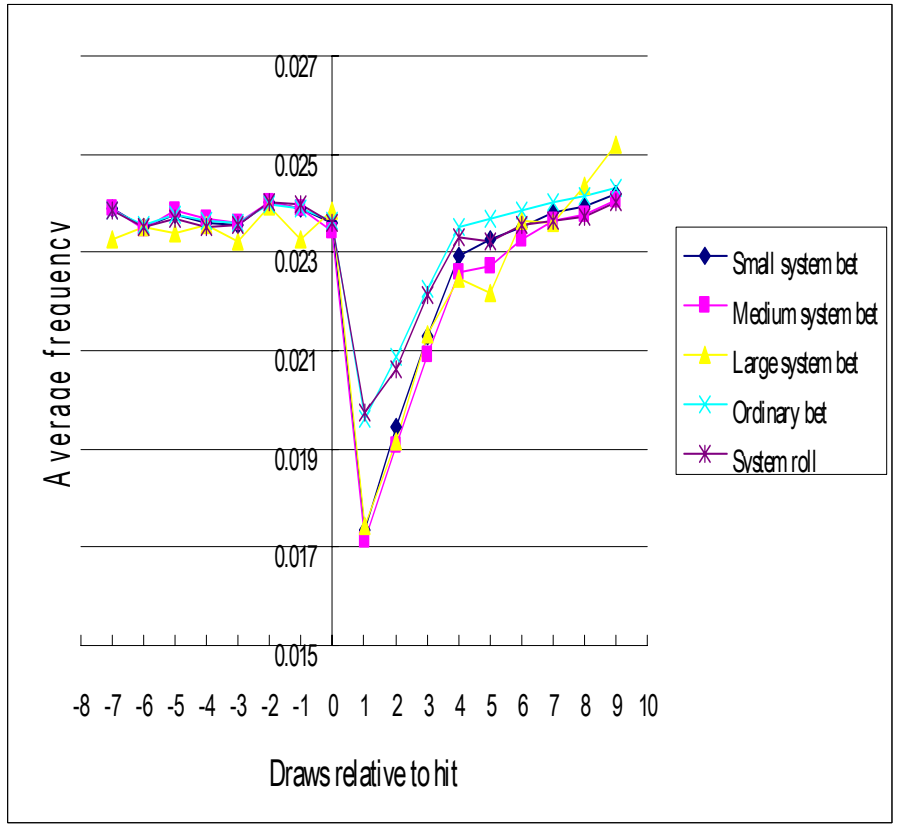


Table 7 Determinants of the probability of the numbers picked across betting types

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

1. *HIT_NOT_P6* equals to *HIT* if the probability distribution does not come from an ordinary bet.

2. *HOT_NOT_P6* equal to *HOT* if the probability distribution does not come from an ordinary bet.

For example :

If the observations come from system bet. Then

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

If the observations come from ordinary bet. Then

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

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

	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	

(VI) Conclusions

This paper analyzes a sample of 1,679,676,226 combinations of lottery ticket numbers consciously chosen by the players of the Taiwan lotto for the period from 2002 to 2003.

- **First, the gambler's fallacy temporarily influences players' selection of lotto 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 picking the numbers that appeared more frequently in the past.**
- **Third, most players avoid picking consecutive numbers, extending the concept of representativeness heuristic. In addition, the win-stay strategy is shown to exist.**
- **Forth, the players using the system bet strategy have stronger misconceptions about random processes than the players using the ordinary bet strategy.**