Ding, Jieyao

Working Paper

What numbers to choose for my lottery ticket? Behavior anomalies in the Chinese online lottery market

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2011,23

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Suggested Citation: Ding, Jieyao (2011) : What numbers to choose for my lottery ticket? Behavior anomalies in the Chinese online lottery market, Preprints of the Max Planck Institute for Research on Collective Goods, No. 2011,23, Max Planck Institute for Research on Collective Goods, Bonn

This Version is available at:
https://hdl.handle.net/10419/57511

Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2011/23

What Numbers to Choose for My Lottery Ticket? Behavior Anomalies in the Chinese Online Lottery Market

Jieyao Ding

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2011/23

What Numbers to Choose for My Lottery Ticket?
Behavior Anomalies in the Chinese Online Lottery Market

Jieyao Ding

September 2011

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

What Numbers to Choose for My Lottery Ticket?

Behavior Anomalies in the Chinese Online Lottery Market

JIEYAO DING [#]

August 2011

Abstract

The Chinese Online Lottery provides field evidence of three anomalies. The first anomaly, which has previously not been documented when there is a financial incentive to overcome, is the guidance effect. Since the target game in this project is a pari-mutuel game, which means people will share the jackpot with other winners, the best strategy should be to choose the least popular numbers among others – information that people could obtain on the webpage. However, to my surprise, instead of doing so, people would choose the most popular numbers among others. The second anomaly tested is the gambler's fallacy. Although it is proved that the gambler's fallacy does exist, the influence lasts only three days, which is much shorter than prior research. Furthermore, the dataset's availability makes it possible to show how the two fallacies unfold over time within a round. This was unlikely before the phenomenon of online betting. The result demonstrates that later entrants are subject to more fallacies than earlier ones. Finally, the paper adds to the evidence showing the additional, culturally contingent pull of special numbers. In China, bettors prefer to choose the lucky number 8, even it won the game in prior rounds, but they are reluctant to choose the unlucky number 14 even it has not been picked for a long while.

Keywords: Lottery Game, Gambler's Fallacy, Guidance Effect, Number Culture
JEL Classification: C93 Field Experiment,
D81 Decision Making under Risk and Uncertainty

The author would like to thank Martin Pollet for creating a program to collect data.

[#] Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str.10, 53113 Bonn, Germany, [Email: ding@coll.mpg.de].

1. Introduction

www.500wan.com is the biggest Chinese online gambling company which supplies more than 20 types of game in mainland of China. The game that attracts my attention is "pick 5 out of 22", since the game's webpage updates information called "hot and cold number information" that displays what were the most/least popular numbers (hot/cold numbers) among those previously entered by bettors in the same round of the game. By grasping the information data, I try to dig out, first, how this information influences bettors' behavior other than winning numbers in prior rounds. More importantly, what I intend to investigate is how bias unfolds across time, if any bias exists in the market. To the best of my knowledge, both points have not been documented in lottery games before.

Principally, predicting winning numbers in lottery games is an impossible mission. Because the drawing procedure is (assumed to be) random and independent, the winning probability should be equal across numbers in each draw. And the frequency with which a number was drawn in a previous game does not provide any hint about whether the number will be chosen again in a current game. However, over the past decades, the anomalies that violated these statements were found among bettors' behavior both experimentally and empirically. Therefore, I shall consider the following anomalies caused by hot and cold number information, as well as previous winning numbers that would possibly appear in the market.

The first bias that could be taken into consideration is the favorite-longshot bias which describes bettors underestimating the winning probability of favorite and overestimating the winning probability of longshot (Ali, 1977). It has often happened in both financial market and racetrack games. In horse racing, when bettors join the game, they could see the temporary odds for each horse. It has been empirically proved that bettors often give a horse a lower subjective probability when it has a higher objective probability of winning, and a higher subjective probability when it has a lower objective probability of winning. Researchers gave out some possibilities for the bias, e.g., it is a Kahneman-Tversky type error, and people enjoy choosing the longshot, even for some irrational reasons, such as the position of the horse (See Thaler and Ziemba, 1988). Furthermore, people argued that the bias is caused by the inside traders (Shin, 1991, 1992, 1993) or the transaction cost (Terrell and Farmer, 1996). In the market I investigate, hot and cold number information is considered to be a similar type of information to the odds shown in horse race. Therefore, it is reasonable to conjecture that bettors in the lottery market are subject to favorite longshot bias after seeing the information.

The second anomaly that was often found in lottery games in prior research is the gambler's fallacy. Gambler's fallacy is the belief that the probability of an event becomes smaller after the event has occurred recently, even though it is objectively known that the probability of the event is independent across trials. This first emerged in Tversky and Kahneman's (1974) experimental work which pointed out that subjects acted as if an event was negatively correlated with a prior event, even though it was common knowledge that all the events were independent. Tversky and Kahneman propose that people view chance as a self-correcting process, which means a deviation in one direction will lead to a deviation in the opposite direction so as to keep the equilibrium. The existence of gambler's fallacy has been proved not only in the lab, but also in the field. Terrell

(1994) examined a pari-mutuel game in New Jersey, where the payoff was shared by all bettors who hit the particular number. He found that the bet on numbers that had won the game in previous days decreased significantly and this effect virtually lasted for a long while. Such negative recency effect not only existed in lottery games, but also showed up in many other areas, for example in greyhound races (Terrell, 1998), in Casinos (Croson and Sundali, 2005), and so forth. On the homepage of the "pick 5" game, the information about which numbers were picked in the past is available as well. It could be the case that Chinese bettors are affected by this information, as in other markets. Therefore, the gambler's fallacy is the second bias tested in the paper.

While there is some evidence of biases, the existing evidence is less clean than one might ideally wish. The existing evidence suffers from the following limitations: first, in horse racing, where the favorite longshot bias was found, not every gambler could hold the idea that the winning probability is objectively equal for every horse, and it is difficult to adjust such a thought in order to compare the betting amount on each candidate. Most research calculates the objective winning probability, which is not possible to be available to bettors in advance. Further, it is hard to find out which participant has inside information that could help to bet more accurately, and the transaction cost is unknown as well. Second, in prior research, only each game's final bet data were available, which implies that researchers were only able to use simulations to know what was going on during the procedure of the game. This is an imperfect alternative for such investigation. A lot of detailed information is lost in this case. Third, either in horse racing or lottery games, it is not easy for bettors to receive clues about what others did during the game. For example, bettors bought a lottery ticket in a lotto shop where it was difficult to have full information about what the others chose. For gamblers in horse racing, they could only deduce what others did from the odds. Therefore, it is hard to assess the influence of the choices of others.

By contrast, in my data set, all the limitations mentioned above can be solved. Compared to horse racing, the objective probability of a number being chosen is identical for each number, i.e., 1/22.[1] Meanwhile, it excludes the possibility of inside traders or information: people have no way to increase their guessing accuracy by means of an internal transaction. Hence, the first limitation is solved. Second, the internet is a perfect medium to investigate the time pattern[2] of betting with real-time serial data that shows what happened in each fixed period of time, e.g., each half hour. It is more accurate than simulation for getting to know how the bets have developed across time. And it is more powerful to conclude how bettors are influenced by the choices of others, since this information is more straightforward.

After all the limitations have been solved, I will investigate the market in the following steps. First, it is necessary to know how the hot and cold number information influences bettors' behavior, as well as whether the gambler's fallacy exists and interacts with this information. If any anomaly exists in the market, the second step would be to display how it develops across time. To the best

[1] There are 26334 possible 5-number combinations between 1 and 22.Therefore, the winning probability of each 5-number combination is 0.0000379.

[2] In this paper, the time pattern represents what numbers are picked over time within each draw of the lottery game, e.g., how many wagers each number has from the beginning until the updated time which is shown in "hot and cold number information".

of my knowledge, this is the first research on the time pattern of bias with a single round of a lottery game with real data instead of using simulation technology. Last but not least, I will attempt to find out whether bettors rely on Chinese number culture in making their choice of number combination.

The paper is organized as follows. Part 2 introduces the game in detail and describes how the data are collected. Part 3 lists the hypotheses that will be tested in the paper. Part 4 gives an overview of betting behavior in the market. Part 5 will show the main findings in the market. Part 6 is the conclusion.

2. Game and Data Collection

2.1. The Game

As mentioned above, www.500wan.com is the first and the biggest online gambling company in mainland China. It supplies almost all the state gambling games and is regulated by the corresponding authority. The final draw is executed by the state lottery institution and shown live. "Pick 5" is a game where bettors choose 5 numbers between 1 and 22. Only the 5 numbers in the combination matter, but not the order of the numbers. Each lottery ticket costs 2RMB/0.31USD and each bettor can buy infinite number of such tickets. The bettor whose five-number combination is the same as the final draw gets to share the jackpot. The money is distributed on a pari-mutuel basis, that is, earning is equal to the jackpot divided by the number of winning wagers.[3] The size of jackpot is 50% of the present round's sale and will not be carried forward. The rule reveals that the more money is put into the game, the bigger the prize is, but the more wagers win the game, the smaller each wager's prize is. There are small earnings as well. In each 5-number combination, the payoff depends on how many numbers match the winning numbers. If 4 numbers match, the payoff is 50RMB/7.6USD per wager, while if 3 numbers match, the payoff is only 5RMB/0.76USD. These payoffs are fixed and informed in advance.

2.2. Data Collection

The data is collected automatically by a JAVA[4] program that works on a server, so that it can collect data continuously. It grasps data from the code of the webpage every half hour and generates a .csv file every day.

The dataset includes two main parts. One part is the general information for each time of draw which contains the combination of winning numbers, the final bet frequency for each number, the total bet each day, etc. The data was collected from October 2009 to August 2010 (the round numbers are shown in Appendix I).

The other part is the time series data, information that is shown in "hot" and "cold" numbers. This information displays the top 10 hot numbers and top 10 cold numbers among bettors who entered game prior to the update of the information, which includes how many bet on hot and cold

[3] In the past, researchers often described the jackpot as being divided by the number of winners. This is not fully correct, since it is possible that one bettor buys more than one lottery ticket. Therefore, I use wagers instead of winners.

[4] The program is available upon request.

numbers and the percentage of the bet (see Appendix II)[5]. The game starts at 19:40 (Chinese time) and ends at 19:40 the next day. Since the information is updated every half hour, for each round (24 hours) there are 48 sets (each set has 22 data points) of observation.[6]

3. Hypotheses

3.1. Timing of Entering

In Terrell and Farmer's (1996) paper, they defined two types of bettor in gambling market: the pleasure bettor and the professional bettor. If people bet for fun, they enter the game as early as possible so as to have more enjoyment from purchasing the lottery ticket. This was recently tested experimentally (Kocher etc. 2009). Professional bettors, the other type, enter the game a short time before the end of the game because they can get the nearly final odds information in order to share the jackpot with fewer bettors. If two types of bettor both show up in the market, there will be significantly more bettors entering during the first and last few hours of the game than in between. It should be noticed that I only assume that pleasure players prefer to enter at the beginning of the game and professional players prefer to enter at the very end of the game. However, it cannot be deduced that all bettors who entered earlier are pleasure bettors, nor that bettors who entered later are professional bettors.

Hypothesis 1: At the beginning and the end of the game, more bettors enter than in between.[7]

3.2. Hot and Cold Number Information

Supposing the aim of purchasing a lottery ticket is to win as much money as possible, given that bettors do not have the ability to change the winning probability as they want, by intuition it will be rational to choose the least popular numbers, so that people can share the jackpot with less winners and earn more money. For the game I investigate here, it specifically means people should focus on the cold numbers shown in the information and choose the five which have the least bets of all.

Hypothesis 2: The less popular (cold) a number was, the more bettors chose it.

The more popular (hot) a number was, the fewer bettors chose it.

3.3. Gambler's Fallacy

It is proved both in the lab (Tversky and Kahneman, 1974; Morrison and Ordeshook, 1975) and in the field (Clotefelter and Cook, 1993; Terrell, 1994) that bettors were subject to gambler's fallacy in the lottery game. In Clotfelter and Cook's paper (1993), they found that after a

[5] From the code of the webpage, I get all 22 numbers' wagers in each half hour. Therefore, this is a full dataset instead a dataset showing only the top 10 and bottom 10.

[6] The last half hour's information could not be seen by the bettors since the new round had already begun. But we could get this data from the code of the webpage for analysis.

[7] Time possibly influences the entrance as well. Since the game starts at 19:40, people would get to sleep and go to work afterwards. Therefore, at the beginning and the end of the game, it is easier for bettors to enter the game.

number was picked in the previous game, the amount of bets on this number fell sharply and it took a few months for this number to recover to original bet level. Later, Terrell (1994) challenged this result with a parimutuel lottery game and found similar results, although the effect was weaker than before. I suppose that in the market I have investigated, bettors are influenced by prior winning numbers as well. But I propose the effect will not last as long, for the following reasons: first, for each day, 5 numbers are picked. Therefore, each number will receive less focus than the games mentioned above, since in their games, only one number was picked for each round of game. Second, in the "pick 5" game, there are only 22 numbers to choose from, and each time, people need to choose a 5-number combination, which means the re-picking rate is higher than before. Therefore, the effect could last for a very short time only.

Hypothesis 3: Bettors are subject to gambler's fallacy. But the effect lasts for a short while.

3.4. Time Pattern of Biases

3.4.1. Gambler's Fallacy

Suppose the market has two types of bettor, as mentioned above. By intuition, pleasure bettors would be influenced more by prior winning numbers while professional bettors would insist that prior winning numbers are independent of the draw in the present round. Therefore, if pleasure bettors enter the game earlier, while professional bettors enter later, the effect of gambler's fallacy should become weaker over time.

Hypothesis 4: The gambler's fallacy effect becomes weaker across time in a round of the game.

Jørgensen et al. found that if a number was on a streak, i.e., if that number won the game continuously, people's attitude would switch to prefer choosing it; this is called "hot-hand fallacy". Therefore, I make the hypothesis that if a number was picked continuously in the "pick 5 game", bettors' attitude towards this number would switch from gambler's fallacy to hot-hand fallacy.

Hypothesis 5: Gambler's fallacy switches to hot-hand fallacy if a number won the game extremely often.

3.4.2. Hot and Cold Number Information

At the beginning of the game, the sample size for hot and cold number information is relatively small. Pleasure bettors prefer their own favorite numbers more than the numbers popular among others.[8] And professional bettors who intend to enter the game later in a round would rely on this information more, since this is an indication for the odds which directly related to how much they could earn in the game. Therefore, information about hot and cold numbers should make people go in the opposite direction. This means that the more popular a number was in the

[8] In the paper by Jørgensen papered al., it is proved that most players did not change their number combination week after week.

information, the less likely it is that people will choose it. And later entrants are affected more than earlier ones, since more professionals enter later.

Hypothesis 6: The influence of hot and cold number information is getting stronger over time.

3.5. Number Preference

It is necessary to notice the influence of Chinese number culture which often gives an indication for how numbers are chosen, especially in gambling. Many Chinese hold the idea that number 8 is a lucky number, while number 14 is an unlucky one.[9] Therefore, it could happen that more people would like to choose number 8, but not number 14, in the lottery market, especially for pleasure bettors.

Hypothesis 7: Number 8 will have more wagers, while number 14 has less.

4. Overview

For the ease of understanding, I first explain the notations that are used in this paper. In a "pick 5" game, there are 22 numbers that bettors can choose from, which are notated as ***i*** (***i***=1, …, 22). Each ***round (r)*** of the game starts at 19:40 Chinese time and ends at 19:40 the next day, which means each round of the game lasts for 24 hours. The point at which the information updated is marked with ***t*** (***t***=1,…,48). At each time ***t***, there is one set of observations that has 22 data points that show how many bets $BET_{i,r,t}$ there are on each number ***i***.

Principally, if bettors are unbiased, 22 numbers should receive nearly equal bets at the end of each round of the game. In order to know whether this is the case, Figure 1 shows the mean of final wagers on each number. The solid line is the mean of wagers if people hold the idea that the winning probability is the same across numbers and randomly choose 5 numbers out of 22. The dash line is the mean of real wagers on each number. It is obvious that people put different bets on each number and the discrepancy is huge, which shows a bias towards numbers. Among all, number 8 has 1342 wagers on average, which is the number chosen most often, while number 1 has only 934 wagers. Numbers 14 and 22 are the bottom 2 and 3 numbers and have 1121 and 1127 wagers on average, respectively. Apparently, the difference cannot be simply ignored. As mentioned before, Chinese number culture does have an effect on choosing numbers, since number 8 has the most wagers, while number 14 is picked much less often than the average.

[9] Number 8 is a lucky number because it has a similar pronunciation with "fa" in Chinese, which means "big fortune". Number 14 is a bad number because it has a similar pronunciation to "yao si" in Chinese, which means "will die".

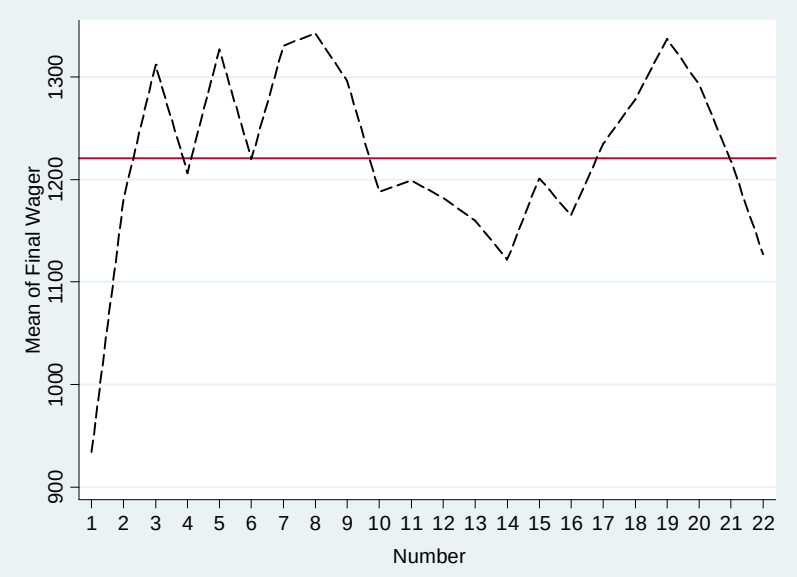

Figure 1 Mean of Final Wager for Numbers[10]

As mentioned before, both professional bettors and pleasure bettors are possibly in the market. Since the game I am studying now is a pari-mutuel game, the best strategy for bettors, especially professional ones, would be to enter the game a short while (e.g., in the last half hour) before the end of the game because people could have the information of the odds near the end and bet on the least popular numbers, in order to share the jackpot with less people to earn more money. On the other hand, pleasure bettors could enjoy more pleasure if they held the lottery ticket from the beginning of the game. Therefore, pleasure bettors would like to enter the game once it starts. If the game has both types of bettor, it is possible to have more bets at both ends of the time line than in between. This is proved in figure 2, which shows the mean of the increase of bets across time. It is obvious that, in each round, at the beginning and the end of game, there are more wagers than in between. The amount of bets increases sharply from T=1 and reach the peak at T=7. Then, it falls down dramatically until around T=16. After that, it recovers gradually and has grows relatively fast from T=37 until the end of game. Hypothesis 1 is supported.

[10] See Appendix III Table 1 for a summary statistic of bets on each number.

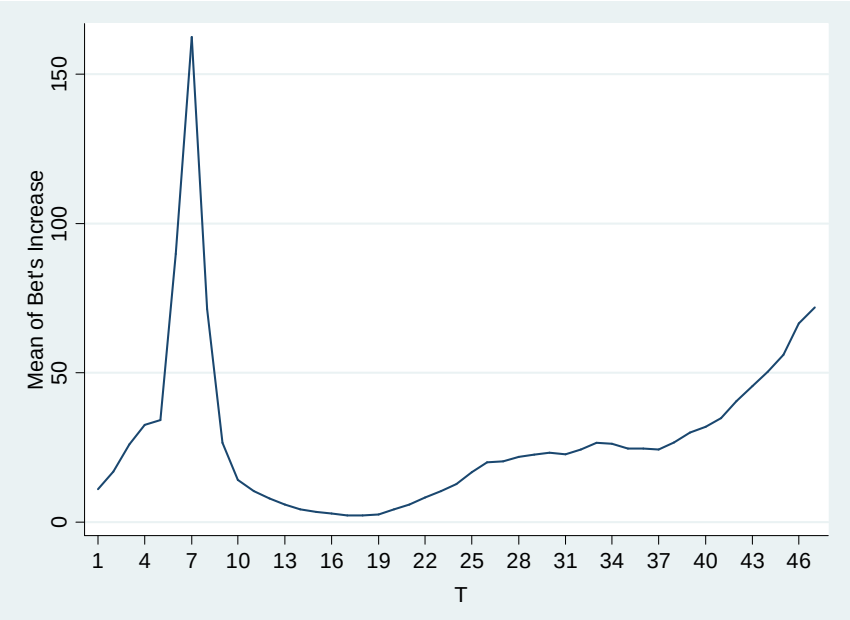

Figure 2 Time Pattern of Betting

5. Results

In this paper, I only focus on the information supplied by the main webpage of the game which includes the winning numbers from the last eight rounds, the previous round's payoff, hot and cold number information and so forth. I do this because my main focus is the influence of hot and cold number information and the time pattern of biases. The main webpage supplies enough information for these research goals.

In this session, I will discuss what I got in two main parts. In 5.1, I mainly introduce regressions of bet on different information (e.g., prior winning numbers, hot and cold number information) in order to show what biases are found in the market. Then, in part 5.2, I will figure out how the biases unfold across time in a single round of the game namely "the time pattern" of biases. And in part 5.3 the Chinese number culture is used as a possible explanation for findings in part 5.2.

5.1. Biases towards Information

In this part, I mainly analyze how the information shown on the homepage influences betting behavior, and I report all the biases found in the market step by step. First, I will test whether bettors are subject to gambler's fallacy in the Chinese market, with each round's final data as prior research. Then, I add hot and cold number data into regression and report the results tested with full data set.

In order to control for the amount of daily sale, I use the bet's proportion $P_{i,r,t}$ instead of the absolute bet for analysis:

$$P_{i,r,t}=BET_{i,r,t}/(\Sigma BET_{j,r,t}+ BET_{i,r,t}) \qquad (i\neq j) \qquad (1)$$

$BET_{i,r,t}$-Bets number ***i*** has at time ***t*** in round ***r***.
$P_{i,r,t}$-Percentage of bets number ***i*** has at time ***t*** in round ***r***.

First, I examine whether bettors are subject to gambler's fallacy, as observed in prior studies. Although bettors could have full information about numbers that won the game before, the main purpose of this paper is to test the influence of information on the homepage of the game which lists the information about the winning numbers from the previous eight rounds, particularly with regard to hot and cold numbers. Therefore, I test whether the winning numbers from the previous eight rounds influence betting behavior with regression (2).[11]

$$FP_{i,r}=\alpha + \beta_1 WIN_{i,r-1}+\beta_2 WIN_{i,r-2}+\beta_3 WIN_{i,r-3}+\beta_4 WIN_{i,r-4}+\beta_5 WIN_{i,r-5}+\beta_6 WIN_{i,r-6}+\beta_7 WIN_{i,r-7}+\beta_8 WIN_{i,r-8} \quad (2)$$

$FP_{i,r}$- number ***i*** 's final percentage of bet in round ***r***, that is $P_{i,,r,48}$ (t=48);
$WIN_{i,r-1}$- Dummy for number ***i*** in round ***r-1***; $WIN_{i,r-1}$=0 if number ***i*** was not picked in round ***r-1***; otherwise $WIN_{i,r-1}$=1. It is the same for all the other WIN dummies.

The result in regression (2) shows that not all the winning numbers in past eight rounds have significant influence on betting behavior, so I do the stepwise estimation from regression (2). Finally, I get the result that only winning numbers in the previous four rounds significantly affect betting behavior, as shown in regression (3).

$$FP_{i,r}= 0.0475632 - 0.0048778 WIN_{i,r-1} - 0.0027007\ WIN_{i,r-2} - 0.0014202\ WIN_{i,r-3} - 0.0002903\ WIN_{i,r-4} \quad (3)$$

SE (0.0000791) (0.0001291) (0.0001291) (0.0001291) (0.0001291) adj R^2=0.2436

In the regression, all the coefficients are statistically significant at high level (see Appendix III, Table 2). The negative marks of the WIN dummies' coefficient indicate that bettors underestimated the probability of a number being picked in the present round if it won the game in any of the past four rounds, which gives support to Hypothesis 3 that bettors are subject to gambler's fallacy as in previous research. Moreover, the coefficients show that the effect of gambler's fallacy fades as the win is further to now. There are 22 numbers on which players can bet, which are drawn with equal probability. If bettors are unbiased, all WIN dummies' coefficients should therefore be 0, and the constant α in the regression should be equal to 1/22=0.0455. Each number should be chosen with percentage 4.55%. The fact that the constant is larger is the result of the fact that all coefficients of WIN dummies are negative. In the present round, the percentage of bets on numbers that won the game in the previous round is 0.0475632-0.0048778=0.0426854, which is 93.91% of the normative percentage, i.e., 6.09 % less. And for the number which only won the game in round r-2, this percentage increases to 98.70%, i.e., only 1.30% less, which means the influence of numbers that won in round r-2 is weaker than the ones in round r-1. This is similar to the results found by Clotfelter and Cook (1993) and Terrell (1994), who found that the winning numbers' impact decreases gradually over time. However, if

[11] Here I only analyze whether bettors are subject to gambler's fallacy with final wager data as in prior research. The time serial data will be regressed later. The time pattern for the influence of gambler's fallacy will be shown in 5.2.

the number won both in round r-1 and round r-2, the bet's percentage is 87.97% of the normative percentage (more details in the discussion), which means that a repeated win convinces bettors that this number will not be picked again for some time. However, it will be proved later that if a number was continuously picked four times, people will start believing that this number will be picked again with a higher probability.

In this market, the influence of winning numbers on betting behavior lasts for only four days, which is a much shorter period than for previous results, e.g., sixty days. I propose that this is caused by the fact that the choice set in the "pick 5" game is much smaller (choose 5 numbers out of 22, ratio=1/4.4), compared to past games (e.g., choose 1 number out of 1000, ratio=1/1000), which leads to faster re-picking of the numbers.

Result 1: Bettors are subject to gambler's fallacy. The overall betting behavior is influenced by the past four rounds' winning numbers if only each round's final data is taken into consideration.

Besides winning numbers in past rounds, the homepage simultaneously supplies information of hot and cold numbers among bettors who entered the present round earlier. It is noticeable that "pick 5" is a pari-mutuel game that means the payoff for each is the jackpot divided by the number of winning wagers. As conjectured before, the best strategy in such a type of game is to choose the least popular numbers among others because there will be less people to share the jackpot.

First, I calculate the following percentage:

$$PC_{i,r,T}=(BET_{i,r,t}-BET_{i,r,t-1})/(\Sigma(BET_{j,r,t}-BET_{j,r,t-1})+(BET_{i,r,t}-BET_{i,r,t-1}))\quad (i\neq j) \qquad (4)$$

($BET_{i,r,t}$- $BET_{i,r,t-1}$) - Bets on number ***i*** between time ***t-1*** and ***t*** in round ***r***;
$\Sigma(BET_{j,r,t}$- $BET_{j,r,t-1})$ - Sum of bets on number ***j (i≠j)***between time ***t-1*** and ***t*** in round ***r***;
$PC_{i,r,T}$ – Relative change of bets on number ***i*** between ***t-1*** and ***t*** in round ***r***. T stands for the time slots between t-1 and t, T=1,….,47.

It is already proved that winning numbers in the previous four rounds influence betting behavior. Now I take both winning numbers in the previous four rounds and hot & cold number information into account. First I regress $PC_{i,r,T}$ on winning numbers in the previous four rounds and the rank of number ***i*** at time ***t-1*** in round ***r***.

$$PC_{i,r,T}=0.0415483 - 0.003705\ WIN_{i,r-1}-0.0018319\ WIN_{i,r-2}-0.0008743\ WIN_{i,r-3}-0.0001049\ WIN_{i,r-4}+0.0004575\ RANK_{i,r,t-1} \qquad (5)$$

SE (0.0001021) (0.0000955) (0.000093) (0.0000922) (0.0000918) (6.42e-06) adj R^2=0.0426

$RANK_{i,r,t-1}$-Rank of number ***i*** 's bets at time ***t-1*** in round ***r*** (e.g., the hottest number's rank=22 and the coldest number's rank=1).[12] This information is updated every half hour and there is no

[12] If there is a tie, I give each number the mean rank as follows: $R=\Sigma R_i/n$. R-rank of number; ΣR_i-sum of ranks ignoring the tie; n-the amount of numbers whose bets are equal. For example, the bets for numbers are number 11=20, number 12=23, number 13=26, number 14=26, number15=26, number16=29. If we ignore the tie, the rank for number 13, 14, 15 should be 3, 4, 5. Then, after considering the tie, the rank should be the mean of three ranks

history of such information available.

In regression (3), I get the result that $WIN_{i,r-4}$ influences bettors' behavior with other variables, but in regression (5), this effect is not significant(see Appendix III, Table 3 and Table 6 for details).[13] Therefore, I regress $PC_{i,r,T}$ on all variables except $WIN_{i,r-4}$ and get the following result:

$$PC_{i,r,T}=.0415234 - 0.0037045WIN_{i,r-1} - 0.0018336WIN_{i,r-2} - 0.0008759WIN_{i,r-3} + 0.0004576RANK_{i,r,t-1} \quad (6)$$

SE (0.0000998) (0.0000955) (0.0000929) (0.0000922) (6.42e-06) adj R^2=0.0426

The result in regression (6) confirms that both gambler's fallacy and guidance effect[14] affect bettors' behavior (see Appendix III, Table 4 and Table 6 for details). The coefficients in the regression show that winning numbers still have a negative effect on betting behavior and the number's rank has a positive effect. This rejects what was suggested in Hypothesis 2. For example, if a number was picked in the past round ($WIN_{i,r-1}$=1) and has the highest rank now ($RANK_{i,r,t-1}$=22), it is still preferable than in the unbiased condition (0.0415234-0.0037045+22*0.0004576
=0.0478861, which is larger than 0.0454545), which means the influence of hot and cold number information is stronger than gambler's fallacy in such a condition. This indicates that hot and cold number information shifts part of the bettors' attention from winning numbers in previous rounds to possible winning numbers in the present round, and the higher the rank of a number is, the stronger the effect is.

Because bettors' choices are subject to gambler's fallacy, which means the rank is partly the result of prior winning numbers, I add interaction items between WIN dummies and RANK into regression (6), as shown in regression (7).

$$PC_{i,r,T}=.0411476-.0030395WIN_{i,r-1} -.0013974WIN_{i,r-2} -.0005271 WIN_{i,r-3} + .0004878 RANK_{i,r,t-1} -.0000726 WIN_{i,r-1}*$$

SE (0.0001201) (0.0001735) (0.0001781) (0.0001827) (8.39e-06) (0.000016)

$$RANK_{i,r,t-1} -.0000423WIN_{i,r-2}* RANK_{i,r,t-1}-.0000306 WIN_{i,r-3}* RANK_{i,r,t-1} \quad (7)$$

(0.000015) (0.0000148) adj R^2=0.0427

The result of regression (7) supports the idea that hot and cold number information interacts with winning numbers in past rounds (see Appendix III, Table 5 and Table 6 for details). Take the case where the winning number from the past round is the current hottest number (and therefore has rank=22). Then the predicted net effect equals 0.0417476-.0030395 + 22*.0004878 - 22*.0000726=0.456332955, which is larger than 0.0454545. It reveals that the current hottest number is to be chosen, even if it was picked in the previous round. This is also the case for hottest numbers that won the game in even earlier rounds (see the discussion for more details).

ignoring the tie that is (3 + 4 + 5)/3=4.

[13] Here, the data I use is different from the one in regression (3). In regression (3), I used the final data here. But here I use the full data. Therefore, in regression (5), the result is different from the one in regression (3). The test in regression (3) is used as a comparable test with prior research only.

[14] I define the effect of hot and cold number information as "guidance effect", since it gives bettors guidance on which number to choose in the game. It could also be defined as the reverse of extreme type of favorite long-shot bias.

Here, Hypothesis 2 is rejected by even stronger evidence. The more popular a number was among earlier entrants, the more people pick it afterwards. This could partly be caused by the cultural attachment to special numbers, which will be discussed later.

Result 2: The hotter a number was among bettors who entered the game earlier, the more bettors who entered later picked this number.

In regression (6) and (7), while the correlation between betting behavior and winning numbers is negative, the relationship is positive between $RANK_{i,r,t}$ and $PC_{i,r,T}$. To my surprise, the influence of hot and cold number information is on the opposite of what the best strategy suggests for parimutuel game. As the definition of guidance effect shows, later entrants followed the direction given by prior bettors in the same round instead of following the best strategy. This phenomenon is similar to the so-called "wisdom of crowds". In the "wisdom of crowds" situation, people have both private information and information about what others did. As more people made the same choices, it becomes more convincing that something important will happen, such as the looking sky story described in Surowiecki's book. [15] In such a situation, people prefer to make a decision on what others did rather than deciding on the basis of their own information. But there is a slight difference between the wisdom of crowds and guidance effect which leads me to prefer to use the guidance effect, i.e., in a lottery game, people know for sure that there is no "private information", because the winning probability is objectively equal for all numbers and nobody could influence the lottery's draw. Therefore, what others did does not provide any hint about something that later entrants do not know. They cannot use such information to convince themselves that what others did will happen with a higher probability.

One more thing worth mentioning is that I run the regression with a time lag as well, which tests whether the previous few slots of hot and cold number information jointly influence the selection of a number combination. The result shows that guidance effect does not have any lag consequence. I give out the following possible reasoning for future research: first, the rank did not change dramatically within a single round of the game. Therefore, it does not make sense to waste time on observing the information for a relatively long time (e.g., more than half an hour). Second, since there is no history of such information, it consumes a lot of effort to record it. So, it is easier to rely on one piece of such information. Third, it is about different types of bettor. Pleasure bettors would like to enter the game as early as possible. Their focus is to enjoy the waiting time until the end of the game more than seeing what happens at the end of the game. Therefore, they would be more likely to pick some numbers they love than rely on this information. As for professional bettors, they only need to look up the information when the game is nearly over, so as to choose numbers that could give them more possible earnings. It is not necessary to observe the information for a long while. Therefore, there is no significant time lag of the influence of hot and cold number information within a round of the lottery game.

It is important to note that the influence of hot and cold number information is different with the

[15] This is an experiment conducted by social psychologists Stanley Milgram, Leonard Bickman, and Lawrence Berkowitz. They put an increasing number of people on a street corner and let them look up at the sky. They found that the more people were standing there at the beginning, the more pedestrians would join the group and look up at the sky as well.

hot-hand fallacy. Hot-hand fallacy refers to a belief in positive autocorrelation of a non-auto correlated random sequence. For example, bettors believe that a lottery ticket's winning probability is higher if they buy the ticket from a shop that has sold a lot of winning tickets; this is defined as "lucky store effect" (Guryan and Kearney, 2008). In this case, the reference is the shop's performance in prior events. If the performance was good in the past, people increase their estimation of the probability that a good performance will occur again in future in such a shop. But what I investigate here is the influence of information that pops up simultaneously with the event. Therefore, there is no indicator to show how well the information will finally work in the present round. People would argue that bettors might hold the idea that the information performed well on predicting in previous rounds of the game. Hence, they believe in the information they see in the present game. Therefore, this phenomenon is referred to as a hot-hand fallacy. But this argument is actually not supported by empirical evidence since the information is not that accurate. In order to know the degree of information accuracy, I calculate it in the following way:

$$\boldsymbol{a} = (R_{win}-R_{min})/(R_{max}-R_{min}) \quad (7)$$

$\boldsymbol{a}$-accuracy of hot and cold number information;

R_{win}-mean rank of winning numbers' $BET_{i,t,r}$;

R_{min}-the minimum mean rank of 5 numbers (the mean rank of 5 coldest/least popular numbers), R_{min} =3;

R_{max}-the maximum mean rank of 5 numbers (the mean rank of 5 hottest/most popular numbers), R_{max} =20.

If $\boldsymbol{a}$=1, it means this information is 100% accurate. The smaller the $\boldsymbol{a}$ is, the less accurate the information is. The result is $\boldsymbol{a}$ =0.48449, which indicates that the information did not give out accurate prediction on winning numbers. If hot-hand fallacy is the explanation, bettors will not follow the information since it is not actually accurate.[16]

Another difference worth noticing is that in gambler's fallacy or hot-hand fallacy, the information of past performance will not influence present payoff because the two events are independent, but the hot and cold number information could influence each wager's payoff. Although the information is in the past, because it is still the information for the present round, the influence is different with the information that results in gambler's fallacy or hot-hand fallacy.

5.2. Time pattern of biases

Now I switch the attention to the time pattern of biases in lottery game. To the best of my knowledge, this is the first research on bias time pattern in lottery game. This will give us a direct picture of how the biases unfold across time in the game and supply more information on how bettors behave.

In order to see the effect of winning numbers on the bet, I first calculate the mean of $PC_{i,r,T}$ by separating $WIN_{i,r-1}$=0 or $WIN_{i,r-1}$=1(gambler's fallacy). Since in the regressions I concluded that

[16] It is possible the case that bettors had illusion about the accuracy of the hot and cold number information which leads them subject to hot-hand fallacy. Or people hold the idea that prior entered bettors have the power to control the winning numbers or predicting the winning numbers which actually is an illusion.

the winning numbers in the past round had the strongest effect, I first show the time pattern of the gambler's fallacy caused by numbers won in the past round.

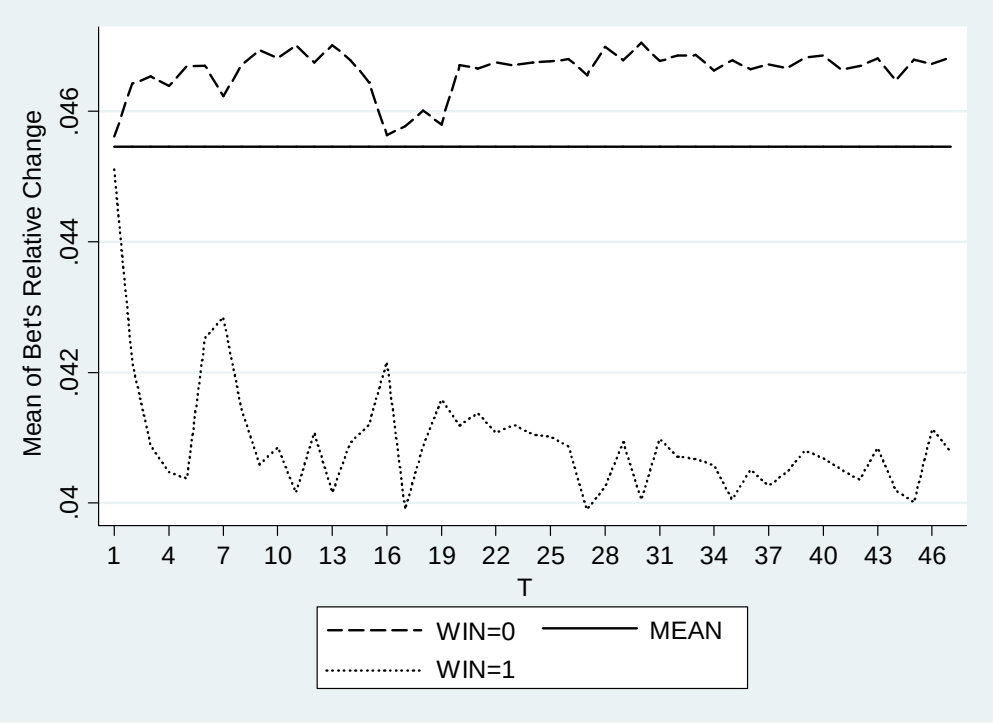

Figure 3 Time Pattern of the Gambler's Fallacy (1)

Figure 3 displays the time pattern of the influence of the previous round's winning number. The solid line is the mean (which is equal to 1/22=0.0455) of the bet's relative change if there is no bias. The lower curve (dash line) shows the mean of $PC_{i,r,T}$ over time if number i was a last-round winning number, which indeed is the time pattern of the gambler's fallacy. The upper curve (dotted line) is the mean of $PC_{i,r,T}$ across time if number i was not a winning number in the previous round. From t=2 to t=6, the relative bet's change for numbers that won the game in the last round (WIN=1) decreases sharply, which indicates that the proportion of earlier bettors who were subject to gambler's fallacy increased fast. From t=5, this trend becomes smooth, which shows that the effect of gambler's fallacy increases, but in a relatively slower speed than before. The dotted line above the average displays that more bettors choose numbers that were not winning numbers in the last round (WIN=0) and $PC_{i,r,T}$ is pretty stable across time. The difference of aptness towards numbers won the game in the last round and numbers that did not win in the last round became gradually salient across time in a round of the game. The figure captures the fact that bettors who enter the game later are subject to more influence by numbers that won in the last round than bettors who entered earlier. This result rejects what Hypothesis 4 suggested. Actually, the influence of winning numbers in the previous round becomes stronger over time.

If follows what was discussed in 3.1. Figure 3 could be interpreted as earlier entrants who are possibly pleasure bettors and are influenced less by the last round's winning numbers than later entrants who are mostly supposed to be professional bettors. This means professionals are possibly not that professional, which could be tested with individual data in future.

Result 3: The later a bettor enters the market, the more serious he is subject to gambler's fallacy.

As shown in regression (3), the previous four rounds' winning numbers have an influence on betting behavior and it should be noticed that a number could possibly win in more than one round. Therefore, in Figure 4, I separately show the time pattern of gambler's fallacy by the frequency with which a number won the previous game.

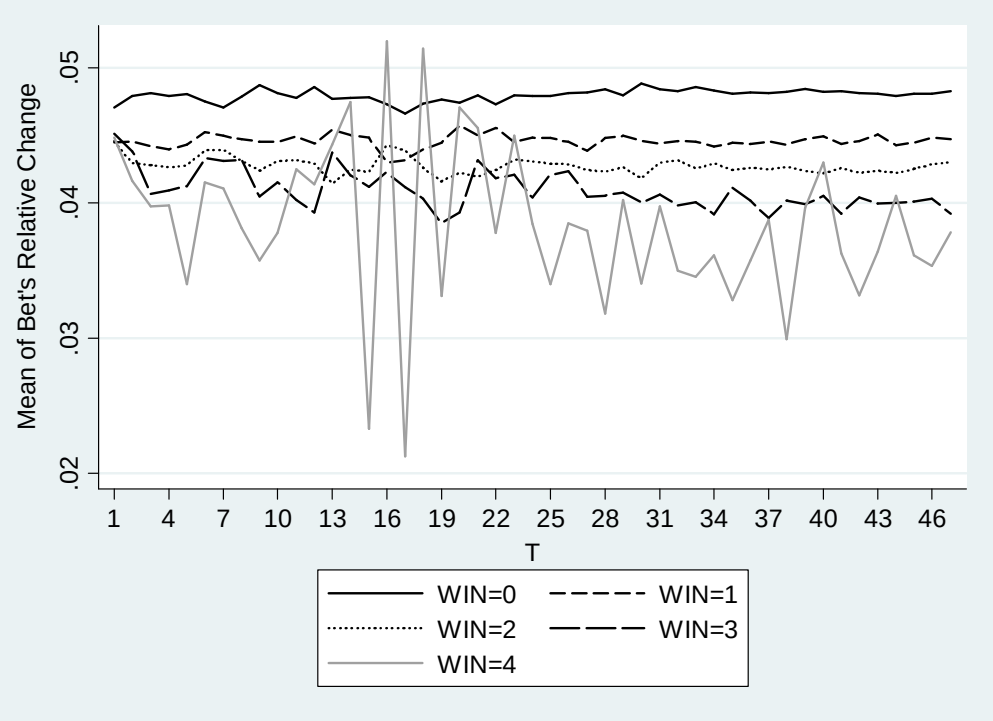

Figure 4 Time Pattern of the Gambler's Fallacy (2)

In Figure 4, each line stands for the time pattern of $PC_{i,r,T}$ by how many times the number won the game in the past four rounds. For example, the black dotted line marked with WIN=2 shows the mean of the bet's relative change for numbers that had been drawn twice in the past four rounds (but may not be continuously picked). The figure reflects the fact that the more times a number won the game in the past four rounds, the fewer bettors chose this number in the present round. The time pattern is pretty stable across time for numbers that have won the game less than 4 times. It is worth noting that people's attitudes towards the numbers that won in all previous four rounds fluctuated considerably. Sometimes, it even went beyond the $PC_{i,r,T}$ of numbers that never won the game in the past four rounds. Consistent with prior research (Jørgensen etc., 2011), the bias switches to hot-hand fallacy for numbers that always won the game.

Result 4: the more times (⩽3) a number won the game in the previous four rounds, the fewer bettors picked it, and the influence is pretty stable over time.
If the number is on streak, i.e., won in all prior four rounds, the effect switches from gambler's fallacy towards hot-hand fallacy.

In order to see how bettors integrate the information of winning numbers in the previous four rounds as a reference, I distinguish the mean of the bet's relative change ($PC_{i,r,T}$) for numbers that did not win the game in any of prior four rounds with numbers that won the game at least once in the past four rounds (Figure 5).

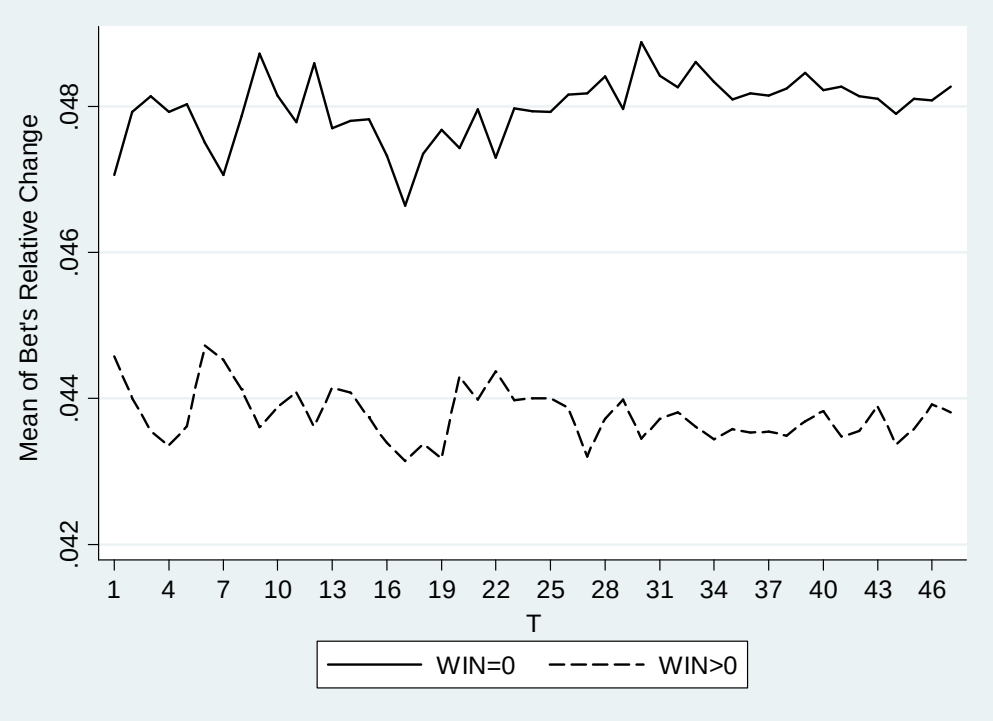

Figure 5 Time Pattern of the Gambler's Fallacy (3)

Figure 5 displays the time pattern of $PC_{i,r,T}$ between two types of number. From the lines it is obvious to see that the time pattern of the gambler's fallacy is pretty stable, as in Figure 4. The discrepancy of $PC_{i,r,T}$ between winning numbers and non-winning numbers starts at the beginning of the game and it did not change too much across time.The difference is statistically significant (paired t-test, p=0.00, two-tailed). This indicates that when we think about whether the number was picked in the past 4 rounds, the attitude towards such information is pretty uniform among bettors, no matter when they entered the game. Combining Figure 3 with Figure 5, one may find that people who joined the game later considered more about what numbers had been picked in the last round, while earlier entrants tended to balance the influence of winning numbers in the past four rounds more.

Now I switch to the time pattern of the guidance effect. The website lists the top 10 most popular numbers (hot numbers) and the top 10 least popular numbers (cold numbers) at the same time. But actually, bettors only need to choose a 5-number combination. Therefore, I provide the time pattern of the top 10 hot and cold numbers, as well as the top 5 hot and cold numbers' influence at the same time (Figure 6).

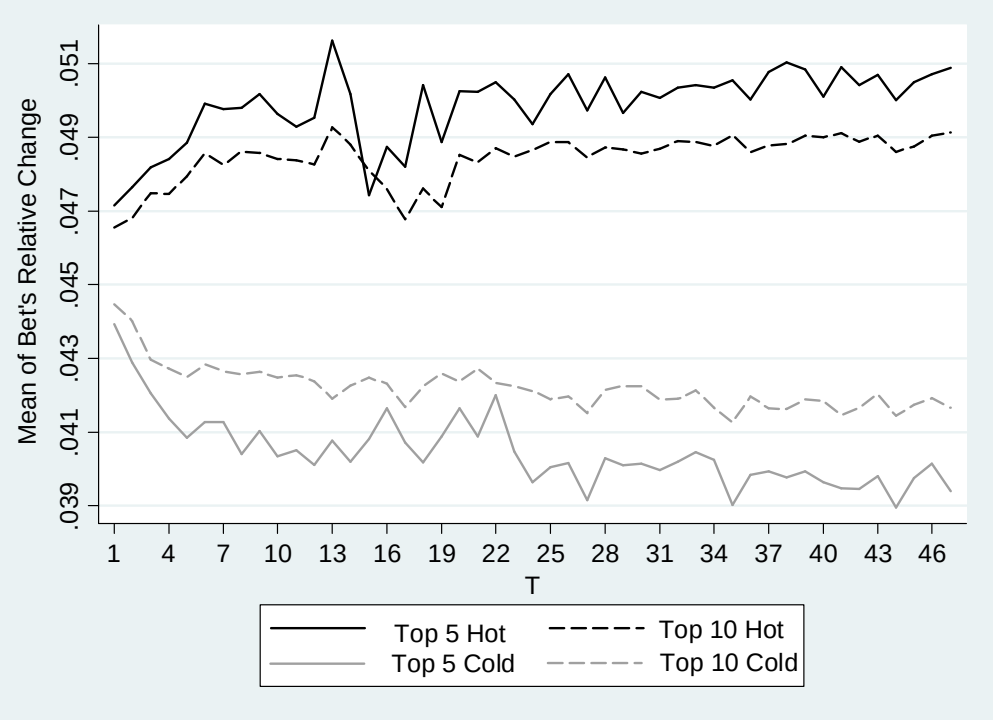

Figure 6 Time Pattern of the Guidance Effect

Figure 6 displays the mean of the bet's relative change ($PC_{i,r,T}$) across time for the top 5 hot, top 10 hot, top 10 cold, and top 5 cold numbers. Again, it shows that hot and cold number information does affect bettors' behavior, as found in the regressions, and that the influence has become stronger across time. At the beginning of the game, there is already a slight difference of the bet's relative change between hot and cold numbers. The discrepancy grows gradually bigger when the information is composed by more bettors' choices. For hot numbers, the top 5 numbers' $PC_{i,r,T}$ has a faster increase than the top 10 numbers' one, which means the hotter a number was among prior bettors, the more people will pick it. The difference of $PC_{i,r,T}$ between top 5 and top 10 hot numbers is statistically significant (paired t-test, p=0.00, two-tailed). Similarly, for cold numbers, the colder a number was, the fewer people preferred it. This difference is significant as well (paired t-test, p=0.00, two-tailed). The time pattern of the guidance effect proves that hot and cold number information has a striking influence on bettor's behavior. The later a person joins the game, the more he is affected by the choices of prior bettors. This suggests that people concern more about what the others have already chosen rather than how many winners will share the jackpot,. They believe that, for a number, the more people chose it, the higher the probability is that this number will win the game. This is consistent with what Mannes (2009) found in lab experiments that concluded that people thought that a larger group's decision would be more accurate than a smaller group's decision. Hypothesis 6 is supported, but the influence is the opposite of what is suggested in Hypothesis 2.

Result 5: The later a bettor enters the game, the heavier he is influenced by the hot and cold number information.

5.3. Number Preferences

From Figure 6 we could see that, at the beginning of the game, there is already a discrepancy between the bet's relative changes of top hot and cold numbers. By intuition, there should be no such big difference of bets among numbers, since in the first half hour the sample size of hot and cold number information is pretty small. But this is not the case in the market. There must be some other reasons that influence bettors' choices especially at the beginning of the game. This is what I want to test next, i.e., preference towards special numbers. For thousands of years, Chinese have been heavily addicted to special numbers, especially businessman who would like to pay more for getting a "good" number, e.g., a special cell phone number, car license number, etc. In Chinese number culture, 8 is an important number which could bring big fortune, while 14 is an unlucky number which people try to avoid. Now, I mainly introduce the influence of Chinese number culture which could partly explain the issue we observed in Figure 6.

In order to illustrate, first, I calculate the mean of ranks for each number in the following way: for each round r and time t, I calculate the increase of bets between time ***t-1*** and ***t*** and sort the changes. Then I give the number with the highest bet's increase with rank=22 and so forth. If there is a tie, I just use the average rank instead. Finally, I calculate the mean of ranks for each number across all data[17]. The result is shown in Figure 7.

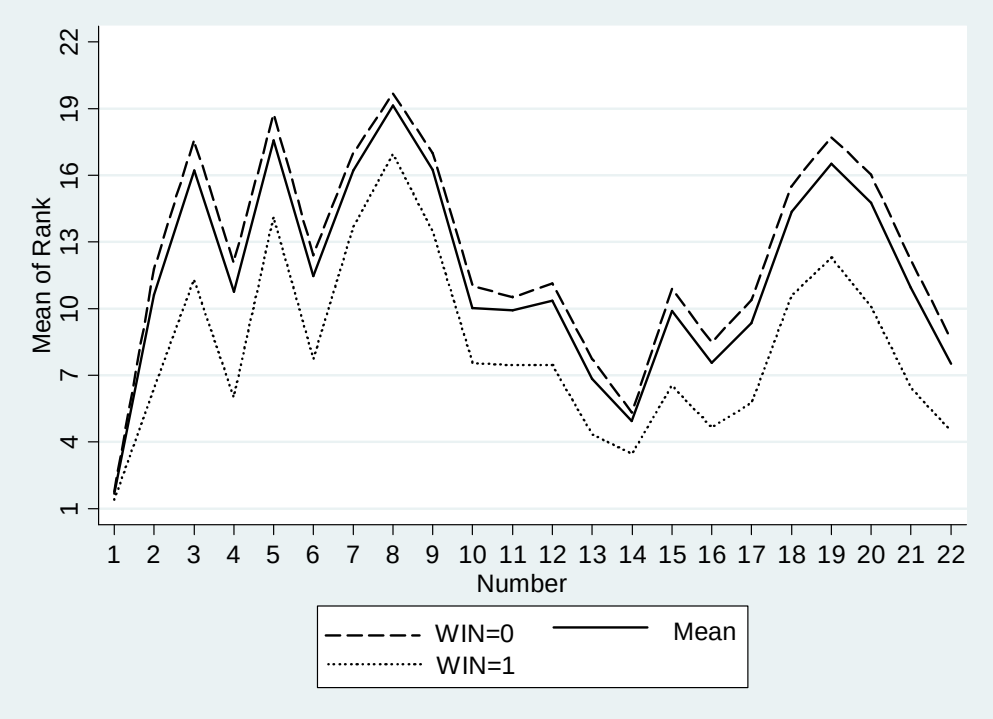

Figure 7 Mean of Numbers' Rank

In Figure 7, the solid line is the mean of overall ranks for each number. The dash line shows the mean of ranks for each number if it did not win the game in prior one round. And the dotted line indicates the mean of ranks for each number if it won the game in the previous round.

[17] This rank is different with RANK dummy in regressions. In regression, RANK stands for the rank of aggregated bet from the beginning until time t, while here the rank of the bet's change in each time slot T is shown.

It is clear in Figure 7 that once the number has won the game in the previous round, its rank decreases in the current round, but in different degrees for different numbers. And for numbers that were not drawn on the previous day, the ranks have a slight increase. From the perspective of preferences towards numbers, the solid line in Figure 7 clearly displays that bettors are edge-number-averse, that is, players are reluctant to choose edge numbers 1 and 22. Meanwhile, number 8 is the most popular number since it has the highest rank in all three conditions. Number 8's rank is significantly higher than number 5, which is the second-most popular one among numbers (t-test, p=0.00, two-tailed). Except edge numbers, number 14 has the lowest rank, which indicates bettors are not fond of it. The difference of rank between numbers 8 and 14 is 14.2. The dotted line shows the mean of ranks if the number won the game in the previous round. This is an indicator of gambler's fallacy on each number. Compared to other numbers, number 8 still has the highest rank, although it is lower than the other two conditions, while number 14 and edge numbers have lower ranks among numbers. Especially number 1 always has the lowest rank no matter whether it has won the game in the previous round or not. The comparison of ranks proves Hypothesis 7.

Besides the comparison of mean ranks, I also explore how the influence of special numbers develops across time (Figure 8). It demonstrates clearly in Figure 7 that the edge numbers 1 and 22 have very low ranks; number 1's rank in particular has nearly no change in any conditions. I refer to this phenomenon as "edge number aversion".

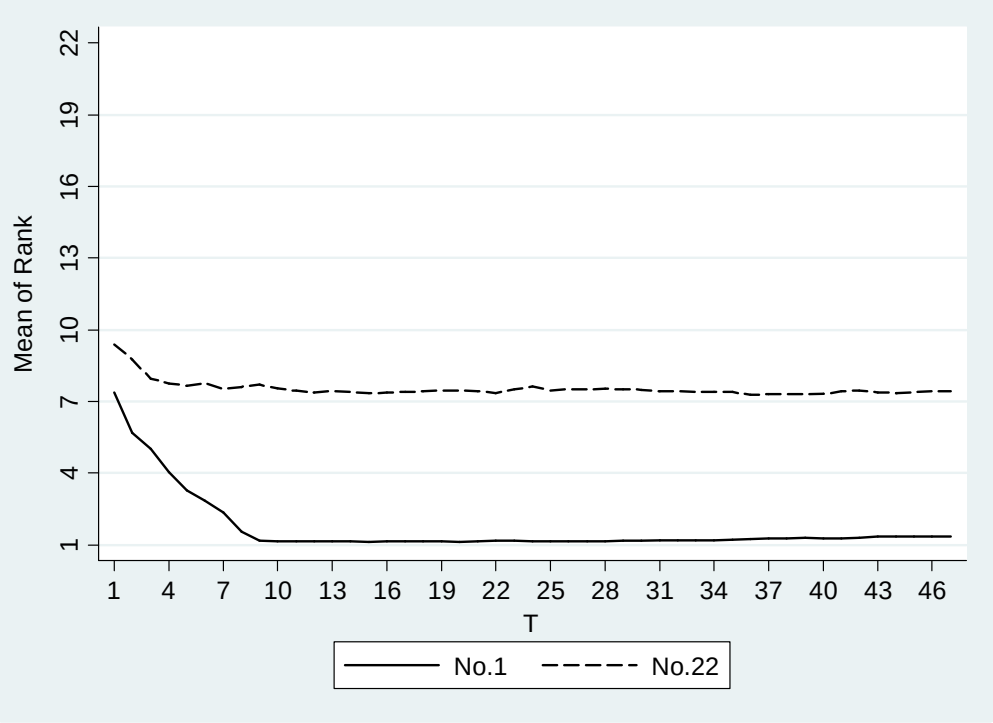

Figure 8 Time Pattern of Edge Numbers' Rank

Figure 8 displays the mean rank of no. 1 and no. 22 across time. It is obvious that from the beginning of the game, bettors already do not like these two numbers, since their ranks are lower

than average, which is equal to 11.5. From Figure 8, it could be read that no.1's rank declines very fast across time and falls to the bottom at t=10. The rank remains there from t=10 until the end of the game. For the other edge number, no.22, this effect is a little bit weaker, but as stable as the time pattern of number 1 from t=4 until the game is over. There is no special cultural reason that people do not like edge numbers. This is a question for further research.

Now we turn to the time pattern of ranks for the lucky number 8 and the unlucky number 14, as shown in Figure 9.

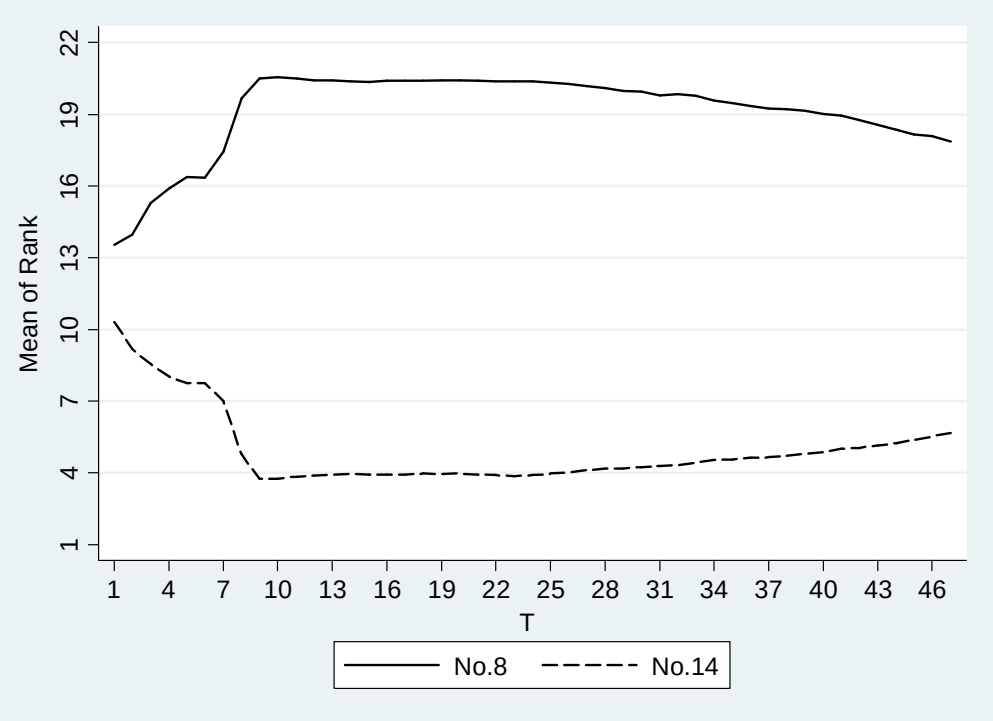

Figure 9 Time Pattern of Lucky and Unlucky Numbers

At the beginning of the game, bettors start showing their preference towards the lucky number 8 and not towards the unlucky number 14. The lucky number 8, which has a higher rank than average, holds this position from the very beginning of the game. This advantage goes up until t= 9 and does not fall too much afterwards until the end of the game. On the other hand, the unlucky number 14's initial rank is already lower than average and drops very fast until t=9. This tendency does not relieve much from then on. First, Figure 9 partly explains why, in the initial part in Figure 6, there is already a gap between hot and cold numbers even though the information is composed by a pretty small sample size. Meanwhile, the time pattern in Figure 9 proves that the bettors who entered the game later have a stronger attachment to special numbers than people who bet earlier.

6. Conclusion

This paper empirically studies anomalies in the Chinese online gambling market. The research is inspired by the hot and cold number information which lists the most and least popular numbers

among earlier-entered bettors on the webpage of the "pick 5 out of 22" lottery game. In the game, people do not follow the best strategy of pari-mutuel game, which is to choose the least popular numbers over others so as to share the jackpot with fewer winners. Instead, they go with the directions given by the information and choose the most popular numbers over others. This is the first main finding of this research which has not been documented before. Besides the guidance effect, consistent with previous research, bettors are subject to the gambler's fallacy as well. However, compared to other research, the gambler's fallacy lasts much shorter in this game.

The second contribution of this paper is the time pattern of anomalies found in the market. As shown in the figures, if only one winning number from a previous round is taken into consideration, the gambler's fallacy grows stronger over time, which indicates that professional bettors are influenced more by prior winning numbers than pleasure bettors if both types of bettor are in the market. Moreover, if one takes numbers into account that were picked in the previous four rounds, the gambler's fallacy becomes stable, which reveals that later entrants put more weight on the winning number from the previous round, while earlier entrants prefer to combine the four previous rounds' winning numbers. On the other hand, guidance effect is increasingly bigger over time as well. The more popular a number has been, the more people like to choose it; and the less popular a number has been, the less people like to choose it. Bettors who join the game later are influenced more, which could be the result of the increasing sample size of the information.

From the time pattern of the guidance effect, we could deduce that, at the beginning of the game, there is already a discrepancy between hot-number and cold-number bets, which is surprising since the information is drawn from a very small sample size. Therefore, I investigate whether such a phenomenon is caused by Chinese number culture and confirm that Chinese number culture does influence people's choice of number combination. From the perspective of average ranks, number 8 always has the highest rank under any comparable conditions, since it is a lucky number in China, while number 14 has a pretty low rank, since it is an unlucky number. Besides, from the perspective of time pattern, number 8's rank is above the average from the beginning of the game until the end, while number 14's rank drops dramatically. Moreover, Chinese bettors are averse to edge numbers, which means they do not like to choose the edge numbers 1 and 22. This phenomenon is subject to future investigation.

The biases found in this paper give hints to both gambling authorities and operators, e.g., lottery ticket sellers. The main attraction of such game for gamblers is that, with an extremely low probability of winning, people could earn a huge amount of money once they get the draw. Although actually it is not possible to forecast the draw combination, people could get hints from others from the hot and cold number information, which could increase the volume of sales. This also adds fun for bettors who join the game for pleasure. It would be interesting to test whether the information about the behavior of others has a similar influence on other markets as well, e.g., on the stock market. Because the stock-market trade has no real end, as each day's final draw in the lottery market does, the influence of the information could be more powerful if there is any effect.

Discussion

1. Gambler's fallacy in regression (3)

In regression (3), I regress the final percentage of the bet on numbers that were picked in the previous 4 rounds, and get the following result.

$$FP_{i,r} = 0.0475632 - 0.0048778WIN_{i,r-1} - 0.0027007\ WIN_{i,r-2} - 0.0014202\ WIN_{i,r-3} - 0.0002903\ WIN_{i,r-4} \quad (3)$$

First, I discuss the effect by how many times a number was picked continuously from round r-1. Suppose that only $WIN_{i,r-1}=1$, that is, the number was a winning number in only one previous round, the net effect should be equal to 0.0475632-0.0048778=0.0426854, which is 93.91% of the normative percentage, i.e., 6.09% less. The smaller the net effect is, the stronger the gambler's fallacy is. The other situations are calculated in a similar way and listed in Table I.

Table I Net Effect of Continuous Winning

Continuous Win	Net Effect	Normative Percentage (%)
1	0.0426854	93.91
2	0.0399847	87.97
3	0.0385645	84.84
4	0.0382742	84.20

Table I displays the fact that the more times a number won the game continuously before, the less people choose this number in the current round, but the effect fades a lot as time passes, e.g., if a number was picked in both round r-1 and r-2, i.e., continuous win=2, the net effect is 0.0399847. If a number was picked once more in round r-3, this net effect decreases to 0.0385645. The difference of the normative percentage is 87.97%-84.84%=3.13%, while if a number was picked once more in round r-4, the change of the normative percentage is 84.84%-84.20%=0.64% which is much smaller than 3.13%. It is obvious that, each time when a number won the game again, the weight that people place on this number decreases. Therefore, although the effect gets stronger, the increase of the speed of effect slows down. My explanation is that the more often a number was picked before, the higher the possibility is that this number will not be picked again. And the more recently a number has won the game, the lower the possibility is that this number could be the winner one more time.

2. Anomalies in Regression (7)

Now I turn to regression (7) which reports the influence of winning numbers in the previous three rounds and numbers' present ranks with interact items.

$$PC_{i,r,1} = .0411476 - .0030395WIN_{i,r-1} - .0013974WIN_{i,r-2} - .0005271\ WIN_{i,r-3} + .0004878\ RANK_{i,r,t-1} - .0000726\ WIN_{i,r-1} * RANK_{i,r,t-1} - .0000423WIN_{i,r-2} * RANK_{i,r,t-1} - .0000306\ WIN_{i,r-3} * RANK_{i,r,t-1} \quad (7)$$

It makes sense that there is an interaction effect between WIN dummies and number rank in the current round because the numbers that bettors choose are affected by previous winning numbers, which would partly cause the change of ranks. This is clearly proved in regression (7). Suppose a

number won the game in *r-1* and its rank is 22 (the most popular number in the present round, e.g., number 8). The percentage of the bet's increase in period T is 0.0417476-.0030395 + 22*.0004878 - 22*.0000726= 0.456332955, which is larger than 0.454545. This means, although the number won the game before, because it was the most popular number among bettors in the present round, it still has a 0.178% net increase. Table II lists all the possible situations in which the number won the game only once before.

Table II Difference between Net Effects (WIN=1)

Rank	WINi,r-1=1	WINi,r-2=1	WINi,r-3=1
22	0.001787955	0.004096655	0.00522435
21	0.001372755	0.003651155	0.00476715
20	0.000957555	0.003205655	0.00430995
19	0.000542355	0.002760155	0.00385275
18	0.000127155	0.002314655	0.00339555
17	-0.000288045	0.001869155	0.00293835
16	-0.000703245	0.001423655	0.00248115
15	-0.001118445	0.000978155	0.00202395
14	-0.001533645	0.000532655	0.00156675
13	-0.001948845	8.71545E-05	0.00110955
12	-0.002364045	-0.00035835	0.00065235
11	-0.002779245	-0.00080385	0.00019515
10	-0.003194445	-0.00124935	-0.00026205
9	-0.003609645	-0.00169485	-0.00071925
8	-0.004024845	-0.00214035	-0.00117645
7	-0.004440045	-0.00258585	-0.00163365
6	-0.004855245	-0.00303135	-0.00209085
5	-0.005270445	-0.00347685	-0.00254805
4	-0.005685645	-0.00392235	-0.00300525
3	-0.006100845	-0.00436785	-0.00346245
2	-0.006516045	-0.00481335	-0.00391965
1	-0.006931245	-0.00525885	-0.00437685

Remember that, if people are unbiased, the net effect $_{PCi,r,t}$ should equal 0.454545. But if people are influenced by information on the webpage, e.g., numbers that have won the game before or hot and cold number information, the net effect should be different from 0.454545. Compared to 0.454545, if the net effect is larger, it indicates that the guidance effect is stronger than the gambler's fallacy. Otherwise, the gambler's fallacy is weaker than the guidance effect. Table II shows the difference between net effect with biases and without biases. The first column is the possible rank a number has, i.e., from 1-22. Columns 2, 3, and 4 show the difference of net effect if a number was picked only in round r-1 or r-2 or r-3. First, place attention on each row. The longer ago it is that a particular number has been picked in the game, the bigger the difference between net effects is, i.e., the stronger the guidance effect is. Especially for Rank 11-17, at the beginning, the competing result between gambler's fallacy and guidance effect is that the gambler's fallacy wins. Later, however, the result switches to positive, which means guidance

effect has become stronger than gambler's fallacy.

Now we turn to each column. The numbers change from positive to negative at Rank=17 in column 2, Rank=12 in column 3, and Rank=10 in column 4. The switching between positive and negative gets slower and slower. This means the competitive nature of the guidance effect becomes stronger and stronger. Therefore, the positive situation could last longer, i.e., when the number was picked in round r-2, it lasts until Rank=1. But when the number was picked in round r-3, it could last until Rank=11.

Appendix I Dataset[18]

The round is counted with a five-digit number. The first two digits stand for the year, e.g., "09" means "Year 2009". The last three digits show the date in the year (e.g., Feb. 24th is the 55th day of the year. Therefore, the last three digits are 055).

09294	09295	09296	09297	09298	09299	09300	09301	09302	09304
09305	09306	09307	09308	09309	09321	09323	09324	09325	09326
09327	09328	09329	09330	09331	09332	09333	09335	09336	09337
09338	09340	09341	09346	09347	09348	09349	09350	09351	09352
09353	09354	10012	10013	10015	10016	10017	10018	10020	10021
10022	10023	10024	10025	10026	10027	10028	10029	10030	10031
10032	10034	10035	10036	10037	10038	10039	10040	10041	10042
10043	10045	10046	10052	10053	10055	10056	10059	10060	10061
10062	10063	10064	10065	10066	10067	10068	10069	10070	10071
10072	10073	10074	10075	10076	10078	10079	10080	10081	10082
10083	10084	10085	10086	10087	10088	10089	10090	10091	10092
10093	10094	10095	10096	10097	10098	10100	10101	10102	10103
10104	10105	10106	10107	10108	10109	10112	10113	10114	10115
10116	10117	10118	10119	10120	10122	10123	10124	10125	10126
10128	10129	10130	10131	10132	10133	10134	10135	10136	10139
10140	10141	10142	10143	10144	10145	10146	10147	10148	10152
10153	10154	10155	10158	10159	10160	10161	10162	10163	10164
10168	10169	10170	10171	10172	10173	10174	10175	10176	10177
10178	10179	10180	10181	10182	10183	10184	10185	10186	10187
10188	10189	10190	10191	10192	10193	10194	10195	10196	10197
10198	10199	10200	10201	10202	10203	10204	10205	10206	10207
10209	10210	10211	10212	10213	10214	10215	10216	10217	10218
10219	10220	10221							

[18] For technical reasons, there are data missing for some rounds.

Appendix II Screenshot for Hot and Cold Number Information

1. Hot Number;
2. Cold Number;
3. Bet Frequency;
4. Ratio (%);
5. Number.

Appendix III

Table 1 Summary Statistic of Bet on Each Number (obs.=400)

Number	Mean	Std.Dev.	Min	Max	Number	Mean	Std.Dev.	Min	Max
1	934.345	209.9722	493	1692	12	1181.618	274.1504	602	1963
2	1180.922	256.9745	570	1896	13	1159.790	245.3461	644	1851
3	1311.67	270.2174	659	2095	14	1121.815	235.3926	601	1865
4	1206.015	248.6293	634	1951	15	1200.922	260.1644	590	1856
5	1326.698	287.5438	714	2088	16	1164.975	255.5573	586	1936
6	1219.435	244.6255	661	1885	17	1234.37	251.6334	678	1939
7	1329.995	269.7681	691	2109	18	1277.603	266.7876	674	2060
8	1342.458	271.089	758	2113	19	1337.063	263.187	729	2220
9	1295.805	256.2953	685	1997	20	1292.348	285.7973	655	2121
10	1187.902	274.4139	619	1967	21	1217.608	277.9737	594	1975
11	1199.08	246.1386	625	1990	22	1127.025	284.9587	515	1825

Regression (3) $FP_{i,r}=\alpha + \beta_1 WIN_{i,r-1}+\beta_2 WIN_{i,r-2}+\beta_3 WIN_{i,r-3}+\beta_4 WIN_{i,r-4}$

Table 2 Result of Regression (3)

```
      Source |       SS       df       MS              Number of obs =    6292
-------------+------------------------------           F(  4,  6287) =  507.37
       Model |  .037363896     4  .009340974            Prob > F      =  0.0000
    Residual |   .11574652  6287   .00001841           R-squared     =  0.2440
-------------+------------------------------           Adj R-squared =  0.2436
       Total |  .153110416  6291  .000024338           Root MSE      =  .00429

------------------------------------------------------------------------------
finalpercent |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          β1 |  -.0048778   .0001291   -37.78   0.000    -.0051309   -.0046247
          β2 |  -.0027007   .0001291   -20.92   0.000    -.0029538   -.0024476
          β3 |  -.0014202   .0001291   -11.00   0.000    -.0016733   -.0011671
          β4 |  -.0002903   .0001291    -2.25   0.025    -.0005434   -.0000372
           α |   .0475632   .0000791   601.23   0.000     .0474081    .0477183
------------------------------------------------------------------------------
```

Regression (5) $PC_{i,r,T}=\alpha+\beta_1 WIN_{i,r-1}+\beta_2 WIN_{i,r-2}+\beta_3 WIN_{i,r-3}+\beta_4 WIN_{i,r-4}+\gamma\ RANK_{i,r,t-1}$

Table 3 Result of Regression (5)

```
      Source |       SS       df       MS              Number of obs =  229196
-------------+------------------------------           F(  5,229190) = 2039.38
       Model |  3.45427476     5  .690854952            Prob > F      =  0.0000
    Residual |  77.6398405229190  .000338758           R-squared     =  0.0426
-------------+------------------------------           Adj R-squared =  0.0426
       Total |  81.0941153229195  .000353821           Root MSE      =  .01841

------------------------------------------------------------------------------
  percentdis |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          β1 |   -.003705   .0000955   -38.78   0.000    -.0038922   -.0035177
          β2 |  -.0018319    .000093   -19.71   0.000    -.0020141   -.0016497
          β3 |  -.0008743   .0000922    -9.49   0.000     -.001055   -.0006937
          β4 |  -.0001049   .0000918    -1.14   0.253    -.0002848     .000075
           γ |   .0004575  6.42e-06    71.27   0.000     .0004449      .00047
           α |   .0415483   .0001021   406.82   0.000     .0413482    .0417485
------------------------------------------------------------------------------
```

Regression (6) $PC_{i,r,T}=\alpha+\beta_1 WIN_{i,r-1}+\beta_2 WIN_{i,r-2}+\beta_3 WIN_{i,r-3}+\gamma\ RANK_{i,r,t-1}$

Table 4 Result of Regression (6)

Source	SS	df	MS		
Model	3.45383249	4	.863458121	Number of obs	= 229196
Residual	77.6402828229191	.000338758		F(4,229191)	= 2548.89
				Prob > F	= 0.0000
				R-squared	= 0.0426
				Adj R-squared	= 0.0426
Total	81.0941153229195	.000353821		Root MSE	= .01841

percentdis	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
β1	-.0037045	.0000955	-38.78	0.000	-.0038917	-.0035173
β2	-.0018336	.0000929	-19.73	0.000	-.0020158	-.0016514
β3	-.0008759	.0000922	-9.50	0.000	-.0010566	-.0006953
γ	.0004576	6.42e-06	71.31	0.000	.000445	.0004702
α	.0415234	.0000998	416.21	0.000	.0413278	.0417189

Regression (7) $PC_{i,r,T}=\alpha+\beta_1 WIN_{i,r-1}+\beta_2 WIN_{i,r-2}+\beta_3 WIN_{i,r-3}+\gamma\ RANK_{i,r,t}+\delta_1 WIN_{i,r-1}*RANK_{i,r,t-1}$
$+\delta_2 WIN_{i,r-2}*RANK_{i,r,t-1}+\delta_3 WIN_{i,r-3}*RANK_{i,r,t-1}$

Table 5 Result of Regression (7)

Source	SS	df	MS		
Model	3.46558079	7	.49508297	Number of obs	= 229196
Residual	77.6285345229188	.000338711		F(7,229188)	= 1461.67
				Prob > F	= 0.0000
				R-squared	= 0.0427
				Adj R-squared	= 0.0427
Total	81.0941153229195	.000353821		Root MSE	= .0184

percentdis	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
β1	-.0030395	.0001735	-17.52	0.000	-.0033795	-.0026995
β2	-.0013974	.0001781	-7.85	0.000	-.0017464	-.0010484
β3	-.0005271	.0001827	-2.88	0.004	-.0008852	-.000169
γ	.0004878	8.39e-06	58.11	0.000	.0004714	.0005043
δ1	-.0000726	.000016	-4.55	0.000	-.000104	-.0000413
δ2	-.0000423	.000015	-2.83	0.005	-.0000716	-.000013
δ3	-.0000306	.0000148	-2.08	0.038	-.0000596	-1.71e-06
α	.0411476	.0001201	342.59	0.000	.0409122	.041383

Table 6 Summary of Regressions

	(1) PC	(2) PC	(3) PC
WIN(r-1)	-0.00370***	-0.00370***	-0.00304***
	(0.0000955)	(0.0000955)	(0.000173)
WIN(r-2)	-0.00183***	-0.00183***	-0.00140***
	(0.0000930)	(0.0000929)	(0.000178)
WIN(r-3)	-0.000874***	-0.000876***	-0.000527**
	(0.0000922)	(0.0000922)	(0.000183)
WIN(r-4)	-0.000105		
	(0.0000918)		
rank	0.000457***	0.000458***	0.000488***
	(0.00000642)	(0.00000642)	(0.00000839)
dummy1			-0.0000726***
			(0.0000160)
dummy2			-0.0000423**
			(0.0000150)
dummy3			-0.0000306*
			(0.0000148)
_cons	0.0415***	0.0415***	0.0411***
	(0.000102)	(0.0000998)	(0.000120)
N	229196	229196	229196
R-sq	0.043	0.043	0.043
adj. R-sq	0.043	0.043	0.043
rmse	0.0184	0.0184	0.0184

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
dummy1=WIN(r-1)*rank
dummy2=WIN(r-2)*rank
dummy3=WIN(r-3)*rank

References

Albert E. Mannes. (2009): “Are We Wise About the Wisdom of Crowds? The Use of Group Judgments in Belief Revision,” Management Science, 55, 1267-1279.

Ali, Mukhtar M. (1977): “Probability and Utility Estimates for Racetrack Bettors,” Journal of Political Economy, 85,803-815.

Clotfelter, Charles, and P.J.Cook (1989): “Selling Hope: State Lotteries in America,” Cambridge, MA: Harvard University Press.

Clotfelter, Charles, and P.J.Cook (1993): “The ‘Gambler’s Fallacy’ in Lottery Play,” Management Science 39, 1521-5.

Croson, Rachel, and Sundali, James (2005): “The Gambler’s Fallacy and the Hot Hand: Empirical Data from Casinos,” Journal of Risk and Uncertainty 3, 195-209.

Erik Snowberg & Justin Wolfers (2010): "Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions?," Journal of Political Economy, University of Chicago Press, vol. 118(4), pages 723-746, 08.

Hurley, W. and L. McDonough (1995): “A Note on the Hayek Hypothesis and the Favorite–Longshot Bias in Parimutuel Betting,” American Economic Review, 85, 949–955.

Jonathan Guryan & Melissa S. Kearney (2008): "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," American Economic Review, American Economic Association, vol. 98(1), pages 458-73, March.

Langer Ellen (1982): “The Illusion of Control,” Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.), Judgment under uncertainty: Heuristics and Biases. Cambridge University Press.

Martin Kocher, Michal Krawczyk and Frans van Winden (2009): “‘Let Me Dream on!’ Anticipatory Emotions and Preference for Timing in Lotteries,” Tinbergen Institute Discussion Papers 09-098/1.

Matthew Rabin & Dimitri Vayanos (2010): "The Gambler's and Hot-Hand Fallacies: Theory and Applications," Review of Economic Studies, Blackwell Publishing, vol. 77(2), pages 730-778, 04.

Morrisey, Rebecca S., and Peter C. Ordoshook (1975): “Rational Choice, Light Guessing, and the Gambler’s Fallacy,” Public Choice 22, 79-89.

Paton (1997): “Why is There a Favourite-Longshot Bias in British Racetrack Betting Markets,” The Economic Journal, 107, 150-158.

Sauer, R. D. (1998): “The Economics of Wagering Markets,” Journal of Economic Literature, 36, 2021–2064.

Shin, Hyun Song (I99I): “Optimal Betting Odds against Insider Traders,” Economic Journal, vol. 101, pp. 1179-85.

Shin, Hyun Song (I992): “Prices of State Contingent Claims with Insider Traders, and the longshot bias,” Economic Journal, vol. 102, pp. 426-35.

Shin, Hyun Song (I993): “Measuring the Incidence of Insider Trading in a Market for State-Contingent Claims,” Economic Journal, vol. 103, pp. 1141-53.

Surowiecki James (2004): “The Wisdom of Crowds:Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business Economies, Societies, and Nations, ” Little, Brown, London.

Thaler, R. and W. T. Ziemba (1988): “Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries,” Journal of Economic Perspectives, 2, 161–174.

Terrell, Dek (1994):“A Test of the Gambler’s Fallacy: Evidence from Parimutuel Games,” Journal of Risk and Uncertainty 8, 309-17.
Terrell, Dek(1998): “Biases in Assessments of Probabilities: New Evidence from Greyhound Races,” Journal of Risk and Uncertainty 17, 151-166.
Tversky, Amos, and Daniel Kahneman(1974): “Judgment under Uncertainty: Heuristics and Biases,” Science 185, 1124-1131.
Woodland, Linda; Woodland, Bill (1994): “Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market,” Journal of Finance,Vol. 49 (1), pp. 269 — 280.
Terrell, Dek and Amy Farmer (1996): “Optimal Betting and Efficiency in Parimutuel Betting Markets with Information Costs,” The Economic Journal 106, 846–868.