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

Jieyao Ding

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

Jieyao Ding

revised version
May 2012

What Numbers to Choose for My Lottery Ticket?

Behavior Anomalies in the Chinese Online Lottery Market

JIEYAO DING [#]

May 2012

Abstract

The Chinese Online Lottery Game 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. Because the lottery game bases on pari-mutuel rule, that is, a bettor who wins the game shares 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, instead of doing so, people would choose the most popular numbers among others. Besides, it is proved that the gambler's fallacy does exist in the game, but the influence lasts only three days which is shorter than the findings in previous research. Second, the dataset's availability makes it possible to show how the two fallacies unfold over time within a round of the game. The result demonstrates that the proportion of later entrants who are subject to the fallacies is higher than the proportion of earlier entrants. Third, 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 was drawn in previous rounds, but are reluctant to choose the unlucky number 14 even it has not been drawn for a long time.

Keywords: Lottery Game, Gambler's Fallacy, Guidance Effect, Number Culture

JEL Classification: C93 Field Experiment,

                D81 Decision Making under Risk and Uncertainty

---

1. Introduction

500wan (www.500wan.com) is the biggest Chinese online gambling company which allows bettors participating more than 20 types of gambling games. The game that attracts my attention is called *pick 5 out of 22* (hereafter, *pick 5*) because on game's homepage, it updates a slot of information called *hot and cold number information* which displays the most and least popular numbers (hot and cold numbers) among bettors who entered current game before the update of the information. What I try to dig out is, first, how the hot and cold number information influences bettors' behavior. Second, how the bias (if any) unfolds across time. To the best of my knowledge, both points have not been documented in lottery games before.

Theoretically speaking, predicting winning numbers in lottery game is a mission impossible. Because the drawing procedure is (assumed to be) random and independent, the winning probability should be equal across numbers in each draw. Neither the frequency a number is chosen by current game's bettors nor the frequency a number was drawn in previous games provides hints about whether the number will be drawn in current game. However, over the past decades, anomalies that violated the probability's independency were found among bettors both in the lab and in the field. Based on previous findings, the following anomalies are considered as the possible consequences caused by hot and cold number information as well as previous winning numbers.

The anomaly in doubt concerning the influence of hot and cold number information is the favorite long-shot bias which describes bettors underestimating the winning probability of favorite and overestimating the winning probability of long-shot (Ali, 1977). The favorite long-shot bias is often found in financial market, horserace betting and so forth. For example, when betting on horserace, bettors often give a horse a lower subjective winning probability when it has a higher objective winning probability, or give a horse a higher subjective winning probability when it has a lower objective winning probability. Researchers give out some possibilities for the bias, e.g., it is a Kahneman-Tversky type error; people enjoy choosing the long-shot; even for some irrational reasons such as the position of the horse (Thaler & Ziemba, 1988). Besides, people argue that the bias is caused by the inside traders (Shin, 1991, 1992, 1993) or the transaction cost (Terrell & Farmer, 1996). In *pick 5*, hot and cold number information is considered to be a similar type of information to the temporary odds shown in horserace. Therefore, bettors in *pick 5* are conjectured of being subject to favorite long-shot bias.

The anomaly concerning the influence of previous winning numbers which was often found in lottery games 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. It is first proved experimentally by Tversky & Kahneman (1974). They 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 figured out not only in the lab, but also in the field. An empirical study (Terrell, 1994) examines the gambler's fallacy in a pari-mutuel game, where the payoff is shared by all bettors who hit a particular number. It is found that wagers on numbers that were drawn in previous days decrease significantly and the effect virtually lasts for a long time. Such negative recency effect not only exists in lottery games, but also shows up in many other domains, e.g., in greyhound races (Terrell, 1998), in Casinos (Croson & Sundali, 2005), and so forth. On the homepage of *pick 5*, the information of numbers that were drawn in previous games is available. Therefore, bettors in *pick 5* are conjectured of being subject to gambler's fallacy.

Although 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, the objective winning probability is not equal across candidates. For example, in horserace, the health condition of the horses would influence their performance. As a result, it is not reasonable to assume that the horses have same chance to win in the race. But because the objective winning probability can only be calculated at the end of the game, for bettors, it is impossible to know it since the game is still going on. Therefore, evaluating whether bettors overestimate or underestimate the winning probabilities is a hard job. Besides, the existence of inside information makes the situation unclear. Because the inside information could help on betting, it is necessary to distinguish the influence of inside information and the temporary odds in order to conclude whether bettors are subject to favorite long-shot bias. But normally, it is hard to find out who have inside information. Second, in previous research, only the data of each round's final bet is available, which implies researchers could only use simulations to know what happens during the procedure of the game. However, this is an imperfect alternative for such investigation because it is not possible to evaluate the influence of procedural information. For example, the temporary odds a bettor relied on in the game are unknown. Third, either in horserace or lottery games, it is difficult to assess the influence of choices made by others. For example, it is possible that a bettor who purchases lottery ticket in a lotto shop talks to other bettors, but there is no record of the conversation.

In contrast, none of the limitations is the case in *pick 5*. First, different with horserace, in *pick 5*, the probability of being drawn is equal across numbers, i.e., 1/22.[1] Nothing and nobody could change this. Therefore, if a bettor overestimates or underestimates the winning probabilities of numbers, it is a bias for sure. Meanwhile, *pick 5* excludes the possibility of inside traders and inside information. Bettors cannot increase their

---

[1] There are 26334 possible 5-number combinations. The winning probability of each 5-number combination is 0.0000379.

winning probabilities by means of internal transaction. Second, the internet is a good platform to investigate the time pattern of betting with real time-serial data that shows what happens in each fixed period of time, e.g., each half an hour, in a round of the game.[2] It is more accurate than using simulation to display how bets develop across time. Third, on the one hand, the posting of hot and cold number information allows every bettor knows what others did. On the other hand, the information allows me to observe the influence of others' behavior in greater detail.

The game will be investigated in the following steps. First, how the hot and cold number information influences betting behavior as well as whether the gambler's fallacy exists and interacts with the hot and cold information will be figured out. If any anomaly is found in the game, the second step is to display how the anomaly develops across time in a round of the game.

The paper is organized as follows. Section 2 introduces the game and data in detail. Section 3 illustrates the theoretical predictions. Section 4 gives an overview of betting behavior. Section 5 shows the anomalies found in the game. Section 6 is the conclusion.

2. The Game and Data Collection

2.1. The Game

Pick 5 is a lottery game that bettors need to choose five numbers from 1 to 22. In a five-number combination, only the five numbers matter, but not the order of the numbers, e.g., a combination of (1,2,3,4,5) is the same as a combination of (2,4,5,3,1). Each lottery ticket costs 2RMB (approx. 0.31USD). Every bettor can buy any number of ticket(s). The final draw is executed by the state lottery institution and shown live. If a bettor's five-number combination is the same as the final draw, he gets to share the jackpot. The money is distributed on a pari-mutuel basis, that is, the payoff to each winning wager is equal to the jackpot divided by the number of winning wagers.[3] The jackpot is 50% of present round's sale and will not be carried to next round if nobody wins in current game. 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. If four numbers out of five match the final draw, the payoff is 50RMB (approx. 7.6USD) per wager. If three numbers match, the payoff is 5RMB (approx. 0.76USD). The small payoffs are fixed and informed in advance.

2.2. Data Collection

A JAVA program automatically collected data from the code of hot and cold number information's webpage each half an hour and generated a file for a round of the game.

---

[2] In this paper, the time pattern represents how the bets unfold from the beginning till the end within a round of the lottery game.

[3] In the past, researchers often described the jackpot as being divided by the number of winners. This is not true in *pick 5*, because it is possible that a bettor buys more than one lottery ticket. Therefore, I use wagers instead of winners.

The data was collected from October 2009 to August 2010 (the round numbers are listed in Appendix I).

The dataset includes two main parts. One part is the general information of each draw which contains the five winning numbers, the final bet frequency of each number, the total bet in each draw and so forth. The other part is the time series data shown in the hot and cold number information which includes how many bets on each number from the beginning till the update of the information and the proportion of bets (see Appendix II).[4] The game starts at 19:40 (Chinese time) and ends at 19:40 the next day. Because the information is updated every half an hour, for each round of the game, 48 sets of observations are collected and each set has 22 data points.[5]

3. Theoretical Predictions

3.1. Timing of Entering

Usually, there are two types of bettor in gambling market, i.e., the pleasure bettor and the professional bettor (Terrell & Farmer, 1996). It is proved experimentally (Kocher etc., 2009) that, if people are betting for fun, they prefer to enter the game as early as possible so as to have longer enjoyment from participating the game; if people are betting for earning money, they would like to enter the game when the game is approaching the end because they can get more accurate information about the odds so as to share the jackpot with fewer bettors if win. If both pleasure bettors and professional bettors show up in the market, there will be more wagers in the first and last few hours of a round than the other time. I only assume that pleasure players tend to enter at early time of a game and professional players tend to enter at late time of a game. But this does not mean that all bettors who enter early are pleasure bettors, nor are all bettors who enter late professional bettors.

Hypothesis 1: At the beginning and the end of the game, there are more wagers than in between.[6]

3.2. Hot and Cold Number Information

Suppose one goal of purchasing a lottery ticket is to win as much money as possible and bettors do not have the ability to increase their winning probabilities, by intuition, it is rational to choose the least popular numbers so that they can share the jackpot with fewer winners and earn more money. In *pick 5*, it means bettors should focus on the cold numbers in the hot and cold number information and choose the five numbers

---

[4] Although the hot and cold number information only lists top 10 hot numbers and top 10 cold numbers, in the code of the webpage, the data of bet on all numbers is available. Therefore, this is a full dataset.

[5] The last hot and cold number information could not be seen by the bettors since the new round had already begun. But this data was collected from the code of the webpage for analysis.

[6] Living habit possibly influences the time of participating. Because the game starts at 19:40, people would get to sleep soon and go to work afterwards. Therefore, at the beginning and the end of the game, it is possibly more convenient for bettors to buy the tickets.

which have the least bets among other bettors.

Hypothesis 2:   The less popular (colder) a number was, the more bettors choose it. The more popular (hotter) a number was, the fewer bettors choose it.

3.3.   Gambler's Fallacy

It is proved both in the lab (Tversky & Kahneman, 1974; Morrison & Ordeshook, 1975) and in the field (Clotefelter & Cook, 1993; Terrell, 1994) that bettors are subject to gambler's fallacy in lottery games. Clotfelter & Cook (1993) found that after a 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's bet to recover to its original level. Later, Terrell (1994) challenged the result with a pari-mutuel lottery game and got similar findings although the effect of gambler's fallacy was weaker. The homepage of *pick 5* lists numbers that were picked in previous eight rounds. I assume that, in *pick 5*, bettors are influenced by previous winning numbers.

Hypothesis 3:   Bettors are subject to gambler's fallacy in *pick 5*.

3.4.   Time Pattern of Biases

3.4.1.   Gambler's Fallacy

Suppose the market has two types of bettor. Professional bettors would be influenced less seriously by previous winning numbers than pleasure bettors because professional bettors are more possible to insist that previous winning numbers are independent of the draw in present round. Therefore, if pleasure bettors tend to enter the game earlier and professional bettors tend to enter later, the effect of gambler's fallacy should become weaker over time.

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

It is found (Jørgensen et al., 2011) that if a number was on a streak (i.e., a number was drawn continuously in previous games), bettors' attitude would switch to prefer choosing it, that is, *hot-hand fallacy*. In *pick 5*, there are some cases that a number was drawn in a few consecutive days. Therefore, I make the hypothesis that if a number was picked continuously in *pick 5,* bettors' attitude towards this number would switch from gambler's fallacy to hot-hand fallacy.

Hypothesis 5:   Bettors switch from being subject to gambler's fallacy to being subject to hot-hand fallacy if a number was on a streak.

3.4.2.   Hot and Cold Number Information

At the beginning of the game, the sample size of hot and cold number information is relatively small. And the pleasure bettors who tend to enter the game earlier often have their own favorite numbers.[7] Therefore, the effect of hot and cold information

---

[7] In the paper by Jørgensen et al. (2011), it is proved that most players did not change their number

is limited. In contrast, when the game is nearly done, the professional bettors enter the game and they would rely on the information because it is an indication for the odds which directly relates to how much they could earn if win. Therefore, it is assumed that more later-entrants are influenced by the hot and cold number information than the earlier-entrants.

Hypothesis 6:   The proportion of bettors who are influenced by the hot and cold number information increases over time in a round of the game.

4. Overview

For the ease of understanding, I first explain the notations that will be used. In a *pick 5* game, there are 22 numbers that a bettor 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 time when the information gets updated is t (t=1,…,48). At each time t, one set of data is collected which includes 22 data points and each of them is a record of the betting frequency ($BET_{i,r,t}$) on number i. T (T=1,...,47) stands for the time slot between time t and $t+1$.

Theoretically, if bettors are unbiased, all numbers should receive nearly equal bets at the end of each draw. Figure 1 presents the mean of final bets on each number.[8]
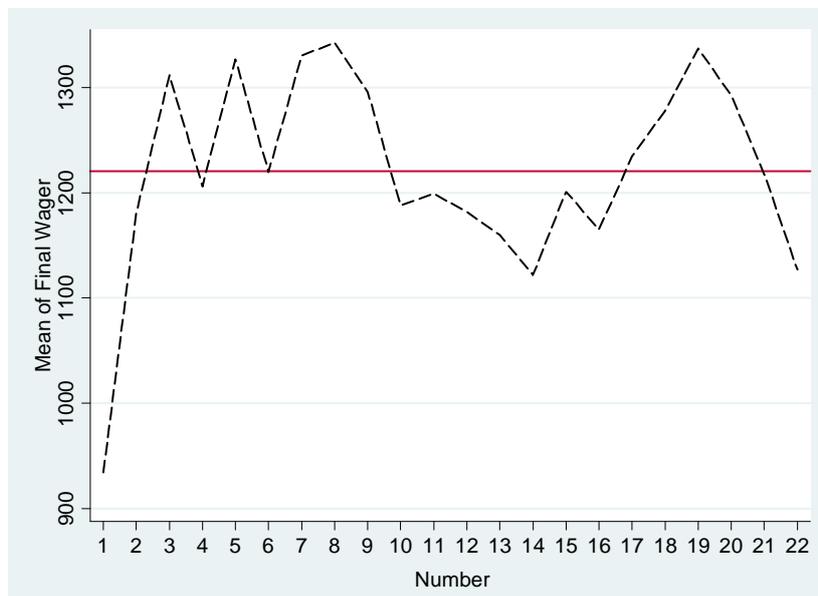

Figure 1 Mean of Final Bets on Each Number

In Figure 1, the solid line is the mean of bets if people hold the idea that the winning probability is the same across numbers and randomly choose five numbers. The dash line is the mean of real bets on each number. The figure illustrates the fact that people

---

combinations week after week.

[8] See Appendix III Table 1 for a summary.

put bets on numbers unequally, which indicates that bettors have biases on numbers. Among all, number 8 gets 1342 wagers on average which is the highest, whereas number 1 gets only 934 wagers which is the lowest. The huge difference of bets between numbers cannot be ignored.

Besides, I will show the distribution of bets across time in a round of the game. On the one hand, bettors could enjoy more pleasure if they hold the lottery ticket from the beginning of the game. On the other hand, the best strategy of a pari-mutuel game suggests bettors to join the game in a short time (e.g., in the last half hour) before the end of the game because the nearly final odds is available and by betting on the least popular numbers bettors can earn more money if win. Therefore, bettors who join for fun would like to enter the game once it starts and bettors who join for money would like to enter at the last few hours. If the game has both types of bettor, it possible receives more bets on both sides of the time line than in between. This is proved in Figure 2 which shows the increase of bets in each time slot. From the figure, it could be found that, at the beginning and the end of the game, there are more bettors enter than in between. The increase of bets expands strikingly from T=1 and reaches the peak at T=7. Then, it falls dramatically until around T=16. After that, it recovers gradually and grows relatively fast from T=37 until the end of game. Hypothesis 1 is supported.
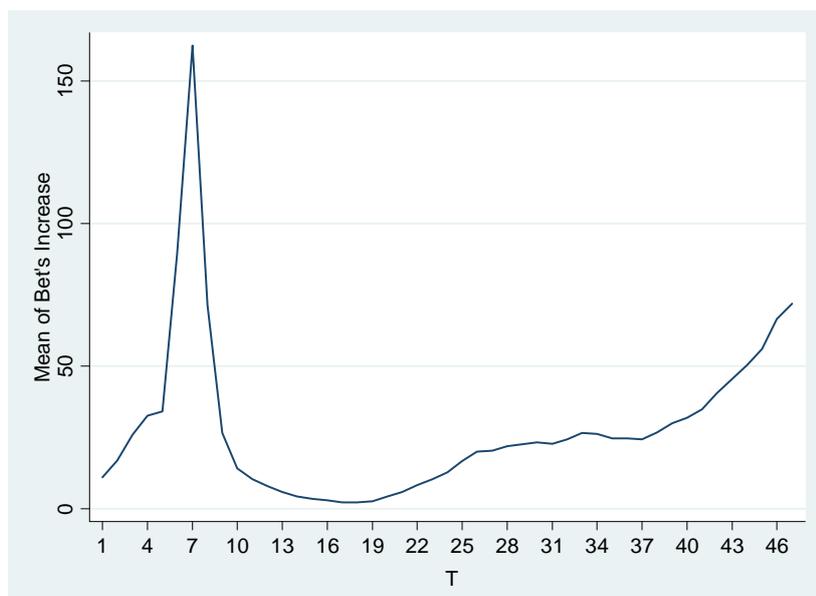

Figure 2 Distribution of Bet across Time

5. Behavior Anomalies

In this section, the observed behavior will be discussed in three main parts. In 5.1, the influence of information (i.e., previous winning numbers; hot and cold number information) will be analyzed and the biases found in the market are shown in detail. In 5.2, how the biases unfold across time in a round of the game namely the time pattern of biases is displayed. And in 5.3, Chinese number culture is introduced for explaining the findings in subsection 5.2.

In this research, the questions under investigation are the influence of hot and cold number information and the time pattern of biases. The information on the homepage of *pick 5* is enough for these research goals, therefore only the information on the homepage will be taken into consideration.

5.1. Biases towards Information

In this subsection, I am going to present how the information shown on the game's homepage influences betting behavior, and report all the biases found in the market. I start with testing whether bettors are subject to gambler's fallacy with each round's final data as has been done in previous research.

In order to make the total amount of daily sale comparable, I use relative bet $P_{i,r,t}$ instead of absolute bet for analyzing.

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

$BET_{i,r,t}$-The amount of bets number i receives at time t in round r.
$P_{i,r,t}$-Percentage of bets on number i at time t in round r.

Now I examine whether bettors are subject to gambler's fallacy as observed in previous studies. Although bettors could have full information about what numbers were drawn in all the previous games, the focus of this research is to test the influence of information on the game's homepage which only lists what numbers were drawn in previous eight rounds. Therefore, I test whether the winning numbers in previous eight rounds influence betting behavior with regression (2).[9]

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

$FP_{i,r}$- Percentage of final bets on number i in round r, i.e., $P_{i,r,48}$;
$WIN_{i,r-1}$- Dummy for number i in round *r-1*; $WIN_{i,r-1}=0$ if number i was not drawn in round *r-1*; otherwise $WIN_{i,r-1}=1$. It is the same for all the other WIN dummies.

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

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

$$SE \quad (0.0000791) \quad (0.0001291) \quad\quad (0.0001291) \quad\quad (0.0001291) \quad\quad (0.0001291) \quad\quad adj\ R^2=0.2436$$

In the regression, all the coefficients are statistically significant (see Appendix III, Table 2). The negative coefficients of WIN dummies indicate that bettors underestimate the probability of being drawn in present game if a number was drawn in any of the past four rounds. This gives support to hypothesis 3, that is, bettors are subject to gambler's fallacy. Moreover, the coefficients show that the effect of

---

[9] Here I only analyze whether bettors are subject to gambler's fallacy with final wager data as has been done in previous research. The time serial data will be considered later.

gambler's fallacy fades as time passes. There are 22 numbers on which players can bet, and all the numbers have equal chance to be drawn. If bettors are unbiased, all the coefficients of WIN dummies should therefore be 0, and the constant in the regression should be equal to 1/22=0.0455, i.e., the amount of bets on each number should be 4.55% of the overall. The constant in regression (3) is larger than 4.55% is the result of the fact that all the coefficients of WIN dummies are negative. In present round, the percentage of bet on numbers that were drawn in round *r-1* is 0.0426854, which is 93.91% of the normative percentage, i.e., 6.09 % less. And for numbers which were only drawn in round *r-2*, the percentage increases to 98.70%, i.e., only 1.30% less. By comparing the two normative percentages, it could be found that the influence of numbers that were drawn in round *r-2* is weaker than the influence of numbers that were drawn in round *r-1*. This is similar to the results found by Clotfelter & Cook (1993) and Terrell (1994). If a number was picked both in round *r*-1 and round *r*-2, the bet is 87.97% of the normative percentage, which means a number's repeated win enhances bettor's belief that this number will not be drawn again in current game. However, it will be proved later that if a number was continuously drawn for four times, some bettors start believing that, in present game, this number will be drawn again with a higher probability.

In *pick 5*, the influence of winning numbers is only effective for 4 days, which is shorter than the findings in previous research, e.g., 60 days. I propose this is due to the fact that the choice set in *pick 5* is much smaller (choose 5 numbers out of 22, ratio=1/4.4) compare to other games (e.g., choose 1 number out of 1000, ratio=1/1000), which leads to a faster re-picking of the numbers.

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

Now I take the influence of hot and cold number information into account and do some tests with the full data.

Since the time series data is applied, the relative change of bets within each time slot is calculated as follows:

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

$(BET_{i,r,t+1}- BET_{i,r,t})$ - The amount of bets on number *i* between time *t* and *t+1* in round *r*;
$\Sigma(BET_{j,r,t+1}- BET_{j,r,t})$ - Sum of bets on number *j* *(i≠j)* between time *t* and *t+1* in round *r*;
$PC_{i,r,T}$ – Relative change of bets on number *i* of T.

It is proved that winning numbers in the previous four rounds significantly influence betting behavior. Besides, hot and cold number information is speculated. So I regress $PC_{i,r,T}$ on winning numbers in previous four rounds and the rank of number *i* at time *t* in round *r* which is shown in the hot and cold number information. The result is presented in regression (5).

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

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

$RANK_{i,r,t}$-Rank of number i's bets from the beginning of the game till time t in round r. The hottest number's rank equals to 22 and the coldest number's rank equals to 1.[10]

The results shown in regression (3) indicates that $WIN_{i,r\text{-}4}$ influences bettors' behavior with the other variables, but in regression (5), this effect is not significant (see Appendix III, Table 3 and Table 6).[11] Therefore, I regress $PC_{i,r,T}$ on all variables except $WIN_{i,r\text{-}4}$ and get the following result.

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

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

The result in regression (6) confirms that bettors are subject to both gambler's fallacy and guidance effect[12] (see Appendix III, Table 4 and Table 6 for details). The coefficients in the regression show that the information of previous winning numbers induces bettors to underestimate the probabilities that these numbers will be drawn again in current game whereas the information of hot and cold numbers induces bettors to overestimate the winning probabilities of hot numbers. The finding concerning the effect of hot and cold number information rejects what was suggested in hypothesis 2. It is not true that the hotter a number was, the fewer bettors will choose it. In contrast, the more often a number was chosen by current game's earlier-entrants, the more bettors will choose it. Even though a number was picked in round r-1 ($WIN_{i,r\text{-}1}$=1), if it has the highest rank in current round ($RANK_{i,r,t}$=22), it is still more preferable ($PC_{i,r,T}$=0.0478861) compare to the unbiased condition ($PC_{i,r,T}$=0.0454545). The hot and cold number information attracts bettors' attention switching from winning numbers in previous rounds to possible winning numbers in present round.

Because bettors are subject to gambler's fallacy, it is possible that previous winning numbers have influence on numbers' current ranks. Therefore I add interaction items between WIN dummies and RANK into the regression as shown in regression (7).

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

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

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

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

The regression result supports the idea that hot and cold number information interacts

---

[10] If there is a tie, the average of ranks is applied.

[11] Only final data is used in regression (3), whereas full data is used in regression (5). Therefore, the influence of previous winning numbers is slightly different.

[12] I define the effect of hot and cold number information as *guidance effect* because bettors consider it as guidance and follow the hot numbers.

with previous winning numbers (see Appendix III, Table 5 and Table 6 for details). Take previous one round's winning number is currently the hottest number (and therefore has rank=22) as example. The predicted net effect equals to $PC_{i,r,T}=0.0478425$ which is larger than under unbiased condition. It reveals that the percentage of bet on current hottest number is quite high even if it was drawn in last round. This is also the case for hottest numbers that were drawn in even earlier rounds. Therefore, hypothesis 2 is rejected by more evidence. The more popular a number was among earlier-entrants, the more bettors choose it afterwards.

Result 2: The more popular a number was among bettors who entered the game earlier, the more bettors who enter later choose this number. The less popular a number was among bettors who entered the game earlier, the fewer bettors who enter later choose this number.

I run regressions with a time lag which tests whether several slots of hot and cold number information jointly influence the selection of numbers. In other words, it is a test on whether the historical information of hot and cold numbers aggregately affects bettors' decisions. The regression result indicates that the guidance effect does not have any lag consequences. I give out the following reasons for future research: first, the ranks of numbers do not change dramatically within each round of the game. Therefore, it does not make sense to waste time on observing the information for a long time. Second, because there is no history of the hot and cold number information on the webpage, a bettor has to put a lot of effort on recording the information. So, it would be easier to rely on one piece of information than to rely on several pieces of information. Third, it is about different types of bettor. If a bettor joins for fun, he would like to enter the game as early as possible. His focus is to enjoy the pleasure. Therefore, he is not likely to wait for the update of the information but likely to purchases the lottery ticket as early as possible. If a bettor joins for money, he only needs to look up the information when the game is nearly over so as to choose the numbers that could give him more earnings. It is not necessary to check the information for more than once.

Two anomalies found in previous research describe similar situations to the guidance effect. Now I explain why I prefer guidance effect for defining the effect of hot and cold number information to the other two. One similar anomaly is the *wisdom of crowds* which describes such a phenomenon. A decision-maker has 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. [13] People prefer to make the same decisions as what others did rather than deciding on the basis of private information. However, in *pick 5*, bettors have no private information because the winning

---

[13] 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 looked up at the sky as well.

probability is objectively equal for all numbers. What others did does not provide any clue about something that later-entrants do not know. Therefore, I do not prefer using wisdom of crowds to describe the influence of hot and cold number information. The other similar anomaly under discussion is *hot-hand fallacy* which refers to a belief in positive autocorrelation of a non-auto correlated random sequence. For example, bettors believe that the winning probability is higher if they buy lottery ticket from a shop that sold a lot of winning tickets before which is called *lucky store effect* (Guryan and Kearney, 2008). In this case, the reference is how the shop performed in the past. The more often a lotto shop sold winning ticket, the more bettors will buy lottery ticket in this shop. But the hot and cold number information in *pick 5* parallels with current event. Because the information publishes the situations of current game, before the winning numbers are drawn, there is no indicator shows the quality of information's prediction. Therefore, it is not sufficient to define the influence of hot and cold number information as hot-hand fallacy. People would argue that bettors might think the information performed well in previous games and as a result, believe in the present information. If this is true, the effect could be defined as hot-hand fallacy. Although this argument is reasonable, it is not supported by empirical evidence. The hot and cold number information is not accurate in predicting winning numbers. Here is the formula for calculating the accuracy of hot and cold number information:

$$a = (R_{win} - R_{min}) / (R_{max} - R_{min}) \qquad\qquad (8)$$

a -Predicting accuracy of hot and cold number information;
R_{win}-Mean rank of bets on present round's winning numbers;
R_{min}-The minimum mean rank of five numbers (the mean rank of five coldest/least popular numbers), $R_{min} = 3$;
R_{max}-The maximum mean rank of five numbers (the mean rank of five hottest/most popular numbers), $R_{max} = 20$.

If $a=1$, it means the hot and cold number information is 100% accurate. The smaller the a is, the less accurate the information is. The result is $a = 0.48449$ which indicates that the information does not give out an accurate prediction on current game's winning numbers. If hot-hand fallacy is the explanation, bettors should not follow the information.[14]

5.2. Time Pattern of Biases

Now I switch attention to the time pattern of biases which gives a picture of how the biases unfold across time in a round of the game and supplies greater details on how bettors behave.

In order to see the effect of winning numbers on betting behavior, I first calculate the

---

[14] I do not exclude the possibility that bettors have illusion about the accuracy of hot and cold number information which leads 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.

mean of $PC_{i,r,T}$ distinguished by the value of WIN dummies, e.g., $WIN_{i,r-1}=0$ and $WIN_{i,r-1}=1$ (gambler's fallacy). Because in the regressions, it is found that winning numbers in round *r-1* had the strongest influence, I first provide the time pattern of gambler's fallacy caused by numbers drawn in the past one round as shown in Figure 3.

In Figure 3, the solid line shows the mean (which is equal to 1/22=0.0455) of the bet's relative change if numbers are not biased; the dash line (lower curve) shows the mean of $PC_{i,r,T}$ if a number was drawn in last round, which indeed is the time pattern of gambler's fallacy; the dotted line (upper curve) shows the mean of $PC_{i,r,T}$ if a number was not drawn in last round. From T=1 to T=4, the bet's relative change for numbers that were drawn in last round decreases sharply which indicates that the proportion of earlier-entrants who are subject to gambler's fallacy increases dramatically. From T=5, this trend becomes smooth. In contrast to hypothesis 4, the influence of previous winning numbers becomes stronger over time. In other words, the proportion of bettors who are subject to gambler's fallacy is higher among later-entrants than earlier-entrants. The dotted line above the average displays that more bettors prefer to choose those numbers that were not drawn in last round (WIN=0) and $PC_{i,r,T}$ is pretty stable across time.

Result 3: The effect of gambler's fallacy caused by previous one round's winning numbers is increasingly strong across time in a round of the game.
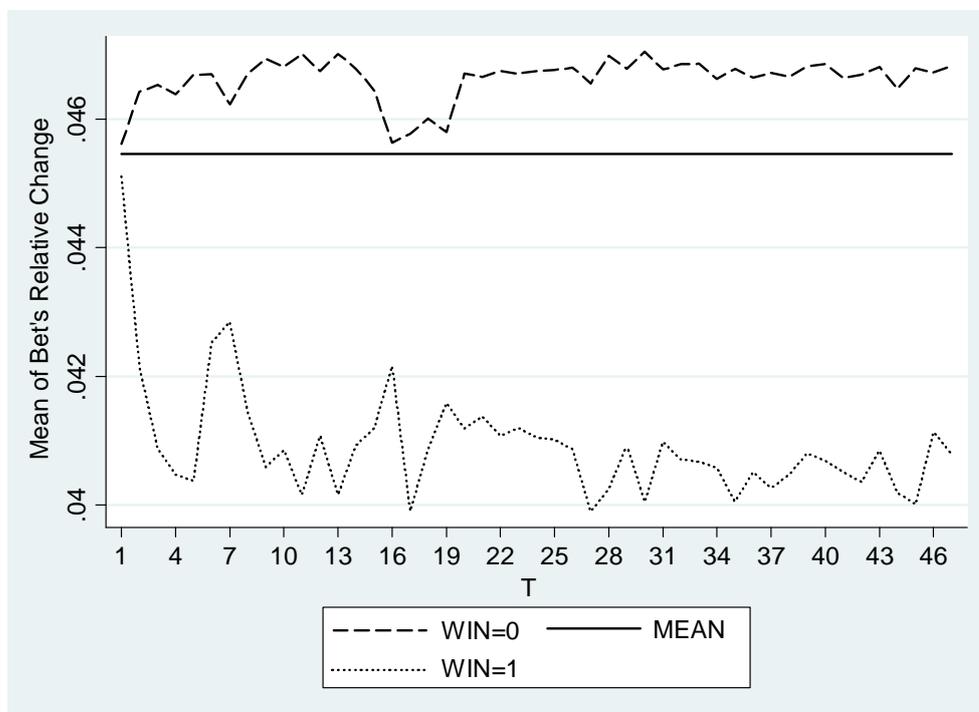

Figure 3 Time Pattern of Gambler's Fallacy (1)

As shown in the regressions, winning numbers of previous several rounds jointly influence betting behavior. Notice that, a number could possibly be drawn in more than one previous game. Therefore, in Figure 4, I show the time pattern of gambler's fallacy by the frequency a number was drawn in previous games separately.

In Figure 4, each line stands for the time pattern of $PC_{i,r,T}$ by how many times a number was drawn in previous four rounds. For example, the black dotted line marked WIN=2 displays the mean of bet's relative change of numbers that were drawn (continuously or not continuously) twice in previous four rounds. The figure displays the fact that the more often a number was drawn in the past, the fewer bettors choose this number in current game. The time pattern is pretty stable for numbers that were drawn fewer than 4 times. However, the line which indicates bettors' attitude towards the numbers that were drawn in all previous four rounds fluctuates across time. Some time, it even goes beyond the line for numbers that were never drawn in the past four rounds. Consistent with previous research (Jørgensen et al., 2011), the effect of hot and cold number information has the tendency to switch to hot-hand fallacy if a number was on a streak.
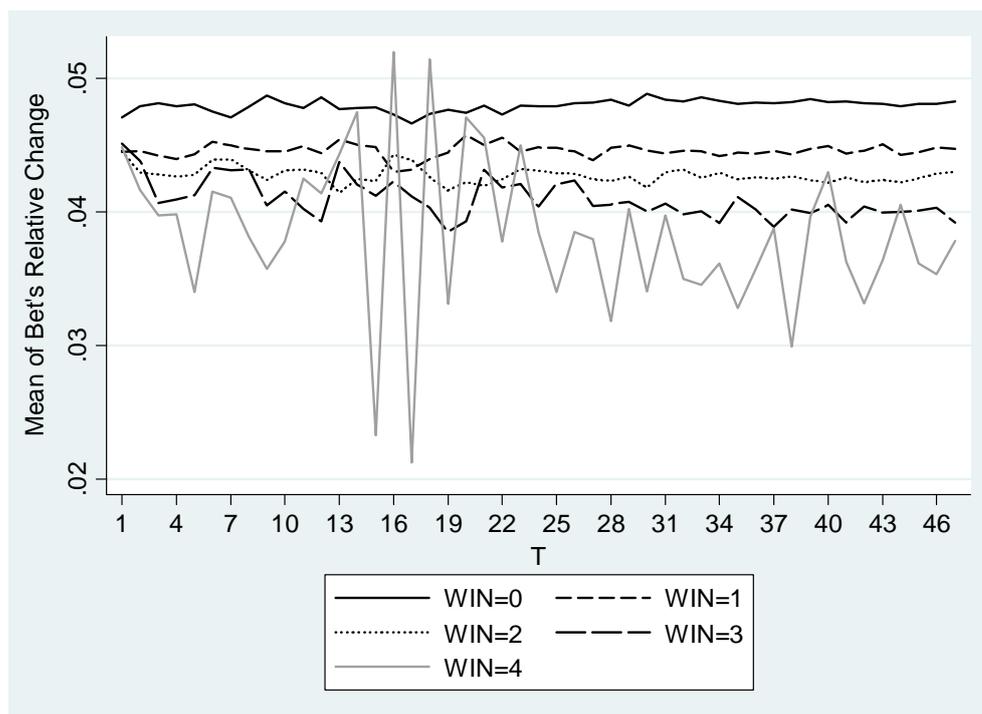

Figure 4 Time Pattern of Gambler's Fallacy (2)

Result 4:   The more often (≤3 times) a number was drawn in previous four rounds, the fewer bettors pick it in current game, and the influence is pretty stable over time. If a number was on a streak, i.e., a number was drawn in all previous four rounds, the effect has the tendency to switch from gambler's fallacy to hot-hand fallacy.

In order to see how bettors integrate the information of winning numbers in previous four rounds as a reference, I compare bet's relative change ($PC_{i,r,T}$) of numbers that were not drawn in any of previous four rounds with the bet's relative change ($PC_{i,r,T}$) of numbers that were drawn at least once in previous four rounds. Figure 5 displays the time pattern of $PC_{i,r,T}$ for these two types of numbers. The discrepancy of $PC_{i,r,T}$ between winning numbers and non-winning numbers starts from the beginning of the game and it does not change too much across time. This discrepancy is statistically significant (paired t-test, $p=0.00$, two-tailed). This indicates that bettors' attitude is pretty consistent concerning whether a number was picked in the past four rounds no matter when they enter the game.
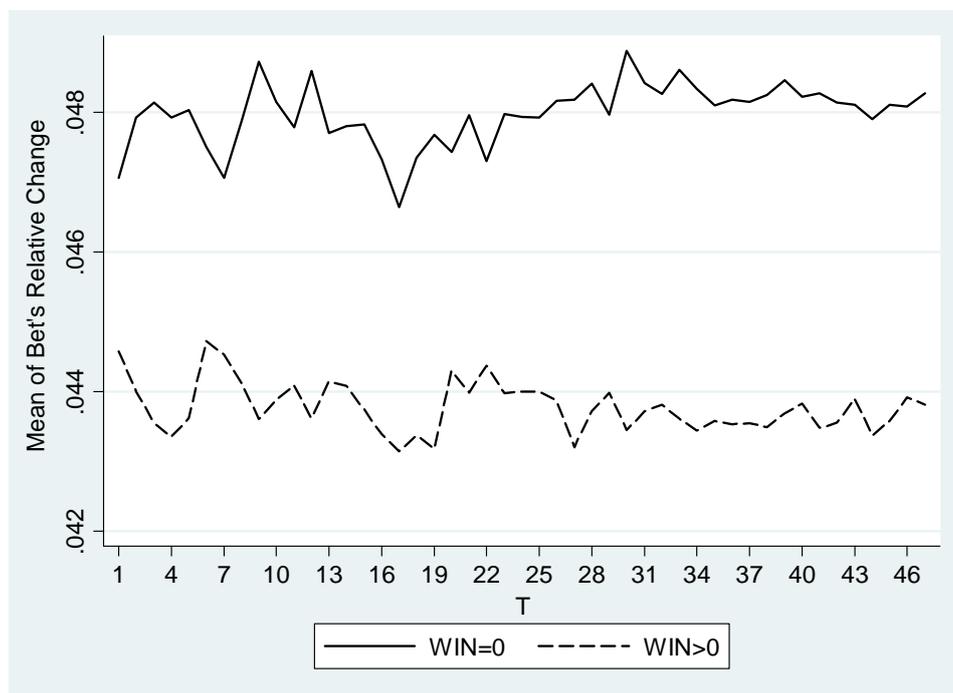
Figure 5 Time Pattern of Gambler's Fallacy (3)

Now I switch to the time pattern of guidance effect. The homepage lists ten most popular numbers (hot numbers) and ten least popular numbers (cold numbers) at the same time. For each lottery ticket, bettors only need to choose five numbers. Therefore, I provide the time pattern of guidance effect caused by top ten hot numbers and top ten cold numbers as well as top five hot numbers and top five cold numbers separately (see Figure 6).
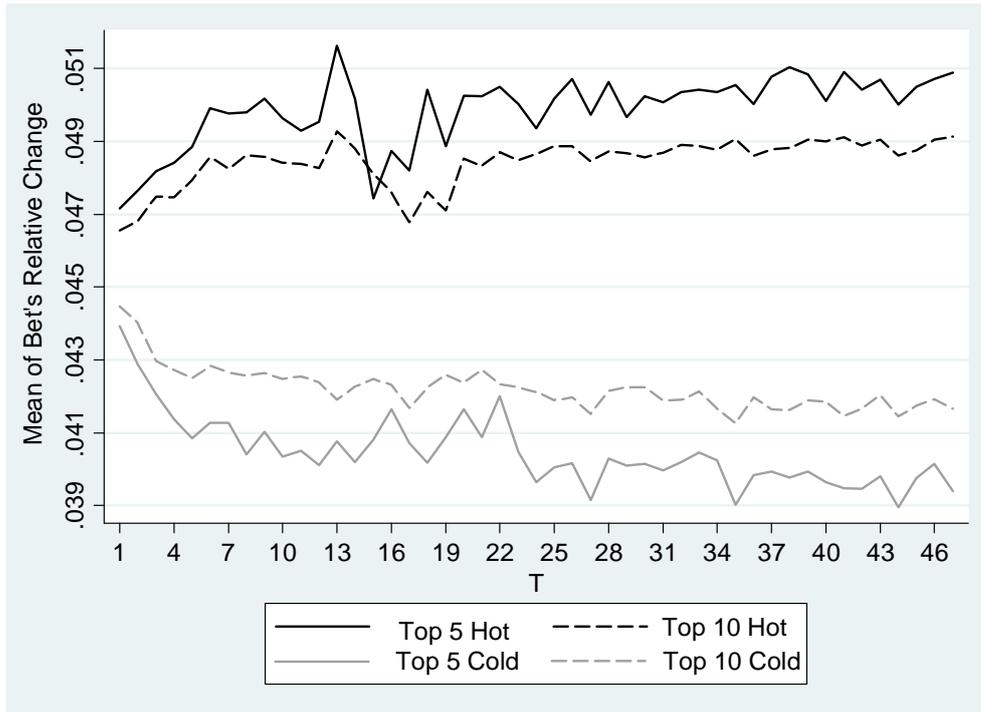
Figure 6 Time Pattern of Guidance Effect

The lines in Figure 6 confirm that hot and cold number information does affect bettors' behavior. And the guidance effect become stronger across time, that is the hotter (colder) a number was, the hotter (colder) the number becomes. 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 when the information is composed of more bettors' choices. For hot numbers, the top five numbers' $PC_{i,r,T}$ has a faster increase than the top ten numbers'. The difference of $PC_{i,r,T}$ between top five and top ten hot numbers is statistically significant (paired t-test, $p=0.00$, two-tailed). Similarly, for cold numbers, the top five numbers' $PC_{i,r,T}$ has a faster decrease than the top ten numbers'. And the difference of $PC_{i,r,T}$ between top five and top ten cold numbers is significant as well (paired t-test, $p=0.00$, two-tailed). The time pattern of guidance effect reveals that the proportion of later-entrants who are subject to guidance effect is higher than the proportion of earlier-entrants. Bettors care about what the others have already chosen. They believe that, for a candidate number, the more bettors chose it, the higher the probability that this number will be drawn in current game. This is consistent with what Mannes (2009) found in lab, that is, people hold the opinion that a larger group's decision would be more accurate than a smaller group's decision. Hypothesis 6 is supported.

Result 5:   The guidance effect is increasingly strong over time in a round of the game.

5.3.   Chinese Number Culture

In Figure 6, at the beginning of the game, there is already a discrepancy between the bet's relative changes of hot and cold numbers. Because in the first one hour, the

sample size of hot and cold number information is quite small (e.g., normally less than 100 bets for each number), the difference should not be like this. So, possibly, there are some other factors that influence bettors' choices especially at the beginning of the game. This is what I want to test next, i.e., the influence of Chinese number culture. For thousands of years, Chinese have been heavily addicted to number culture, especially businessmen who would like to pay a price for getting a good number or avoiding a bad number. In Chinese number culture, for example, 8 is the first candidate which people conjecture that could bring big fortune, whereas 14 is a number that people try to avoid. In this section, I will provide evidence on the influence of Chinese number culture on the number selection in *pick 5*.

First, the mean rank for each number is generated in the following way.[15] At time *t+1* in round *r,* the increase of bets between time *t* and *t+1* is calculated for each number *i*. Then all the increases are sorted. The number with the highest increase is assigned with rank=22 and so do others. If there is a tie, the average rank is applied. Finally, each number's mean rank over games is calculated as shown in Figure 7.
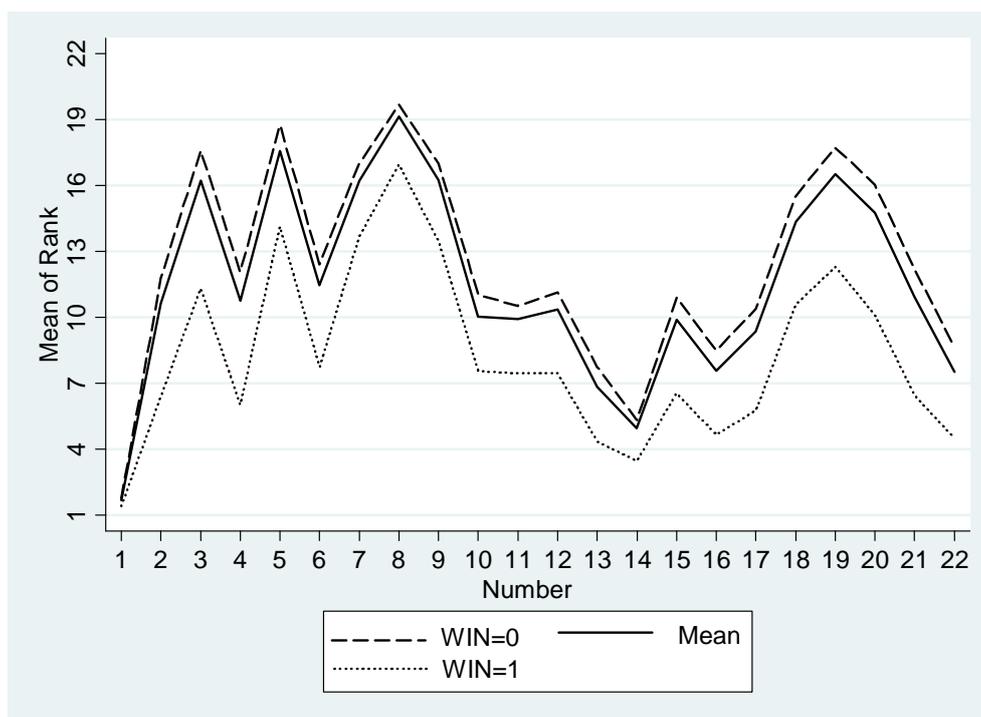
Figure 7 Mean of Numbers' Rank

In Figure 7, the solid line is the mean of overall ranks; the dash line is the mean of ranks if a number was not drawn in previous one round; and the dotted line is the mean of ranks if a number was drawn in previous one round. The lines show that the ranks are not equal across numbers. First, bettors are edge-number-averse, that is, players are

---

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

reluctant to choose edge number 1 and edge number 22. In particular, number 1 always has the lowest rank under all situations. Second, Chinese number culture plays role on number selection. Among all, number 8 is the most popular number because it has the highest rank under all three conditions and it receives significantly more bets than number 5, which is the second-most popular number (paired t-test, $p=0.00$, two-tailed). Except edge numbers, number 14 has the lowest rank which indicates bettors are not fond of it. In addition, the dotted line is an indicator for the effect of gambler's fallacy. Compared to other numbers, number 8 has the highest rank as well. It means even number 8 was drawn in last game, bettors still believe that number 8 will be drawn in current game with a higher probability.
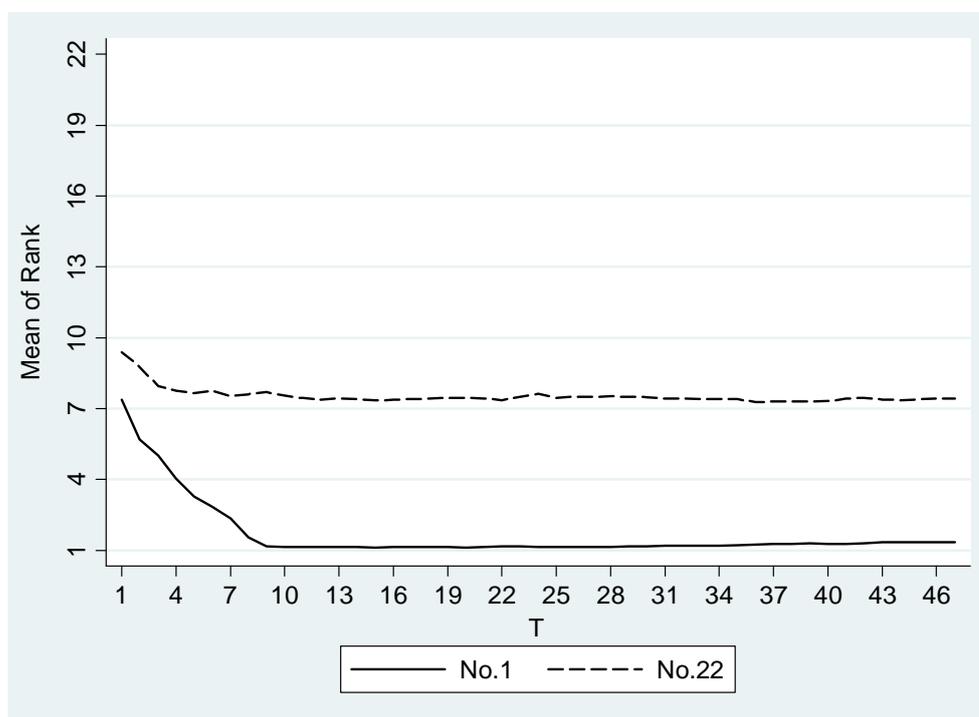
Figure 8 Time Pattern of Edge Numbers' Rank

Besides the comparison of ranks, next, I will show the time pattern of ranks of number 1, 8, 14 and 22. This gives a detailed view on how bettors consider these numbers. First, the ranks of edge numbers are explored as shown in Figure 8.

Figure 8 displays the mean ranks of edge number 1 and edge number 22 across time. From the figure, it could be seen that from the beginning of the game, the ranks of these two numbers are already lower than the average, i.e., 11.5, which means the proportion of bettors who choose either edge number is quite low. Number 1's rank drops sharply from the beginning, falls to the bottom at T=10 and remains there till the end of the game. For edge number 22, this effect is a little bit weaker, and the time pattern is quite stable since T=4 until the game is over. There is no special cultural issue that may account for the edge number aversion. This is a question that needs further investigation.

Now I turn to the time pattern of number 8's rank and number 14's rank as shown in Figure 9.
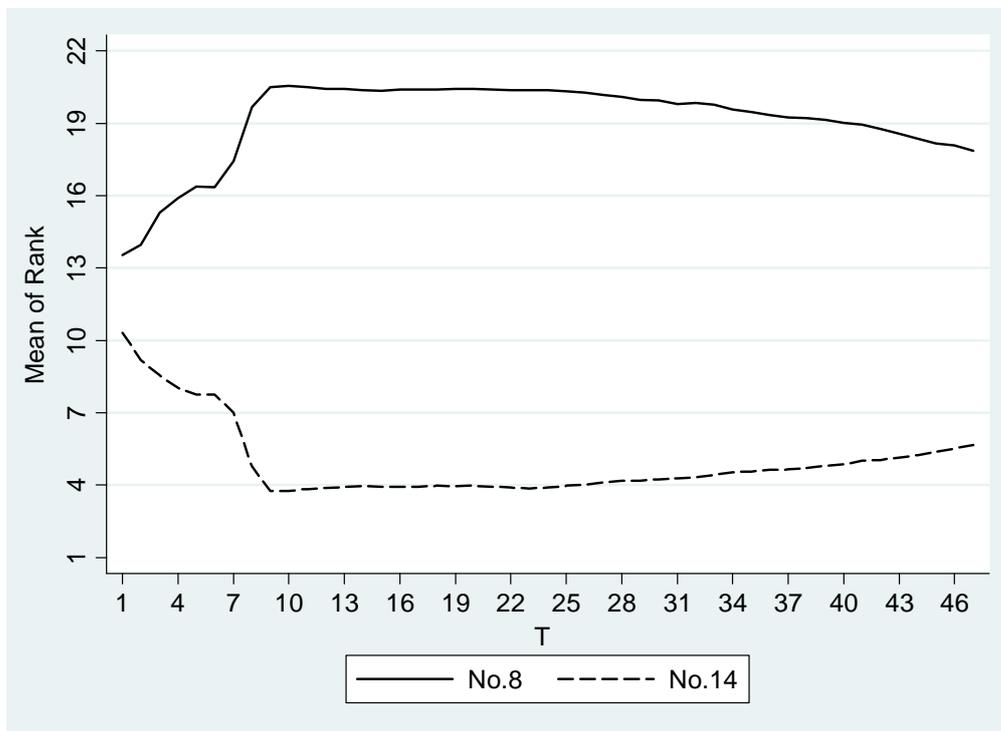
At the beginning of the game, there are already quite a lot of bettors express their attitude towards number 8 and number 14. The lucky number 8, which has a much higher rank than the average, holds this position from the beginning of the game. This advantage goes up till T= 9 and does not fall much afterwards until the end of the game. In contrast, number 14's initial rank is already lower than the average and drops dramatically till T=9. The rank does not change much from then on. First, the biases caused by Chinese number culture explains why at the beginning of the game, there is already a gap between hot and cold numbers' $PC_{i,r,T}$ (as shown in Figure 6) even though the sample size of the information is pretty small. Second, the time pattern in Figure 9 indicates that the influence of number culture is more salient on later-entrants than on earlier-entrants.

It is proved above that, for bettors in *pick 5*, Chinese number culture is one of the concerns on number selection. Finally, take number 8's role on gambler's fallacy and guidance effect as an example, I show the result from a robustness check. It found that, after taking number 8 out, the effect of gambler's fallacy becomes stronger. From this view, number 8 is really a lucky number because it alleviates the proportion of irrational decisions caused by previous winning numbers. But from the view of guidance effect, number 8 is not a lucky number because it breaks the truth that the winning probability is the same across numbers.

6. Conclusion

This paper empirically investigates the anomalies in a Chinese online lottery game. 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 homepage of the lottery game *pick 5*. The main contributions of the paper are as follows.

The first contribution is the finding of guidance effect. In the game, people do not follow the best strategy of a pari-mutuel game, i.e., choose the least popular numbers among others so as to share the jackpot with fewer winners. In contrast, bettors follow the direction given by the hot and cold number information, that is, choose the most popular numbers among others. This phenomenon is termed as guidance effect. To the best of my knowledge, the influence of such information has not been documented before.

The second contribution is the time pattern of biases. On the one hand, in a round of the game, if only winning numbers from previous one round is taken into account, the effect of gambler's fallacy grows stronger over time. But, if numbers drawn in previous four rounds are taken into consideration, the effect of gambler's fallacy becomes relatively stable over time. On the other hand, the guidance effect is increasingly stronger over time. In other words, the more popular a number was, the more bettors will choose it. The less popular a number was, the fewer bettors will choose it.

Third, the paper explains the behavioral anomalies from the view of Chinese number culture and finds that bettors rely on Chinese number culture when picking numbers. For example, number 8 which is a lucky number in Chinese culture always has the highest rank under any comparable conditions, whereas number 14 which is an unlucky number is chosen less frequently than the other numbers. Besides, from the perspective of time pattern, number 8's rank is above the average through the whole game, while number 14's rank drops dramatically from the early time of the game. Moreover, Chinese bettors are averse to edge numbers, which means they do not choose edge numbers 1 and 22 quite often. But this is not caused by number culture. The underlying reason needs further investigation.

The finding of anomalies in the lottery game provides some suggestions for both lottery ticket sellers and policy makers. For lottery sellers, from the perspective of increasing sale volume, supply information concerning others' decisions could be considered as a method. Because the winning probability in a lottery game is extremely low, bettors would feel quite hopeless and helpless if they only know the game's rule but nothing else. The posting of seemingly related information such as hot and cold number information gives bettors a way out. Bettors who rely on the information would get the feeling that they have super-power for predicting winning numbers and decide to purchase the lottery tickets. But from the view of rational behavior, posting such information is definitely not a wise choice. In the analysis, all the evidence points to the fact that bettors are subject to fallacies after seeing the

information and make some irrational decisions. For policy makers, it is necessary to figure out how to train bettors behave rationally. Besides, the finding of irrational behavior reminds policy makers taking caution when evaluate the effect of disclosing information to the public. For instance, it would be helpful to test whether the information of others' behavior has a similar effect in the stock market. Because the stock market has no ending as each day's final draw in the lottery game, the influence of information concerning others' choices (if any) could be more powerful and could lead to a market's fluctuation.

Appendix I Dataset[16]

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 ordinal number of 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		

---

[16] For technical reason, some rounds' data is missing.

Appendix II   Screenshot for Hot and Cold Number Information
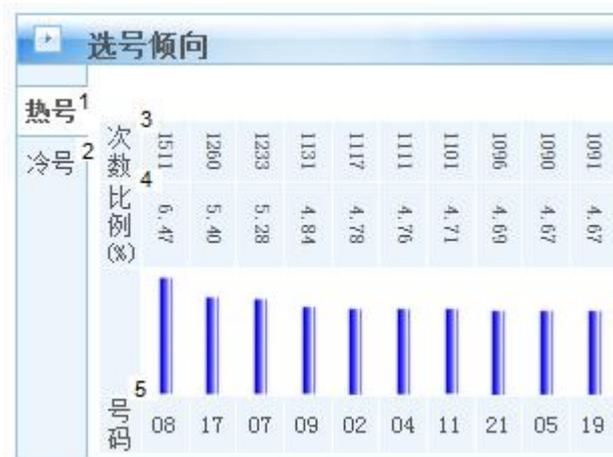
1. Hot Number;

2. Cold Number;

3. Bet Frequency;

4. Ratio (%);

5. Number.

Appendix III Statistical Tests

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

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 |  77.6398405  229190   .000338758         R-squared     =   0.0426
-------------+------------------------------            Adj R-squared =   0.0426
       Total |  81.0941153  229195   .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) $PC_{i,r,T} = \alpha + \beta_1 WIN_{i,r-1} + \beta_2 WIN_{i,r-2} + \beta_3 WIN_{i,r-3} + \gamma RANK_{i,r,t}$

Table 4 Result of Regression (6)

```
      Source |       SS           df       MS            Number of obs =   229196
-------------+------------------------------            F(  4,229191) =  2548.89
       Model |  3.45383249         4   .863458121       Prob > F      =   0.0000
    Residual |  77.6402828  229191   .000338758         R-squared     =   0.0426
-------------+------------------------------            Adj R-squared =   0.0426
       Total |  81.0941153  229195   .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) $PC_{i,r,T} = \alpha + \beta_1 WIN_{i,r-1} + \beta_2 WIN_{i,r-2} + \beta_3 WIN_{i,r-3} + \gamma RANK_{i,r,t} + \delta_1$
$WIN_{i,r-1} * RANK_{i,r,t} + \delta_2 WIN_{i,r-2} * RANK_{i,r,t} + \delta_3 WIN_{i,r-3} * RANK_{i,r,t}$

Table 5 Result of Regression (7)

```
      Source |       SS           df       MS            Number of obs =   229196
-------------+------------------------------            F(  7,229188) =  1461.67
       Model |  3.46558079         7   .49508297        Prob > F      =   0.0000
    Residual |  77.6285345  229188   .000338711         R-squared     =   0.0427
-------------+------------------------------            Adj R-squared =   0.0427
       Total |  81.0941153  229195   .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)             (2)             (3)
                         PC              PC              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 erros 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 & Philip J. Cook (1989), "Selling Hope: State Lotteries in America", Cambridge, MA: Harvard University Press.

Clotfelter, Charles & Philip J. Cook (1993), "The 'Gambler's Fallacy' in Lottery Play", *Management Science* 39, 1521-5.

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

Guryan, Jonathan and 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.

Hurley, William and Lawrence McDonough (1995), "A Note on the Hayek Hypothesis and the Favorite–Longshot Bias in Parimutuel Betting", *American Economic Review*, 85, 949–955.

Jørgensen, Claus Bjørn & Sigrid Suetens & Jean-Robert Tyran (2011), "Predicting Lotto Numbers ", *CEPR Discussion Papers* 8314, C.E.P.R. Discussion Papers.

Kocher, Martin & Michal Krawczyk & Frans van Winden (2009), "'Let Me Dream on!' Anticipatory Emotions and Preference for Timing in Lotteries", *Tinbergen Institute Discussion Papers* 09-098/1.

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

Morrison, Rebecca S., & Peter C. Ordeshook (1975), "Rational Choice, Light Guessing, and the Gambler's Fallacy", *Public Choice* 22, 79-89.

Rabin, Matthew and 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.

Sauer, Raymond D. (1998), "The Economics of Wagering Markets", *Journal of Economic Literature*, 36, 2021–2064.

Shin, Hyun Song (I991), "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.

Snowberg, Erik and 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.

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, Richard H. & William T. Ziemba (1988), "Parimutuel Betting Markets: Racetracks and Lotteries", *Journal of Economic Perspectives*, American Economic Association, Vol2 (2), pages 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.

Terrell, Dek & Amy Farmer (1996), "Optimal Betting and Efficiency in Parimutuel Betting Markets with Information Costs", *The Economic Journal* 106, 846–868.

Tversky, Amos, & Daniel Kahneman(1974), "Judgment under Uncertainty: Heuristics and Biases", *Science*, Vol185, 1124-1131.

Vaughan Williams, Leighton & David Paton (1997), "Why is There a Favourite-Longshot Bias in British Racetrack Betting Markets", *The Economic Journal*, 107, 150-158.

Woodland, Linda M. & Bill Woodland (1994), "Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market", *Journal of Finance*, Vol. 49 (1), pp. 269 — 280.