

Are Lottery Players Affected by Winning History? Evidence from China's Individual Lottery Betting Panel Data

Jia Yuan ^{*†}

University of Macau

September 2011

Abstract

I explore a unique individual level lottery betting panel data and show that lottery gambling is significantly affected by lottery winning history although this winning history is shown to be merely an exogenous random shock. This panel data records lottery players' collective lottery betting behaviors on a Chinese online lottery purchase website. Especially, this website lists each player's lottery investment performance: the ratio between the lottery return and the lottery investment in the past three months. This ratio is shown to be an independent random shock across players, and therefore, provides a clean identification for the regression as it is orthogonal to all the unobserved heterogeneities. Based on the data with around 400,000 observations, I find that this random shock significantly affects lottery players' purchasing behaviors. Specifically, collective lottery gamblers are significantly more likely to join a lottery package proposed by someone with higher winning rate; lottery players are spending more money on the proposers with higher return rates. These findings contribute to the economics literature by identifying a new gambling irrationality, other than the "Lucky Store Effect" by Guryan and Kearney (2008, AER), and the "Hot Hand Fallacy" by Clotfelter and Cook (1993, Management Science).

Keywords: Gambling, Irrationality, Lottery Betting, Gambling Fallacy

*Correspondence Author: Jia Yuan, L212, Faculty of Business Administration, University of Macau, Taipa, Macau. Email: jia yuan@umac.mo

†I would like to thank the seminar participants at the Fudan University, Shanghai University of Finance and Economics, Zhejiang Univeristy, and Peking University for the comments. I also thank Yongliang Jia for his excellent assistance. All errors are my own.

1 Introduction

Are people's lottery purchase behaviors affected if someone just won a more-than-average money? Or are lottery gamblers affected by other's winning history if this winning history is a random shock? Theory predicts that it should not happen if lottery players are rational. In this paper I explore a unique individual level collective lottery gambling panel data and show that lottery gambling is significantly affected by lottery winning history although this winning history is shown to be merely an exogenous random shock.

In this paper, I explore a unique individual level lottery betting panel data. The panel data is obtained from Taobao Lottery directly and traces all the collective lottery betting behaviors at individual level.¹ The collective lottery business works as follows: one player (Any registered Taobao user are allowed to.) proposes a lottery package with all the information listed, such as the total amount of lotteries planned to purchase, the lottery number combinations, the commission fee of the proposer and other information. After the package is put online for sale, other followers may decide whether to join the package and if so, how many shares to purchase. If the package wins money, the winning money will be split evenly by shares after the proposer takes the commission out first. The details of this online collective lottery purchase business will be discussed in section 2.

One important feature of this online lottery game is that on its front webpage where all lottery packages are listed for sale, Taobao Lottery provides each proposer's lottery winning

¹Taobao Online Lottery is a subsidiary of Taobao Lmt. Taobao is also called China's Amazon, the biggest online shopping service provider in China, or even in Asia. In 2010, Taobao had over 370 million registered customers and generated over 400 billion RMB in sales (Over \$ 60 billion) with an over 100 percent annual growth rate. RMB is the currency in China. The exchange rate between RMB and US dollar is 6.5 RMB to 1 dollar in April, 2011. Besides the normal C2C (Customer to Customer) and B2C (Business to Customer) business, Taobao provides a platform, called "Taobao Lottery", for online lottery gambling in China, which now has become the largest online lottery gaming service provider in China.

record as public information: the ratio between the lottery return and the lottery investment in the past three month. This ratio is shown to be merely an independent random shock across the players. Other lottery players utilizes this information to make their decisions. Based on the data with around 400,000 observations, I find that this random shock significantly affects lottery players' purchasing behaviors. Specifically, collective lottery gamblers are significantly more likely to join a lottery package proposed by someone with higher winning rate; lottery players are spending more money on the proposers with higher return rates.

Due to the data reason or the difficulty of regression identification, it is challenging to convincingly identify gambling fallacy. In the literature, there are few well documented. Clotfelter and Cook (1993) and Terrell (1994) document that the amount of money bet on a particular number in a pick-3 or pick-4 game falls sharply after the number is drawn, and that the betting money only gradually returns to its former level after several months. This phenomenon is called "hot hand fallacy": people expect a positive correlation between the winning lottery numbers. Guryan and Kearney (2008) document another gambling fallacy: "Lucky Store Effect". They find that consumers erroneously increase their estimate of the probability a ticket bought from the winning store will itself be a winner. Based on aggregate data, Guryan and Kearney (2008) carefully exclude the impact of advertising effect and convincingly show that there is a positive impact of hitting jackpot on the sale of lottery store.

This research builds on the previous literature in several ways. First, this paper identifies a new gambling fallacy, other than the "Hot Hand Fallacy" and the "Lucky Store Effect". It provides evidence and draws more general conclusion on how lottery betting behaviors are

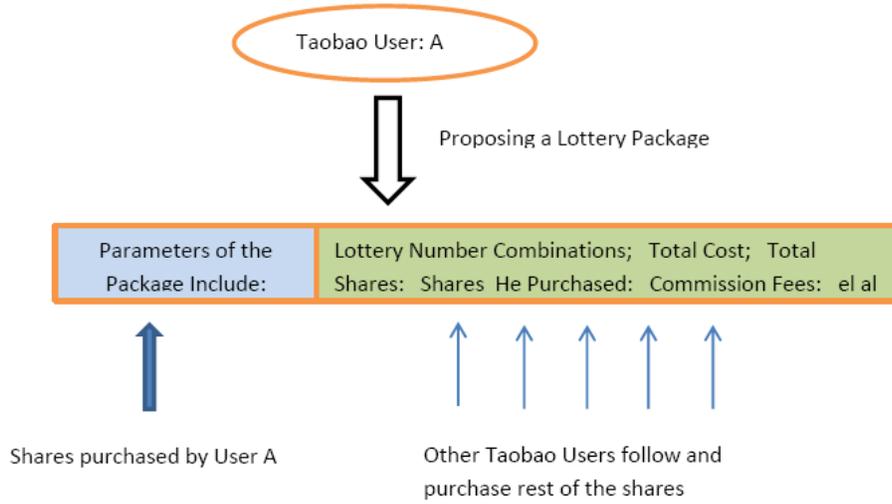
erroneously affected by random lottery winning history. Actually, "Lucky Store Effect" only represents a special case of the findings in this paper. It analyzes the case when the return rate is extremely high, while the current research shows the evidence of the impact of all lottery purchase return rates on lottery betting behavior.

Second, probably more important, the identification of the current regression is much cleaner. As the randomness of the lottery return rate is orthogonal to all the unobserved shocks, it provides a standard OLS regression condition. Therefore, the regression here provides an unbiased estimation on how lottery return rate affects on lottery betting behavior.

Third, to the best of my knowledge, this is the first empirical paper exploring individual lottery betting data. The advantage of studying individual level lottery behavior is that it provides richer information than the aggregate level data dose. The novelty of this panel data is the richness of the information, including how collective lottery gamblers share risks, how individual lottery gamblers pick lottery number combinations, how individual lottery gamblers spend in each draw, and even how individuals set commission fees when they provide number-picking services. Section 2 introduces how this Collective Online Lottery Gambling business model works in China and describes the dataset in details.

The rest of the paper is organized as follows: section 2 describes the collective lottery business in China. Section 3 sets up the regression model. Section 4 shows the dataset. Section 5 presents the regression result and discuss the results. Section 6 concludes.

Figure 1: How Collective Lottery Purchase System Works



2 Industry Background of Taobao Collective Lottery and Relevant Literature

2.1 Taobao Online Collective Lottery

Taobao, also called China’s Amazon, is the biggest online shopping service provider in China, or even in Asia. In 2010, Taobao had over 370 million registered customers and generated over 400 billion RMB in sales (Over \$ 60 billion) with an over 100 percent annual growth rate.² Besides the normal C2C (Customer to Customer) and B2C (Business to Customer) business, Taobao provides a platform, called ”Taobao Lottery”, for online lottery gambling in China, which now has become the largest online lottery gaming service provider in China.

Any registered Taobao customer can buy tickets for any of the lotteries listed on Taobao

²RMB is the currency in China. The exchange rate between RMB and US dollar is 6.5 RMB to 1 dollar in April, 2011.

Lottery. However, Taobao Lottery is more than just an online lottery store. Taobao Lottery also provides a platform for group or collective lottery purchases. Figure 1 illustrates how Taobao's collective lottery purchase system works.

Anyone with a Taobao account can propose a "lottery package". A lottery package is a collection of lottery numbers chosen by a Taobao user being sold for a price set by that same user. The user announces how many shares are in a package and corresponding price per share ($= \text{Total Cost} / \text{Total Shares}$). For example, the user can propose a package, which only consists of two lottery tickets, and the total cost is 2 RMB, 1 RMB for each ticket. Meanwhile, if this same user divides this package into 100 shares, he will attract more people to join. Therefore, there will be 100 shares for this lottery package and each share will be worth 0.02 RMB. Thus, at most, 100 lottery gamblers can participate and purchase this lottery package.

Besides the lottery number picks, total cost, total shares and share price, the user also has to reveal how many shares he bought before the package was open for sale. In other words, he must reveal his own investment to his lottery package. The last parameter the creator of a lottery package must reveal is his commission fee: the percentage of the total winning prize before the prize is divided among the investors according to their shares in this package. For example, suppose that a lottery package wins 100 RMB and the package proposer's commission fee is 3 percent. Before others share the prize, the proposer will get 3 RMB first, leaving 97 RMB for sharing.

If the open shares are not sold out, the lottery package is regarded as a failure and everyone gets his investment back. If the package is successfully sold out, the lotteries will be purchased automatically. If the package wins prize, Taobao Lottery will automatically

calculate the return for each investor based on his shares in this package and the money will be credited to each investor's account. Most important, it is free and very easy for any registered Taobao user to initiate a lottery package. Any registered Taobao user can either initiate a lottery package by himself, or simply participate in another's lottery package.³

3 Regression Model

The purpose of the regression is to reveal the relationship between lottery players' preference and the factors which may influence people's behaviors. I adopt a Logit model to examine this influence.⁴ To better understand the model, it is helpful to examine how lottery players behave in this business.

When a lottery player plans to join a collective lottery package, he goes to the "Lottery Package Hall" on the website. On this hall, all the lottery packages are listed for the player to pick. The player can filter the packages by setting the requirement on lottery package proposer's winning ratio or the value of commission fee. After comparing various packages, the lottery player may choose to join one package. In the end, some lottery packages may be successfully sold out, which means that there are enough the mass lottery players who are interested in purchasing the shares of the package. Some lottery packages may not be successfully sold out because there are not enough mass lottery players interested in this package. We examine the relationship between lottery players' behavior and lottery winning history by exploiting this difference of whether the lottery package is successfully "Sold out".

³Actually, the authors have conducted experiments on this lottery platform. The authors participated as both a lottery package proposer and as a lottery package follower. It turns out the transaction cost difference between these two roles is almost zero.

⁴Actually, adoption of other models such as Probit does not affect the results.

What this research especially focuses on is whether the lottery winning history in the past influences people's lottery choice.

In the first part of this section, I propose the benchmark regression model. All the modified regressions later examined in this paper are based on this benchmark model. In the second part of this section, I briefly discuss the randomness of the lottery return rate.

3.1 The Benchmark Model

If a lottery package is not successfully sold out there will be a zero assigned to this observation. Otherwise, the value will be 1. Therefore, we first employ a Logit model to investigate which factors are influencing the success rate of a lottery package. The estimation function is shown as follows:

$$Pr(y = 1) = F(\beta_0 + \beta_1 * RATE + \gamma * X)$$

Above is the standard Logit model. y is the indicator variable showing whether the lottery package is successfully sold out. $RATE$ represents the lottery winning return ratio. β_1 is the coefficient the research is interested in. X is the vector of all other variables of interests, which may include p : the share price of the lottery package; C : the commission rate set by the package; M : the size of the lottery package; $SELF_BUY$: the portion of the package purchased by the lottery proposer himself.

As discussed latter, if lottery gamblers are rational in their lottery gambling decision, the coefficient of β_1 should not be significant because $RATE$ is merely a random shock. The estimation for β_1 will unbiased, which is an important result of the current regress. A simple

analysis should tell that the impact of *SELF_BUY* should be zero, as no matter what the commitment the proposer makes, it does not change the nature of chance and add any edge for himself. The impact of p may be ambiguous as it depends on the risk preference of the lottery players. If he loves risks, the coefficient should be positive. Finally, lottery players should be negative on C as it only takes money away from them and actually does not improve any chance for them to win the lottery.

3.2 The Randomness of Winning Ratio

The randomness of return rate in the past three month is important for the identification as it is orthogonal to the noises and thus eliminates the impact of unobserved shocks. The proof is straight forward and the basic idea is very simple: there is no way for lottery players to gain any higher return in the lottery game and there is no extra information which can tell more about the performance of each lottery player than the nature of chance does. Here, I only discuss the main ideas.

There are two claims. First, excluding the small probability event of hitting jackpot and only focusing on the fixed return winning event, across the lottery players, the expected value for all their return rate should be equal. Second, excluding the small probability event of hitting jackpot and only focusing on the fixed return winning event, the variable of return rate of any individual lottery proposer is independent with any other variables. Combining these two claims together, the immediate conclusion is that return rates is a random shock when we exclude the event of hitting jackpot.

The reason to exclude the impact of grand prize in the lottery is that the jackpot prize is

correlated to prize pool, which is further correlated to lottery players' participation. Then, the return rate may be correlated to other variables. For example, experienced lottery players may purchase more lotteries if the prize pool is bigger and jackpot prize is bigger. Then, without controlling for the jackpot prize, experienced players may have higher expected return rate. But in reality, the probability of hitting jackpot is trivial. Therefore, the above claim is true with probability almost equaling to one.

Now, suppose the lottery game is only dealing with fixed prize with fixed probability. For a person, who may have various rules to guide him to play the game, it is impossible for him to beat any other player in terms of lottery win return rate, *ex ante*. And there is also no variable which can be used to forecast his performance, or his lottery winning return rate, other than the nature of chance. In other words, the random variable of his return rate is independent with any other variable and only governed by chance.

The above reasoning suggests that there is no extra information about lottery players's return rate. Therefore, lottery players should not make their decisions based on this information. And this is exactly what this research is going to explore.

4 Data

Data source is a collection of webpages containing all the collective lottery betting activities from the Taobao Lottery website. The lottery game of interest in this research is the most popular national lottery game named Double Color Ball.⁵ I implement an algorithm to

⁵Double Color Ball lottery is one of the two most popular lotteries in China. The other one is the China Super Lotto. In both of these two lotteries, players pick and try to match several numbers from around 40 numbers in two divisions, similar to the lottery game in other countries such as the Super Lotto in the U.S. and Lotto 6/49 in Canada.

extract the lottery betting information directly from the webpage html code and compile it into usable data format. The algorithm enables me to trace all individual lottery players' collective betting behavior related to the Double Color Ball in Taobao Lottery website for two months. I collect all the variables of interests, including each proposer's information, the number combinations of each lottery package, the total cost, the total shares, the share prices, the amount of shares bought by each proposer, the number of participants, whether or not the lottery package is sold out, the winning prize, and so on. One important feature about Taobao Lottery is that what buyers observe on the website when they purchase a lottery package is the same information that we observe as outsiders.⁶

In the end, I obtain a panel of data with 388,132 collective lottery packages or observations. This data contains 52,489 registered Taobao users and traces their histories of collective lottery proposals for two months, from 2011-04-12 to 2011-06-05 representing 24 rounds of lottery draws. Table 1 presents the simple statistics of the Data.

I trace 52,489 lottery package proposers for 24 draws and obtain 388,132 lottery packages. Each package clearly specifies the information including the total cost of the package, shares of the package divided into, the shares purchased by the proposer himself, the number of the followers in this package, the commission fee for the proposer, selling progress of the package, the amount of money won by the package, the time when the package was proposed, and most importantly, the lottery investment return rate of the lottery package proposer in the past three months.

The first row of table 1 shows an important variable of the data: the average return rate

⁶Actually, we registered an account and found that the webpage layout is the same no matter whether we log in or not. There is no extra insider information as a buyer.

of lottery investment. The mean value of the return is 70%, which means that for each 1 RMB lottery purchase the lottery player is losing 30 cents on average. The std is 11,800% which suggests huge variation of the return rates. The median return rate is 15.6% and the 90th percentile is return rate is 43%. This suggests that the distribution of the return rate is skewed to the left. In other words, most of the lottery package proposers are mediocre except few ones. The highest return rate is 4,098,400%, which suggests this player may have hit jackpot in the past three months.

5 Results and Discussion

Table 2 presents the regression results for the benchmark model. The first column records the Logit regression result when I only regress the lottery package success indicator on return rate and the round of draw dummies. As discussed above, the return rate is completely a random shock across individuals. Therefore, the estimated coefficient is an unbiased estimation of the impact of return rates on lottery betting behaviors. The estimated result is significant positive, 0.012. In other words, if the return rate increases by one percent, a representative lottery player will be more likely to purchase his package by 1.2 percent.

To check the robustness of the regression result, I add more variables into the regression. Column 2 shows the results when the commission rate is added into the regression. Column 3 adds the square of the commission rate. The results in Column 2 and 3 show similar regression results about the impact of return rate on lottery betting behaviors. There is no significant difference. In column 4 and column 5, I further add variables of the package share price and its square into the regression. In the last column 6, I regress all the possible

variables including the self-committed investment made by lottery package proposer himself: the ratio of the share purchases by himself and the total shares. The first result presented by Table 2 is that the impact of return rate on lottery betting behavior does not change much. The result is robust.

Before we go through coefficients of other variables, it is helpful to first discuss the correlation between the variables other than return rate with the lottery betting behavior. In theory, the commission rate will drive lottery package followers away because it affects their expected return negatively; if lottery players are not risk neutral, share price of the lottery package may also influence players' behavior, because higher price implies less amount of shareholders and higher amount of winning money return if the package hits jackpot, given the size of the lottery package. However, the variables such as the the portion of shares purchased by the proposer should not affect lottery players' behavior. Lottery package proposer may use it as a signal to show his confidence about his own lottery package. But this signal is fake in the sense that either does it not change the the expected return of each share or the winning probability of the lottery as it is only influence by chance.

In Table 2, column 6 shows the regression results when all the observed variables are included in the regression. First of all, all the coefficients are significant. The pair of commission coefficients, both the first order term and the second order term, suggests that the impact of commission rate set by the proposer is concave. Because the rate is small than .1, the first order term dominates. Therefore, the overall correlation between commission rate and the propensity to join the lottery package is negative. This is consistent with the above reasoning, although the estimation may potential be biased due to the lack of controlling unobserved heterogeneities.

Table 2 also suggests that lottery gamblers are sensitive to share price. The response is a concave function as revealed by the estimated coefficients for the share price and its second order term. As most of the share prices are smaller than 100 RMB, the first order term also dominates the second term here. Therefore, overall correlation between the share price and lottery players' propensity to join is negative. This implies that lottery players prefer lower risks and share lottery purchase with others, although the expected returns for all lottery share prices are equal.

Finally, table 2 shows that the success rate of a lottery package is significantly dependent on the portion of shares purchased by the proposer. This may suggest that the commitment, or confidence, shown by the proposer is an important factor that induces more lottery gamblers to join a package. The last result is not consistent with theory. This may potentially be another type of lottery fallacy.

To summarize, the above results show there is a significant correlation between the lottery gambling behavior and other factors such as the commission rate of the proposer, the prices of the lottery package and the self-committed investment made by the lottery package proposer. Discussing these correlations is also interesting if they are unbiased. For example, the expected return of lottery followers should not be affected by the self-committed purchase. However, the positive estimated result suggests that lottery players are more likely to join a lottery package in which the proposer makes bigger investment himself. This should also be one kind of gambling fallacy. However, the main issue behind this argument is the lack of control of the unobserved heterogeneities in the regression. This contrasts to the case above when I discuss about the impact of random return rate on the gambling behaviors.

The bottom line here is that the lottery investment return has a significant impact on

lottery betting behavior, which contradicts theory.

5.1 Robustness Check

In this section, I try a broad variety of model setups to check the robustness of the above results against different possible factors which may potentially affect the above regression.

Fringe Package Proposers I first check whether the results in table 2 are robust against the behaviors of fringe lottery package proposers. The possible influence comes from the fact some lottery package proposers may not be serious players. Data shows that some players propose multiple packages, sometimes more than 40 packages in a single draw. It is reasonable to speculate that these non-serious proposers may be more likely to result in failures for their lottery package proposals. Therefore, I revise the regression with a control for fringe proposers to examine the robustness of the previous result.

Table 3 records the estimation results when different criteria is used to eliminate fringe proposers. The first column presents the regression results when I eliminate all the proposers who submit more than 20 lottery packages in a single round without round dummy. The average number of packages submitted by each proposer in each round is less than 2. If a proposer submits more than 20 packages, it is reasonable to question his seriousness for the game. In the second column, I further impose the round dummy in the regression to check whether controlling the round affects the regression results. In the third column, I change the definition for non-serious players: I eliminate the proposers who submit more than three packages in a single round. The third case is the most restrict as it only considers

the players with one or two submission in each round. The results in Table 3 are similar to the results without controlling for fringe proposers. The estimated impacts of return rates in these three cases are 0.98 percent, 1.02 percent and 1.08 percent respectively, and all are significant. All these suggest that the regression results are robust against the possible impact from non-serious package proposers.

The Timing Impact Another possible factor which may potentially affect the estimation is when the proposer put his package online. It is possible that the earlier a lottery package is put online, the more accessibility the lottery players may have and therefore the higher probability the package has to be successfully sold out. Therefore, I further add the variable of time left into the regression and examine the robustness of the previous regression.

Table 4 presents the regression results. I calculate the time difference between the time when the package is put online for sale and the time when the collective lottery purchase is closed for the corresponding round. I use this to control for the possible impact of the timing issue on the lottery players' betting behavior. The second column in table 4 records the regression results when timing issue is taken into consideration. For comparison, I put the regression results without controlling timing issues in the first column. The results show that the previous conclusion that lottery betting behaviors are affected by random shock return rates do not change here. The coefficients are 1.01 percent and 1.02 percent respectively for the case with controlling for the timing issue impact and without. It suggests that there is no much significant difference. In other words, the preference of the lottery players to the proposers with higher return rates is reflected all the time.

Regression on Each Round

The previous regressions take all rounds of lotteries altogether and addressing the heterogeneity across different rounds by adding a round dummy in the regression. It is possible that although in certain round the impact of return rate is not significant, the impacts in other rounds are so strong that the average impact is significant positive as shown in the previous regressions. Therefore, to examine whether the above result is robustly hold in each round, in this subsection, I further break up the data into single round and examine the above results by running regression for each round.

The data consists of 24 rounds of lottery draws and therefore, I run 24 separate regressions. Table 5 presents the estimation results for these 24 rounds of regressions. First, the results show that the impacts of return rates on lottery betting behaviors in each round are all significantly positive. It suggests that higher return rate does induce lottery players to follow the lottery package and this is robust no matter which round of lottery draw is considered. Second, there seems a variety of propensity of lottery players to follow the high return rate proposers in each round. In certain round, estimated coefficient is only 0.25 percent and in certain round, the estimated coefficient is over 1 percent. Although all significantly positive, these results suggest that the impact of return rate on lottery betting behavior may not be constant over time. People may change their propensities to follow high return rate proposers. In summary, the above result on how lottery return rate affects lottery betting behavior is robust here.

5.2 Impact of Winning History on the Lottery Expenditure

In this section, I further look at how the actual lottery expenditure is affected by the lottery winning history. From the lottery followers' perspective, higher winning rate of lottery proposers should not induce them to purchase more lottery share of him if they are rational. From the lottery proposer's perspective, if he expects that the mass lottery players's lottery purchase is not affected by his lottery winning performance, the strategy to set bigger lottery packages should not work. For himself, if excluding the income effect as return is not high enough if he does not hit jackpot, his own lottery expenditure should also not be affected by his lottery winning history. Therefore, in the end, the lottery winning history should have no impact on lottery expenditure.

I employ the simple OLS regression to examine the impact. As emphasized above, the randomness of the return rate is orthogonal to other unobserved shocks, which provides an unbiased estimation.

I look at two dependent variables: The first one is the total lottery expenditure of the lottery followers other than the proposer himself. This variable is interesting because it reveals lottery players' reaction to the lottery package proposer's winning history; the second variable examined here is the ratio between the lottery package shares purchased by others and the shares purchased by the package proposer himself. This relative ratio variable is also revealing how lottery winning history affects lottery players' behavior.

Table 6 shows the estimation results of how lottery history affects the lottery expenditure. The the first two columns, I regress the lottery expenditure on return rate with and without controlling for round dummy, respectively. The results are similar. It turns out that when

lottery package proposer's return rate increase by one percent, the corresponding lottery investment of other people on him will increase by around .7 percent, significantly. This pattern is also kept in the third column and the fourth column. In these two columns, I regress the ratio of other's lottery investment and the proposer's own investment on return rate. In the third column, I adopt the absolute value of the ratio and in the fourth column, I use the log value. Both of these two cases show similar results: lottery package proposers tend to use higher leverage, in terms of the ratio, to attract other's investment, when he has higher rate of return. This is not consistent with theory as the return rate is only a random shock.

5.3 Replicating the "Lucky Store Effect": The Impact of Big Win Shock

In this subsection, I attempt to replicate the results of "Lucky Store Effect", identified by Kearney (2002). They found that in the U.S., stores that had sold winning lottery tickets experienced significant increase in lottery sales. Kearney (2002) show that in the week after a store sells a winning lottery ticket it will see a 12-38 percent increase in lottery ticket sales. They name this phenomenon the "Lucky Store Effect". In the current context, each lottery package proposer is functioning similar to the lottery store in Kearney (2002). I try to examine whether an unusual lottery win of a lottery package can generate similar consequence in the following draws. Specifically, I investigate whether a "Big Win" for certain lottery package proposer will affect the lottery betting behaviors in the following

draws. Here I have a similar regression compared to Kearney (2002) based on individual level lottery betting data.

I define two types of "Big Win". In the first case, I define "Big Win" as the lottery package winning return rate in a single round is higher than 100 percent. From table 1, the average lottery winning return rate is 70 percent and the 90 percent of the return rates are lower than 43 percent. In other words, 100 percent lottery return rate is well above the averages. Therefore, it is reasonable to take it as a "Big Win". In the second case, I further increase the bar for the criteria. I define "Big Win" as the lottery package winning return rate in a single round is higher than 500 percent. And people should be noted that the whole conclusion of the regression is not dependent on this critical value.

It is helpful to have a discussion about the "Big Win" definition and its implication before we go through the results. First, it is a relative number as opposed to the absolute value defined in Guryan and Kearney (2008). Because higher amount of lottery purchase has higher probability to win jackpot, the definition based on absolute value creates endogeneity issue. However, a relative ratio of lottery return return rate, meanwhile also an independent random shock, does not have this issue. Second, based on aggregate data, Guryan and Kearney (2008) carefully and convincingly exclude the impact of advertising effect as jackpot itself creates buzz to attract more people to come. The story in this online collective lottery business is different. Before mass lottery players decide whether and whom they will follow, they may check the profile of the proposers by just one click on the lottery package proposer's link. Then, they can easily read his recent lottery purchase performance, calculate and evaluate his return rate. As only the average three month lottery return rate is put on the Lottery Hall, these recent lottery performance will not create advertising buzz unless the potential

lottery players dig it a little bit by a simple click. Therefore, it is reasonable to believe that advertising effect is not an issue in the current regression.

Table 7 records the estimation results. In the first two columns, I run the Logit regression to examine the impact of "Big Win" on the success rate of lottery package. The third and fourth column, I examine the impact of "Big Win" on the lottery expenditure. The first column shows that if the definition of "Big Win" uses the 100% criteria. It turns out that after a lottery package proposer wins a big prize, in the next round, his package is more likely to be sold out successfully. If I change the definition of "Big Win" into 500% criteria, the result is similar. Therefore, if we take the lottery package proposer as a "Lottery Store", it suggests that more people are likely to join this "store" in the next round of lottery draw after this "store" wins big in the previous round. This is the "Lucky Store Effect".

I further look at the impact of "Big Win" on the log-expenditure in the third column and the fourth column under two different definition of "Big Win". The third column shows that after a lottery package proposer wins a big prize, in the next round, the lottery expenditure of his lottery package from other people will increase significantly under both definition of "Big Win". What is more, the effect of "Big Win" under 500% criteria definition is bigger than that under 100% criteria definition.

All above results are replicating the "Lucky Store Effect" in the literature. Actually, they show that "Lucky Store Effect" is only a special case of the findings in this paper as it only represents the case with high return rates. All these results suggest that lottery gamblers tend to switch their bets to lottery packages whose proposers just recently won big, although, theoretically, this is not rational.

6 Conclusion

Examining the misconception of the randomness is an important empirical question as it hinges the fundamental issue in people's decision making when they are facing uncertainties. The current China's online collective lottery business provides a good social experiment to examine people's lottery betting behaviors and further touches their conception of randomness in reality. As opposed to most of the past research mainly based on experimental data, I investigate this issue using an individual level lottery betting panel data obtained by extracting and tracing data directly from individual lottery bettor with the help of auto algorithm.

I mainly focus on examining the lottery betting reactions to other people's lottery betting history. As argued above, this lottery betting history is completely a random shock governed by chance. This randomness gives extra advantage for current study as it provides the standard orthogonal regression condition, which eliminates the impact of unobserved heterogeneities. Therefore, the identification is cleaner than most of other research. The estimation result is interesting. People are influenced by this random shock. And by further trying a broad variety of model setups, the result is very robust. People are erroneously purchasing the lottery package shares if the proposer has a higher lottery investment return rate.

An immediate question following the above findings is what explains this phenomenon. This question is out of the scope of current paper as I only want to show solid evidence

of lottery betting irrationality. The question is worth investigating though, as it is related to the fundamental understanding about people's decision making under uncertainty. My speculation would be the misconception or misinterpretation of probability for the lottery players. This is left for future study.

Another issue about the current research is that this longitudinal data contains much richer information than what is explored in this research. Current research only exploits the variation across lottery players to examine the gambling fallacy behavior. But it is more interesting to explore the dynamics of lottery betting behavior. This is out of the scope of current research and also left for future research.

References

- CLOTFELTER, C., AND P. COOK (1993): “The Gambler’s Fallacy in Lottery Play,” *Management Science*, 12(39), 1521–1525.
- GURYAN, J., AND M. S. KEARNEY (2008): “Gambling at Lucky Stores: Evidence from State Lottery Sales,” *American Economic Review*, 98(1), 458–73.
- KEARNEY, M. S. (2002): “State Lotteries and Consumer Behavior,” *NBER Working Paper*, (9330).
- TERRELL, D. (1994): “A Test of the Gamblers Fallacy: Evidence from Pari-Mutuel Games,” *Journal of Risk and Uncertainty*, 3(8), 309–317.

Table 1: Simple Statistics of the Data

	Mean	SD	Min	Median	90th Percentile	Max
Return Rate in the Past Three Months	.71	118.80	0	.156	.43	40,984
Cost of Each Package	412.2	5,346	8	16	140	10,000
Money Each Package Win	39	16,076	0	0	20	10,000,000
Shares of Each Lottery Package	495	6,965	1	50	240	2,079,700
Price of Each Share	1.71	45.04	0.2	.4	1	10,000
Number of Followers	11.23	57.5	1	6	19	8,322
Shares Bought by the Proposer Himself	76	442	1	25	100	75,000
The Portion of Self-purchase	.542	.253	.01	.594	.833	1.00
Commission Fee	5.6%	4.3%	0	6%	10%	10%
Progress of Lottery Package	.862	.30	.01	1.00	1.00	1.00
Packages Success Rate (Total) in Each Round	81.5%	–	–	–	–	–
# of Draws	24	–	–	–	–	–
# of Proposers	52,489	–	–	–	–	–
# of Observation	388,132	–	–	–	–	–

Note: The money is in term of RMB. The data records start on 12th April 2011 and end on 5th June 2011.

Table 2: What Influences Lottery Betting Behavior: Benchmark Model

	(1)	(2)	(3)	(4)	(5)	(6)
Return_Rate	0.0119*** (0.000230)	0.0116*** (0.000230)	0.0115*** (0.000229)	0.0113*** (0.000228)	0.0113*** (0.000228)	0.0101*** (0.000227)
Commission		0.0473*** (0.000967)	-0.0767*** (0.00445)	-0.0817*** (0.00446)	-0.0825*** (0.00446)	-0.316*** (0.00610)
C^2			0.0122*** (0.000428)	0.0127*** (0.000429)	0.0127*** (0.000429)	0.0217*** (0.000582)
Price				-0.0249*** (0.00124)	-0.0322*** (0.00151)	-0.0174*** (0.00174)
P^2					0.0000457*** (0.00000387)	0.0000234*** (0.00000574)
Self_comm						8.235*** (0.0283)
_cons	1.217*** (0.00566)	0.968*** (0.00746)	1.059*** (0.00820)	1.091*** (0.00833)	1.098*** (0.00838)	-1.495*** (0.0128)
N	372391	372391	372391	372391	372391	372391

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: What Influences Lottery Betting Behavior: Robustness Check on the Fringe Bettors

	(1) status	(2) status	(3) status
Commission	-0.316*** (0.00610)	-0.335*** (0.00617)	-0.342*** (0.00687)
C^2	0.0218*** (0.000582)	0.0233*** (0.000588)	0.0239*** (0.000655)
Price	-0.0144*** (0.00135)	-0.0132*** (0.00133)	-0.00891*** (0.00225)
p^2	0.00000683*** (0.00000137)	0.00000608*** (0.00000160)	-0.00000120 (0.0000108)
Self_comm	8.159*** (0.0285)	8.322*** (0.0292)	7.971*** (0.0333)
Return_rate	0.00983*** (0.000227)	0.0102*** (0.000231)	0.0108*** (0.000277)
_cons	-1.457*** (0.0129)	-1.893*** (0.0306)	-1.662*** (0.0339)
N	368888	368888	298854

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: What Influences Lottery Betting Behavior: Robustness Check on the Timing Issues

	(1)	(2)
	status	status
status		
Commission	-0.335*** (0.00617)	-0.333*** (0.00619)
C^2	0.0233*** (0.000588)	0.0231*** (0.000589)
Price	-0.0132*** (0.00133)	-0.0117*** (0.00132)
p^2	0.00000608*** (0.00000160)	0.00000500* (0.00000221)
Self_comm	8.322*** (0.0292)	8.419*** (0.0295)
Return_rate	0.0102*** (0.000231)	0.0101*** (0.000229)
Time_left		0.00999*** (0.000314)
_cons	-1.893*** (0.0306)	-2.172*** (0.0320)
N	368888	368888

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: What Influences the Lottery Betting Behavior: Analysis for Each Round

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	status	status	status	status	status	status	status	status	status
Commission	-0.323*** (0.0293)	-0.324*** (0.0288)	-0.277*** (0.0291)	-0.278*** (0.0312)	-0.310*** (0.0310)	-0.338*** (0.0303)	-0.213*** (0.0314)	-0.284*** (0.0322)	-0.300*** (0.0325)
C^2	0.0241*** (0.00281)	0.0248*** (0.00276)	0.0195*** (0.00279)	0.0211*** (0.00300)	0.0232*** (0.00298)	0.0249*** (0.00290)	0.0138*** (0.00302)	0.0203*** (0.00309)	0.0203*** (0.00310)
Price	0.0120 (0.0123)	0.00222 (0.0118)	0.0201 (0.0127)	0.0208 (0.0172)	-0.00122 (0.0147)	-0.00946 (0.00514)	-0.00224 (0.0102)	0.00784 (0.0122)	-0.00537 (0.0125)
p^2	-0.0000841 (0.000105)	-0.0000809 (0.000132)	-0.000124 (0.0000924)	-0.000160 (0.000173)	-0.0000797 (0.000145)	0.00000778 (0.00000848)	-0.0000282 (0.0000565)	-0.000143 (0.0000974)	-0.0000413 (0.0000931)
Self_comm	7.594*** (0.135)	7.376*** (0.131)	7.980*** (0.132)	7.436*** (0.145)	8.076*** (0.143)	8.238*** (0.141)	7.647*** (0.140)	7.678*** (0.145)	7.903*** (0.145)
Time_left	0.0147*** (0.00189)	0.00900*** (0.00182)	0.00761*** (0.00115)	0.0126*** (0.00201)	0.00819*** (0.00193)	0.00892*** (0.00121)	0.0153*** (0.00196)	0.0111*** (0.00202)	0.0132*** (0.00129)
Return_rate	0.00463*** (0.000924)	0.00380*** (0.000902)	0.00501*** (0.000910)	0.00257** (0.000810)	0.00735*** (0.00104)	0.00836*** (0.00102)	0.00823*** (0.00110)	0.00542*** (0.000985)	0.0105*** (0.00118)
_cons	-1.963*** (0.0832)	-1.795*** (0.0825)	-2.020*** (0.0828)	-1.755*** (0.0914)	-1.884*** (0.0852)	-2.106*** (0.0850)	-2.089*** (0.0879)	-1.700*** (0.0884)	-2.250*** (0.0950)
N	13,830	14,104	15,482	12,911	13,956	14,505	13,129	12,989	12,909

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Impact of Return Rates on Lottery Expenditure

	(1)	(2)	(3)	(4)
	Log_Sum_others	Log_sum_others	Ratio_money	Log_ratio_money
Commission	0.0629*** (0.00246)	0.0633*** (0.00246)	-0.324*** (0.0103)	-0.127*** (0.00223)
C^2	-0.00361*** (0.000233)	-0.00365*** (0.000233)	0.0154*** (0.000986)	0.00497*** (0.000213)
Price	0.0931*** (0.00116)	0.0928*** (0.00116)	-0.0350*** (0.00493)	0.0140*** (0.00106)
p^2	-0.000360*** (0.00000749)	-0.000359*** (0.00000749)	0.000138*** (0.0000318)	-0.0000663*** (0.00000685)
Self_comm	-2.662*** (0.0122)	-2.671*** (0.0123)		
Time_Left	0.000572*** (0.000113)	0.000396** (0.000121)	0.0123*** (0.000472)	0.00930*** (0.000109)
Return_Rate	0.000747*** (0.0000407)	0.000745*** (0.0000407)	0.000723*** (0.000172)	0.000681*** (0.0000372)
_cons	3.355*** (0.00847)	3.386*** (0.0139)	2.142*** (0.0224)	-0.247*** (0.0111)
N	244415	244415	244415	244415
adj. R^2	0.196	0.196	0.028	0.136

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Replicating the Lucky Store Effect

	(1)	(2)	(3)	(4)
	Package Success	Package Success	Log_Sum_others	Log_Sum_others
main				
Commission	-0.316*** (0.00680)	-0.316*** (0.00678)	0.0676*** (0.00269)	0.0683*** (0.00269)
C^2	0.0217*** (0.000649)	0.0219*** (0.000648)	-0.00394*** (0.000255)	-0.00397*** (0.000255)
Price	-0.0160*** (0.00154)	-0.0161*** (0.00154)	0.107*** (0.00142)	0.107*** (0.00142)
p^2	0.00000155*** (0.000000284)	0.00000157*** (0.000000283)	-0.000441*** (0.00000971)	-0.000443*** (0.00000974)
Self_comm	8.422*** (0.0321)	8.412*** (0.0321)	-2.723*** (0.0136)	-2.726*** (0.0136)
Time_left	0.0101*** (0.000341)	0.0103*** (0.000341)	-0.000439*** (0.000131)	-0.000354** (0.000132)
Big_Win_100	0.908*** (0.0318)		0.391*** (0.0109)	
Big_Win_500		1.920*** (0.139)		0.671*** (0.0418)
_cons	-1.679*** (0.0314)	-1.635*** (0.0313)	3.412*** (0.0153)	3.419*** (0.0153)
N	317,619	317,619	210,033	210,033
adj. R^2			0.196	0.192

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$