Representativeness Biases and Lucky Store Effects

Jaimie W. Lien[±] Jia Yuan[†] Jie Zheng[‡]

This version: October 13th, 2015
Initial version: May 5th, 2014

Abstract:[1]

Recent theories (Rabin, 2002; Rabin and Vayanos, 2010) propose that both the Gambler's Fallacy and the Hot Hand Fallacy are driven by Representativeness Bias, also known as the Law of Small Numbers (Tversky and Kahneman, 1971). The Lucky Store Effect (Guryan and Kearney, 2008), in which the popularity of a lottery retailer surges after selling a winning ticket, bears an intuitive similarity to the Hot Hand Fallacy in terms of the belief that 'previous winners will win again', yet has been previously thought of as irreconcilable with Representativeness Bias. This paper develops theory and empirical evidence on this issue. We extend the Law of Small Numbers model of Rabin (2002), to the context where a decision-maker chooses among different lottery ticket stores after observing a history of prior outcomes. We show the conditions under which the decision-maker tends to believe that the store which has won previously has a higher chance of winning again, even if this event has only occurred once before. We then provide new empirical evidence on the Lucky Store Effect and Gambler's Fallacy, from a large peer-to-peer online lottery marketplace for the Chinese national lottery. We find that lottery players exhibit Gambler's Fallacy beliefs when picking lottery numbers, while believing in Lucky Stores when choosing which online lottery store to purchase their tickets from, which is consistent with the result in our model that decision-makers will tend to believe in the Lucky Store Effect when the perceived uncertainty about true probabilities is greater.

Keywords: Representativeness Bias, Lucky Store Effect, Gambler's Fallacy, Hot Hand Fallacy
JEL Codes: D01, D03, D81, L86

---

[1] ± Lien: Department of Decision Sciences and Managerial Economics, The Chinese University of Hong Kong, Hong Kong; Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, jaimie.lien.tsinghuasem@gmail.com ; † Yuan: Faculty of Business Administration, University of Macau, Taipa, Macau, jiayuan@umac.mo ; ‡ Zheng: Department of Economics, School of Economics and Management, Tsinghua University, Beijing, China, zhengjie@sem.tsinghua.edu.cn; We are especially grateful to Vincent Crawford and Colin Camerer for detailed comments which helped us to improve the manuscript. For helpful comments and insights, we thank Chong-En Bai, Jeff Carpenter, Ronaldo Carpio, Jimmy Chan, Xiaogang Che, Soo Hong Chew, Ian Crawford, John Hey, Brad Humphreys, David K. Levine, Yew Kwang Ng, Paul Schweinzer, Jason Shachat, Satoru Takahashi, Dimitri Vayanos, Cheng Wang, Ruqu Wang, Zaifu Yang and seminar participants at the Tsinghua SEM brown bag lunch, Fudan University, Nanyang Technological University, CESS Nuffield College, University of Oxford, Durham University Business School, University of York, and the 2014 Tsinghua Conference on Theoretical and Behavioral Economics. All errors are our own.

1. Introduction

In predicting the outcome of random variables, when do people tend to believe in positive correlation over realizations, and when do they tend to believe in negative correlation? The Gambler's Fallacy refers to the well-documented tendency individuals have, to believe that stochastic outcomes over time are more negatively serially correlated than they are - while the Hot Hand Fallacy refers to the seemingly contradicting phenomenon of believing that stochastic outcomes over time are more positively serially correlated than they actually are. As Rabin (2002) and Rabin and Vayanos (2010) propose, both fallacies can be explained by an underlying belief in representativeness, the idea that small samples from a distribution ought to closely reflect the distribution itself. Their theory thus suggests that these two seemingly contradictory fallacies are not ad-hoc anomalies or context-specific deviations from rational beliefs, but are part of the same underlying phenomenon – the Hot Hand Fallacy arising from Gambler's Fallacy beliefs when the decision-maker has to make an estimate of the true probabilities.

In this paper, we provide new theoretical and empirical field evidence to support the relationship between the Gambler's Fallacy and Hot Hand Fallacy. On the theoretical side, we show that the Law of Small Numbers model in Rabin (2002) applied to the task of cross-sectional comparisons, can generate a belief that 'winners will win again' even when only one previous period of observation has occurred. This phenomenon, first documented empirically in Guryan and Kearney (2008) as the Lucky Store Effect in Texas State Lottery sales, was originally interpreted as irreconcilable with representativeness-based belief fallacies such as Rabin (2002) and Rabin and Vayanos (2010), due to the focus in the literature on streaks of outcomes over time.[2]

We show that a model following Rabin (2002) in a cross-sectional decision setting, will indeed generate a tendency to believe that winners will win again, even over trivially short sequences of outcomes. That is, a decision-maker does not require a long sequence of outcomes to infer positive serial correlation. Our extension of Rabin (2002) implies that the Lucky Store Effect can be understood as a special case of a generalized class of Hot-Hand-Fallacy-like beliefs, sharing in common the tendency to believe that 'previous winners will win again'. A key contribution of our study is to fill in a gap in the literature on representativeness bias, which is due to the previous literature's focus on probabilistic outcomes over *time* rather than in either a cross-section or a panel structure of outcomes. By considering inferences over the cross-sectional dimension, our paper seeks to clarify the ties between the Gambler's Fallacy, Lucky Store Effect, and Hot Hand Fallacy, in both the theoretical and empirical domains.

Contributing to the empirical evidence on the Lucky Store Effect, and its co-existence with Gambler's Fallacy beliefs, we analyze a rich individual-level dataset from an online collective lottery betting market for China's national lottery. This setting provides a natural experiment for identification of belief fallacies. Under the structure of the online collective betting marketplace, our lottery consumers observe two sequences of realized random outcomes, which are in fact independently and identically distributed over time and in the cross-section: 1. the outcome of the numerical lottery draws which occur in each round over time, and 2. the win rates of the individual lottery ticket sellers whom they can purchase ticket shares from online. Due to the fully random nature of lottery number draws, which simultaneously determine the lottery outcome and sellers'

[2] Guryan and Kearney note with regard to their empirical findings, "There is no streak to move someone from the Gambler's Fallacy to a belief in a Lucky Store….The behavior we document makes this particular representativeness-based explanation unlikely [referring to previous work, Gilovich, Vallone and Tversky, 1985; Camerer, 1989; Rabin, 2002]. Lottery players appear to infer the luck of a store after a single win; they do not require a surprising streak of wins before they move away from an expectation of negative serial correlation." Indeed, in the introduction of their paper, Guryan and Kearney call for a model of "*when* we should expect to observe either of two well-documented but seemingly contradictory misperceptions of randomness: the hot hand and gambler's fallacies – one an expectation of positive and the other of negative serial correlation."

returns independently of one another, we can observe lottery consumers' choices in response to prior exogenous outcomes in each domain.

We find that consistently with the influence of Representativeness Bias (Tversky and Kahneman, 1971), individual gamblers subscribe to the Gambler's Fallacy in response to previous numerical draws, which we observe through their choices of lottery numbers. However, this same group of individuals subscribes to the Lucky Store Effect in response to previous successes of sellers, which we observe through their choices of which ticket sellers to purchase from. This is despite the fact that holding the number of different tickets purchased constant, no particular seller or numerical ticket should have any advantage in winning the lottery. In addition to showing the two fallacies exist simultaneously in the marketplace, our study provides supporting evidence on when we should expect each type belief fallacy to occur. In our setting, the Gambler's Fallacy arises over basic probabilistic events, while the Lucky Store Effect arises over individual performances. Our theoretical model derives the analogous insight that situations where true probabilities are well-known tend to produce Gambler's Fallacy beliefs, whereas perceptions of high uncertainty over true odds tends to produce Lucky Store Effects.

The empirical pattern we observe among consumer's choice of ticket sellers indeed replicates Guryan and Kearney's finding that lottery stores that have just sold a jackpot ticket ("Lucky Stores") subsequently receive an unusual jump in lottery sales. However, our paper adds a further dimension to the analysis, since our data is at the individual ticket purchase level as well as at the store level. Thus, we can detect that the same ticket buyers who produce Lucky Store Effects simultaneously subscribe to the Gambler's Fallacy in lottery number choices. This provides the first direct empirical evidence to our knowledge, of the co-existence of the Gambler's Fallacy and Lucky Store Effect.

Previous studies have examined the simultaneous existence of the Gambler's and Hot Hand Fallacies, most commonly over a single domain of serial outcomes, as predicted by Rabin (2002) and Rabin and Vayanos (2010). Croson and Sundali (2005) examine roulette playing behavior in the casino, finding evidence for the Gambler's Fallacy in bet choices among binary categorizations on the roulette wheel (red/black, odd/even, low/high), and evidence for the Hot Hand Fallacy in terms of bet amount adjustments to prior wins and losses, which they use to proxy for the strength of beliefs. Rao (2009) finds a transition from the Gambler's Fallacy to the Hot Hand in an experiment where subjects had to predict the outcome of basketball shots in NBA games. Galbo-Jorgensen, Suetens, and Tyran (2015) examines Danish Lottery data, finding that gamblers exhibit the Gambler's Fallacy over immediate previous outcomes, and the Hot Hand Fallacy over longer term streaks in outcomes. Asparouhova, Hertzel and Lemmon (2009) find evidence using a laboratory experiment for the model of Law of Small Numbers beliefs as in Rabin (2002). The empirical findings in our study are complementary to these works in that we also find simultaneous evidence for beliefs in reversals and in streaks, although our focus differs in that the cross-sectional domain is the source of the belief in streaks in our study. By focusing on the cross-sectional domain, the belief in streaks by our lottery players can occur even over very short sequences.[3]

Finally, our study contributes to the literature which examines the existence and perception of the time-dimensional Hot Hand in various domains, by proposing the Lucky Store Effect as a cross-sectionally driven Hot Hand Fallacy. Gilovich, Vallone and Tversky (1985) finds that although basketball players and fans tend to believe in the idea of a Hot Hand in the domain of shooting success, there is no empirical evidence that a Hot Hand exists. Camerer (1989) finds that the betting market for basketball also believes in the Hot Hand. Reexamining this question, Miller and Sanjurjo

[3] Some of the papers in this literature have mentioned the intuitive connection between Guryan and Kearney's Lucky Store Effect and the Hot Hand Fallacy. However, no rigorous theoretical treatment of the connection has yet been given. Thus, our paper fills an important existing theoretical gap in the literature by specifying a model which can connect the two phenomena, and providing empirical evidence.

(2015a) find that contrary to previous analyses, a Hot Hand does exist in NBA shooting, thus questioning whether belief in the Hot Hand can be truly considered a fallacy (see also Miller and Sanjurjo (2015b)). Moving from the sports domain to the financial domain, Powdthavee and Riyanto (2015) find that people are willing to pay 'experts' for predictions of random outcomes, over which the experts clearly have no predictive ability. Their experiment provides a controlled environment for testing whether decision-makers are reasonably attracted to previously successful funds or fund managers, where potential Hot Hands among funds are of interest in the finance literature (see for example, Hendricks, Patel and Zeckhauser, 1993; Jagannathan, Malakhov and Novikov, 2010). Our findings support that 'investors' prefer to put their bets on previously successful funds or fund managers, even when there is no true skill involved (see also Yuan, Sun and Siu, 2014, which measures the willingness to pay). Finally, Huber, Kirchler and Stockl (2010) find in a laboratory experiment, that subjects who rely on 'experts' to predict random outcomes tend to believe in the Hot Hand, while those who choose to predict on their own tend to believe in the Gambler's Fallacy.

The remainder of our paper proceeds as follows: Section 2 provides a theoretical model showing that the Lucky Store Effect can be generated by a belief model with Representativeness Bias, and provides intuitive conditions for the effect to hold; Section 3 introduces the data from the online lottery betting market; Section 4 describes our empirical strategy; Section 5 presents the empirical results and robustness checks; Section 6 summarizes and discusses the relationship between the Hot Hand Fallacy and Lucky Store Effect. Detailed proofs and tables for robustness checks are provided in the Appendix.

2. Model Background

The main objective of our theoretical model is to demonstrate that the general framework in Rabin (2002), which uses Representativeness Bias as its foundation, can generate an immediate Hot Hand Fallacy result after just one period of observation (ie. a Lucky Store Effect). We choose the simpler framework of Rabin (2002) rather than the more generalized version in Rabin and Vayanos (2010) due to the fact that the true data generating process in our field data actually matches the setup in Rabin (2002) very closely.[4]

In Rabin (2002), Law of Small Numbers believers think of the true probabilities over possible outcomes, as a representative but *finite* number of outcomes (or 'balls') in a pot. For example, in the simple case of a fair coin flip, a Law of Small Numbers believer with pot size N believes there is an initial distribution of outcomes in the pot: N/2 heads, N/2 tails.[5] The negative serial correlation in outcomes implied by the Gambler's Fallacy is generated by random draws being made from the pot without replacement. Thus, the larger the pot size, the less biased the decision-maker is in the Law of Small Numbers sense; as the pot size goes to infinity, the decision-maker is a classical Bayesian.

The model of belief fallacies in Rabin (2002) is of a decision-maker's beliefs under a true independent and identically distributed (iid) process, but can be extended to cases where the true data generating process is in fact not iid (see Rabin and Vayanos, 2010). In order to model the Lucky Store effect, we maintain this basic framework of Rabin (2002) in analyzing consumers' beliefs among *a set* of different lottery ticket stores the consumer is to choose from. Our Lucky Store Effect resembles the 'over-inference effect' described in Rabin (2002), where inference about the true distribution of outcomes after seeing a sequence of realizations from a single seller, can lead to a Hot Hand Fallacy.

[4] Rabin and Vayanos (2010) allow for the possibility that the true probabilities (not just the beliefs of the decision-maker) change over time, a flexible feature useful for financial market settings, but which is not necessary for our simple lottery market setting.

[5] Note that this implies that the true probabilities must be able to be expressed accurately as a finite fraction of outcomes where the number of balls is an integer. For example, in the case of modeling a belief about a fair coin flip, the pot size N must be divisible by 2.

2.1 Model Setup

In deciding which store to purchase a lottery ticket from, the consumer is uncertain about each store's true winning rate, and uses previously occurring observations to infer which of the stores are most promising. Of course, a consumer with standard beliefs realizes that the ex-ante likelihood of winning in each round is independent of the store at which the ticket is purchased – an assumption about the consumer's beliefs which we relax here in accordance with Rabin (2002).

There are n stores, indexed by $i = 1,...,n$, where each store has M_i possible outcomes in its pot. For simplicity, we assume that all stores have the same $M_i = M$ number of outcomes in the pot. The "type" of store i is denoted as $\theta_i \in \{r_1,...,r_K\}$ which represents that particular store's winning probability. The priors $\pi_i(\theta_i)$ on each store's winning rate are such that $\sum_{k=1}^{K} \pi_i(r_k) = 1$.

Let m index the number of outcomes the consumer observes from each store. Let the outcome profile of store i be denoted $y_i = (y_{i,1},..., y_{i,m})$ where $y_{i,t} \in \{a,b\}$ $(1 \le t \le m)$ denotes the tth lottery outcome of store i where a denotes winning and b denotes not winning. The consumer updates his prior in a Bayesian manner, for each store's win rate θ_i after observing a set of outcomes $y_{i,t} \in \{a,b\}$. The outcomes are drawn independently across stores. The consumer then purchases from the store which in his calculation, is most likely to win in the next lottery round, given these updated beliefs.

2.2 General Result

The general result for n stores, K winning rates and general priors is summarized in Proposition 1.

Proposition 1: Let $M^* = \dfrac{2\left(\sum_{k=1}^{K} \pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K} \pi_i(r_k)\cdot r_k\right)}{\left(\sum_{j=1}^{K}\sum_{k=1}^{K} \pi_i(r_j)\pi_i(r_k)\cdot(r_j - r_k)^2\right)}$. Then $\forall i, j = 1,...,n$, when $M > M^*$, $P(y_{i,2} = a \mid y_{i,1} = a) > P(y_{j,2} = a \mid y_{j,1} = b)$; when $M < M^*$, $P(y_{i,2} = a \mid y_{i,1} = a) < P(y_{j,2} = a \mid y_{j,1} = b)$.

Proof: The proof is provided in Appendix A.

Proposition 1 shows that the Lucky Store Effect depends critically on whether the size of the pot for each store exceeds some threshold level, M^*, which is a function of the set of possible (perceived) winning rates of the different stores, r. In other words, the size of the Rabin (2002) psychological pot determines whether the Lucky Store Effect dominates, or whether a belief in reversals stemming from Gambler's Fallacy beliefs will dominate. In particular, a notable relationship is that the minimal pot size for the Lucky Store Effect, M^*, conditional on the possible winning rates, is *decreasing* in the gap between the possible ex-ante win rates.

In other words, when the decision-maker is trying to make inference between stores which are ex-ante perceived as being very different, in order have the Lucky Store belief, he need not be as classically Bayesian (ie. as "unbiased" in the Law of Small Numbers sense), compared to the case where the win rates are ex-ante perceived to be quite similar. This is because observing a winning store when stores are potentially very different, is more "informative" in the belief updating process, and can counteract the decision-maker's underlying tendency to believe that a winning ticket will not be drawn from the winning store's pot again.

Another way of phrasing this is that when win rates across stores are perceived to have little variation, a Law of Small Numbers believer will tend to believe in reversals in outcomes, rather than in Lucky Stores. The origin of such belief in reversals under cross-sectional comparisons, is the Gambler's Fallacy in the Law of Small Numbers framework.

2.3 Example: 2 Stores, Uniform Priors

To make ideas more concrete, we show the detailed analysis for 2 stores ($n = 2$), two possible winning probabilities $\theta_i \in \{r_1, r_2\}$, where $r_1 > r_2$ and just one previous lottery outcome realization ($m = 1$, the condition for the Lucky Store effect). We assume a uniform prior across the possible winning probabilities, in this case $\pi_i(r_k) = \frac{1}{2}, \forall i, k$. While the general case is shown in Appendix A, this special case is sufficient to fully understand the intuition of the Lucky Store effect.

Just as in the generalized set up, the consumer uses Bayes' rule to calculate the posterior likelihood of each possible winning rate for each store $\theta_i \in \{r_1, r_2\}$. We consider the non-trivial scenario in which only one of the two stores wins, consistent with one store winning the jackpot. Since the pots of results are independent across the two stores, although the set of observation by the consumer is $\{y_{1,1}, y_{2,1}\}$, this is equivalent to merely incorporating the information about $y_{i,1}$ into the consumer's belief about store i, for $i = 1, 2$.

For the case in which store i has won, the consumer's belief about store i being the lucky store (that is $\theta_i = r_1$) is

$$P(\theta_i = r_1 \mid y_{i,1} = a) = \frac{\frac{1}{2} \cdot r_1}{\frac{1}{2} \cdot r_1 + \frac{1}{2} \cdot r_2} = \frac{r_1}{r_1 + r_2} > \frac{1}{2}$$

His belief about store i being the unlucky store (that is $\theta_i = r_2$) is

$$P(\theta_i = r_2 \mid y_{i,1} = a) = \frac{r_2}{r_1 + r_2} < \frac{1}{2}.$$

For the case in which store i has not won, similarly we have

$$P(\theta_i = r_1 \mid y_{i,1} = b) = \frac{\frac{1}{2}(1 - r_1)}{\frac{1}{2}(1 - r_1) + \frac{1}{2}(1 - r_2)} = \frac{1 - r_1}{2 - r_1 - r_2} < \frac{1}{2}$$

and

$$P(\theta_i = r_2 \mid y_{i,1} = b) = \frac{1 - r_2}{2 - r_1 - r_2} > \frac{1}{2}.$$

Now the consumer determines which of the stores is more likely to sell the winning ticket in the next lottery round. Without loss of generality, we assume store 1 won in round 1 and store 2 did not.

The likelihood that Store 1 will win in this round, given that Store 1 won in the previous round is given by

$$P(y_{1,2} = a \mid y_{1,1} = a)$$
$$= P(\theta_1 = r_1 \mid y_{1,1} = a) \cdot P(y_{1,2} = a \mid \theta_1 = r_1, y_{1,1} = a) + P(\theta_1 = r_2 \mid y_{1,1} = a) \cdot P(y_{1,2} = a \mid \theta_1 = r_2, y_{1,1} = a)$$

$$= \frac{r_1}{r_1+r_2} \cdot \frac{r_1 M - 1}{M-1} + \frac{r_2}{r_1+r_2} \cdot \frac{r_2 M - 1}{M-1}$$

$$= \frac{(r_1^2 + r_2^2)M - (r_1 + r_2)}{(r_1 + r_2)(M-1)}.$$

The likelihood that Store 2 will win in this round, given that Store 2 did not win in the previous round is given by

$$P(y_{2,2} = a \mid y_{2,1} = b)$$

$$= P(\theta_2 = r_1 \mid y_{2,1} = b) \cdot P(y_{2,2} = a \mid \theta_2 = r_1, y_{2,1} = b) + P(\theta_2 = r_2 \mid y_{2,1} = b) \cdot P(y_{2,2} = a \mid \theta_2 = r_2, y_{2,1} = b)$$

$$= \frac{1-r_1}{2-r_1-r_2} \cdot \frac{r_1 M}{M-1} + \frac{1-r_2}{2-r_1-r_2} \cdot \frac{r_2 M}{M-1}$$

$$= \frac{(r_1 + r_2 - r_1^2 - r_2^2)M}{(2-r_1-r_2)\cdot(M-1)}.$$

All that remains is to compare the attractiveness of store 1 which won in the previous round to the attractiveness of store 2 which did not win in the previous round. If $P(y_{1,2} = a \mid y_{1,1} = a)$ minus $P(y_{2,2} = a \mid y_{2,1} = b)$ is positive, then the Lucky Store effect holds.

$$P(y_{1,2} = a \mid y_{1,1} = a) - P(y_{2,2} = a \mid y_{2,1} = b)$$

$$= \frac{(r_1^2 + r_2^2)(2-r_1-r_2)M - (r_1+r_2)(2-r_1-r_2) - (r_1 + r_2 - r_1^2 - r_2^2)(r_1+r_2)M}{(r_1+r_2)(2-r_1-r_2)(M-1)}$$

$$= \frac{(r_1 - r_2)^2 M - (r_1+r_2)(2-r_1-r_2)}{(r_1+r_2)(2-r_1-r_2)(M-1)}$$

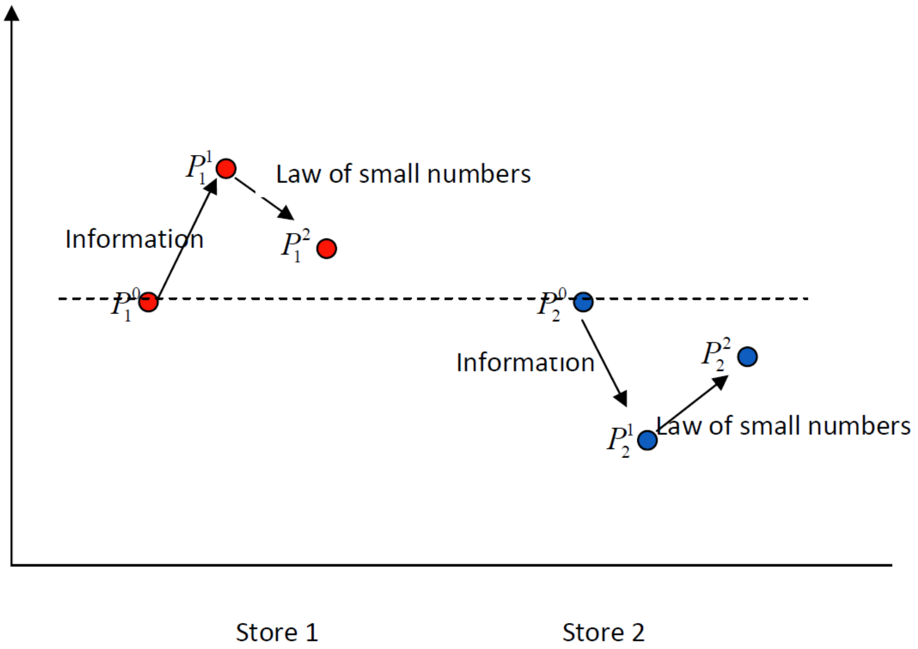
This expression is positive for sufficiently large values of M, specifically $M > \frac{(r_1+r_2)(2-r_1-r_2)}{(r_1-r_2)^2}$.

In other words, as long as the size of the pot from which the outcomes are drawn for each store is sufficiently large, the store that sold the winning ticket in the previous round appears more attractive to the consumer. Note that this cutoff value for M is smaller when the possible winning rates r_1 and r_2 are farther apart.[6] We note that the probability structure of a lottery game, where the odds of winning are quite small, a large M (number of balls) will typically be needed in order to represent the true odds in a pot. Thus the aforementioned condition on M should not be a binding constraint.

[6] For example, in the case where $r_1 = 0.85$ and $r_2 = 0.2$, the minimum M required is 2.36 (or 3 since integers are required). For $r_1 = 0.55$ and $r_2 = 0.5$, the minimum M is 399. The larger of these range of pot sizes are reasonable given the true lottery odds which are quite small, combined with the assumption in Rabin (2002) that true odds are represented using an integer number of outcomes in the pot.

Figure 1: Strength of the Direction of Beliefs

y-axis: belief on future winning probability if Store 1 is the winner in the previous round



The intuition for this result is that under reasonable values of the pot size, the information effect of a particular store winning (thus making that store more attractive), exceeds the effect of the law of small numbers (which makes that store less attractive, as it is seen as having a 'winning ball' removed from the pot). That is, although Store 1 has won the lottery in the previous round, making it less likely for that same store to win in the next round (by the law of small numbers), the positive news that Store 1's previous round win reveals about its true winning rate in the mind of the consumer, is indeed stronger. This is shown in Figure 1 for illustrative purposes only (not drawn to scale).

3. Field Setting and Data

We now turn to the empirical evidence for these representativeness-based belief fallacies in the field. We examine buyers' purchasing tendencies in a popular online market for lottery tickets. The online lottery market is a formalized collective lottery purchase arrangement, in which ticket sellers propose number combinations to the buyer side of the market. Buyers have the opportunity to purchase from one of thousands of sellers offering collective betting contracts.

The setting is unique in that we can observe both individual sellers' winning statistics, and the exact number combinations being offered in their collective betting contract. When considering ticket buyers' decisions on which sellers to purchase from, the lottery outcomes in each round provide an exact natural experiment for testing the Gambler's Fallacy and Lucky Store Effect – in each round the winning numbers and the winning ticket sellers are randomly chosen. To test the Gambler's Fallacy we use an approach similar to that used in Clotfelter and Cook (1993) and Terrell (1994) which found Gambler's Fallacy in lottery number choices. We follow their approach in examining the popularity of previously winning lottery numbers. In testing for the Lucky Store Effect, we examine the popularity of previously winning lottery ticket sellers, as in Guryan and Kearney (2008).

Sections 3.1 to 3.3 describe the lottery game, the online marketplace, and the data. Section 4 describes our empirical approach which shows that both the Gambler's Fallacy and the Lucky Store Effect are prevalent in the data.

3.1 The SSQ Lottery Game

The gaming rules of the SSQ Chinese national lottery are similar to those of other popular lotteries, such as the Powerball in the United States and the LottoMax in Canada. SSQ stands for Shuang Se Qiu (双色球), which means dual-colored balls in Chinese. Each ticket is sold for 2 Yuan (about $0.32 USD). It requires players to pick numbers from two groups of numbers. Players need to pick six integers without replacement from the range 1 to 33, which are referred to as the "red numbers". Players also need to pick one integer from 1 to 16, referred to as the "blue number". To win the Grand prize jackpot, a player needs to match all 7 numbers randomly drawn as the winning number combination.

The SSQ has 6 levels of prizes, which are shown in Table 1. The Grand prize is shown in the first row. The second prize is won by matching all 6 red numbers but not the blue one. Note that the first and the second prizes are pari-mutuel, since the final reward depends on the number of winners and the prize pool for each payout. The third to sixth prizes are non pari-mutuel fixed prizes.

Table 1: SSQ Policies

Award level	Winning conditions		Prize distribution
	Number of Red balls correct (out of 6)	Blue ball correct?	
Grand prize	6	Yes	If the rollover money from the last jackpot is less than 100 million Yuan, then the grand jackpot winners will split the rollover from the previous draw and the 70% from the "high prize pool". If the prize is more than 5 million Yuan, each winning ticket will only be worth 5 million Yuan. If the rollover money from the last jackpot is at least 100 million Yuan or more, there is a two part prize portfolio. The winners split the rollover money from the previous draw from the "high prize pool". With each prize, a maximum of 5 million Yuan is paid (total of 10 million Yuan).
Second prize	6	No	30% of current Grand prize
Third prize	5	Yes	Fixed amount of 3000 Yuan per winning lottery ticket
Fourth prize	5	No	Fixed amount of 200 Yuan per winning lottery ticket
	4	Yes	
Fifth prize	4	No	Fixed amount of 10 Yuan per winning lottery ticket
	3	Yes	
Sixth prize	2	Yes	Fixed amount of 5 Yuan per winning lottery ticket
	1	Yes	
	0	Yes	

3.2 The Online Lottery Market

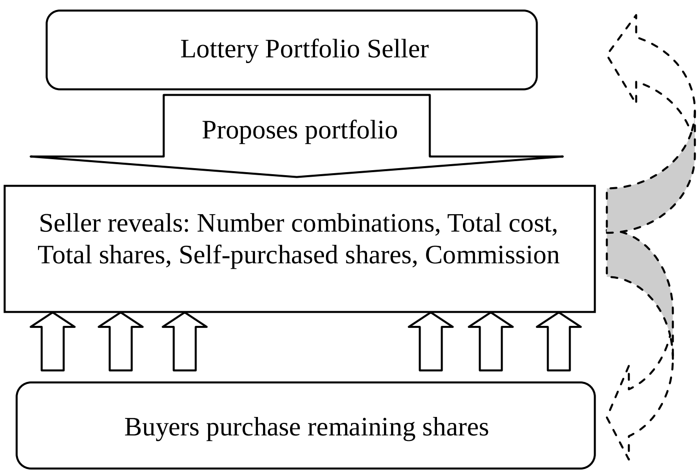
The lottery industry in China is one of the largest lottery systems in the world, and SSQ is one of the most popular lottery games in China. The online peer to peer lottery market we examine takes place on the website of Taobao Lottery, the largest online official retailer of SSQ lottery tickets. A small number of other studies have examined different aspects of behavior in this Taobao online lottery market. Siu, Sun and Yuan (2014) examine how much customers are willing to pay in commission fees to buy tickets from previously winning ticket sellers. Lien and Yuan (2015b) study sellers' responses to customers' biased beliefs about lottery outcomes, finding that their choices in commissions, portfolio size, and other features are consistent with responding to consumers with biased beliefs. In the current paper, we focus on lottery ticket buyers' beliefs as revealed by their choices of lottery sellers and number combinations.

The nature of this peer to peer lottery market can best be understood in the context of what is known as "collective purchases" of lottery tickets in the US. In US lottery games (ex. Powerball), it is common for family, friends, or co-workers to pool their money together to buy a certain number of lottery tickets. In the event of a win, those who had contributed money to the pool receive a share of the winnings (see Humphreys and Perez, 2013 for a discussion).The lottery market in Taobao Lottery is a formalized version of this arrangement, where institutional rules are set and enforced by the website. Collective bettors do not need to know one another personally, but can interact in the online marketplace.

There are very low if any barriers to entry to becoming a seller. The only prerequisite, which is also in place for buyers on the site, is having a Taobao account which in itself is free of charge. Once a user decides to become a seller, it is required however, that they self-invest on at least 1% of the shares in their own proposed portfolio, or the same percentage as the commission rate they charge buyers, whichever is larger. This requires that sellers have a stake in the portfolios they are offering on the market. Sellers can self-invest up to 100% of the portfolio, equivalent to the case of non-collective purchase. Sellers are also allowed to propose more than one portfolio. In each lottery round there are about 10,000 sellers participating, with the great majority of them posting a single portfolio.

Figure 2: Structure of Taobao Online Lottery Market

(In event of lottery win, returns made to buyers proportionally according to purchased shares)



Portfolios may be open for sale directly after the previous round of lottery is over. Once a portfolio is posted online, the seller cannot change it. Similarly, once buyers purchase shares, they cannot request a refund. Selling of shares closes 3 hours before the actual lottery draw. If by that time, a seller's portfolio is not 100% sold out, the portfolio is cancelled and all investors have their money automatically refunded to their account.

Once a portfolio is sold out, Taobao receives the money invested in it. Taobao then purchases tickets on behalf of the seller, and is responsible for distributing the winnings among investors in the event of a win. Taobao thus eliminates any concerns from buyers about the trustworthiness of the seller. For providing this service, Taobao receives a commission from the official lottery authority, and they have an additional benefit of advance cash flow when the lottery investment is made. Figure 2 illustrates the structure of the Taobao online lottery marketplace.

3.3 Data

Our data consists of 4,529,730 observations over 25 rounds of lottery games. We observe each seller's portfolio on the market including the exact numbers chosen, for each of these rounds. We also observe several portfolio-specific control variables critical to our analysis, shown in Table 2.

Table 2: Main Variable Definitions

SimIndex	Similarity Index of a lottery ticket, measuring how similar a lottery portfolio is to the previous winning lottery number combination
WinRate	Winning rate of the seller, in other words his total amount won last round, divided by his total wager last round
Commission	Commission rate charged by seller, as a percentage of the total winnings of the portfolio
Size	Total amount of money in the lottery portfolio
Shares	Total number of shares in the portfolio
Price	The price of a single share
SelfBuy	The percentage of shares purchased by the seller

The sample summary statistics for these and other variables are shown in Table 3. In each round over 2 million lottery portfolios are put on the market by some 41,000 sellers. The win rate of sellers is consistent with the theoretical expected rate of the lottery game. Portfolio size and number of tickets included vary widely. There is also substantial variation in self-investment behavior and commissions charged.

Table 3: Summary Statistics

Variable	Minimum	Mean	Median	Std dev	Maximum	Observations
Similarity Index	0	.098	.083	.082	1	283083
WinRate (all)	0	.58	0	35.41	14930.84	248523
Excluding Jackpot	*0*	*.229*	*0*	*1.87*	*376.25*	*248460*
Commission	0	.0575	.08	.0437	.10	301982
Size (all)	8	774.32	18	8035.93	937770	301982
Successful Only	*8*	*53.46*	*14*	*400.31*	*58848*	*233481*
Shares	1	911.67	50	13859.6	2491060	301982
Price	.2	3.91	.5	87.55	12400	301982
Self-Investment	.01	.55	.6	.27	1	301982
Number of Tickets in Each Package	4	387.16	9	4017.96	468885	301982
Number of Buyers for Each Package	1	9.76	5	46.39	6118	301982
Portfolios in Each Round	18899	20132	20260	1015	22242	301982

Time Expose (Hours)	.01	25.67	24.02	19.532	70.49	301982
Sold-out Portfolios			77%			301982
Total Number of Sellers:						41418

4. Measurements and Empirical Strategy

Our main objective is to understand how lottery buyers' purchase decisions are affected by the return rate of the seller and the composition of lottery numbers in the portfolio. Both of these factors should have no effect on ticket buying in the case of standard beliefs about the lottery process. In fact, since the SSQ lottery is pari-mutuel, subscribing to common belief fallacies is harmful to the buyer in expectation, since adhering to popular choices results in greater chance of prize-sharing (a point raised in Terrell, 1994).

To examine the Gambler's Fallacy, we take an approach similar to Clotfelter and Cook (1993) and Terrell (1994) in examining lottery buyers' number choices. Due to the fact that ticket sellers are offering number combinations rather than just individual numbers, we would like to have a measure of how similar a particular combination is to the previous winning ticket. Thus for each ticket sold in the online market, we calculate a Similarity Index to summarize its numerical resemblance to the previous winning ticket. We consider only the similarity level to the most recent winning ticket. As previous literature (Clotfelter and Cook, 1993; Terrell, 1994) has shown, the immediately previous ticket tends to be most influential and its effect tapers over time as new winning tickets are generated in each round of the lottery. If lottery ticket buyers have rational beliefs, they should not be affected by the composition of the lottery numbers in the portfolio. As in Terrell (1994), the Grand and First prizes of our lottery setting are pari-mutuel, so that an expected monetary penalty exists for choosing according to belief fallacies when other players do so as well.

To examine the Lucky Store Effect, we simply compute each seller's winning rate in the most recent realized round. According to the Lucky Store effect as well as our theoretical analysis in Section 2, the outcome of even the most recent round itself should have a positive effect on the popularity of a winning seller, if buyers have Representativeness Bias.

As our results in Section 5 show, controlling for features of the lottery portfolios such as size, commissions, and other control variables, buyers gravitate towards portfolios which are dissimilar to immediately previous winning lottery tickets (Gambler's Fallacy), and simultaneously gravitate towards portfolio sellers who have had positive win rates in the previous round (Lucky Store Effect). These findings are robust to empirical specifications and the presence of portfolio-specific factors.

4.1 Detecting the Gambler's Fallacy

Each lottery ticket consists of seven numbers with six red balls and one blue ball. We use the following notation to denote a lottery number combination for a single $Ticket_i$, where the 7th number is the blue ball:

$$Ticket_i = \{ b_1, b_2, b_3, b_4, b_5, b_6 | b_7 \}$$

We aim to measure how similar a lottery ticket i is to the winning numbers in round t. We use $S\_Index_{i,t}$ to denote this Similarity Index.

We assume that the winning numbers in round t are the following:

$$\{ w_1^t, w_2^t, w_3^t, w_4^t, w_5^t, w_6^t | w_7^t \}$$

Let $I(\cdot)$ be an indicator function, where if the event in the bracket is true, the value is one, and the value is zero otherwise.

We define $S\_Index_{i,t}$ as follows:

$$S\_Index_{i,t} = \frac{1}{2}\left(\frac{\sum_{n=1}^{6}\sum_{m=1}^{6} I(b_n = w_m^t)}{6}\right) + \frac{1}{2}\, I(b_7 = w_7^t)$$

The Similarity Index consists of two parts: the first term is the similarity on the red numbers and the second term is the similarity on the blue division, where for simplicity we use the same weight on the red and blue sections. Thus, if the seventh blue number b_7 is the same winning number w_7^t in round t, by the definition above, this similarity contributes to a half of the total index. On the red ball component, according to the above definition, if any of $\{ w_1^t, w_2^t, w_3^t, w_4^t, w_5^t, w_6^t\}$ appears in $\{ r_1, r_2, r_3, r_4, r_5, r_6\}$, it will contribute to $\frac{1}{2} * \frac{1}{6} = 1/12$ of the total index. It is easy to see that the $S\_Index_{i,t}$ for the lottery ticket $\{ w_1^t, w_2^t, w_3^t, w_4^t, w_5^t, w_6^t | \, w_7^t \}$ is one. This highlights the emphasis on the color of the numbered balls reflected in the name of the lottery game (Dual colored balls), however our results are robust to changes to this exact specification of the Similarity Index.

4.2 Detecting the Lucky Store Effect

The Lucky Store Effect implies that lottery players will believe that previous sellers of winning tickets are more likely to sell winning tickets again. In other words, players will respond to the past returns of lottery sellers, in spite of the fact that no seller has any true advantage over other sellers. It is easy to calculate the return rate of the sellers from our data. The return rate of seller j based on his performance in the period immediately prior is defined as follows:

$$WinRate\_j = \frac{TotalWin\_j}{TotalWager\_j}$$

where $TotalWin\_j$ is the total amount of winning money in the previous round for seller j and $TotalWager\_j$ is the total lottery investment by seller j in the previous round.

4.3 Empirical Approach for Sales Measures

We analyze the data at the lottery ticket portfolio level, interpreting portfolios as "stores" throughout our analysis.[7] A key measure of the attractiveness of lottery tickets and ticket sellers to consumers is whether the lottery portfolio is successfully sold out or not. A significant proportion of portfolios are not successfully sold (and thus cancelled) in each round, and we can infer buyers' beliefs by examining the fraction of a portfolio's shares which are actually sold in the market. We allow buyers' beliefs to be inferred by the market popularity of the portfolios, as determined by the sales progress towards being fully sold out in the online marketplace. The results thus reflect the aggregate behavior of the buyer side of the market.

Our baseline empirical specification is a standard Tobit model for sales progress, to analyze how buyers choose lottery portfolios, given the Win Rate of sellers and the Similarity Index of the lottery portfolios. We also provide several robustness checks which show that these baseline results remain for alternative specifications of the dependent variable and subsets of our sample.

The main Tobit regression model, which is implemented at the portfolio level is as follows:

$$PROGRESS^* = \beta_0 + \beta_1 WinRate + \beta_2\, S\_Index + \beta_3\, CONTROLS + \varepsilon$$

$$PROGRESS = \begin{cases} PROGRESS^* & IF \quad PROGRESS^* < 100 \\ 100 & IF \quad PROGRESS^* \geq 100 \end{cases}$$

[7] "Stores" here can be interpreted as the entity which consumers can decide to make a purchase from, which accords with the concept of portfolio, even though a single seller can technically offer multiple portfolios.

$PROGRESS$ represents the progress rate of a portfolio, or the fraction of the shares sold for this portfolio at closing time. $PROGRESS^*$ is the latent variable which we cannot observe, due to the fact that once a portfolio is complete sold out, no further shares can be sold. $WinRate$ and S_Index are the variables defined previously in Sections 4.1 and 4.2.

$CONTROLS$ is a vector of portfolio-specific control variables: $COMMISSION$ is the seller's commission fee, $SIZE$ is the total amount of money to be collected if the lottery portfolio is completely sold out, *SHARES* represents the total shares in the portfolio, *PRICE* is the value for each single share, *SELFINVEST* measures the percentage of the shares purchased by the seller himself. These variables which are chosen by sellers, and we control for them in the empirical specifications since buyers' decisions may also be influenced by these features. These are also the key variables buyers observe when they make the decision of which ticket portfolios to purchase.

5. Empirical Results

5.1 Sales Progress Results

The Gambler's Fallacy and Lucky Store Effect can be seen clearly from the Tobit regression results. Table 4 shows the results for the entire sample, with the Similarity Index as the measure of the Gambler's Fallacy. To account for actual differences in portfolio or "store" features observable to the consumer, which may affect their buying choices, we include portfolio size, commission, price per share and the seller's self-investment in the portfolio as control variables. These variables are added sequentially to the model of portfolio sales progress, so that the impact of store features can on our coefficients of interest can be observed.

The coefficient for WinRate is significantly positive across all specifications, indicative of the Lucky Store Effect. That is, the higher the previous round's win rate of the seller, the greater proportion of tickets he sells in the current round. We note that the directions of the coefficients on the control variables match the standard intuition. Portfolios where the seller invests more have higher sales progress. Commissions, being a type of price, are negatively associated with sales progress. The size of the portfolio also has a negative relationship with progress, which is intuitive, since more tickets need to be sold for each percentage point of sales progress. Importantly, in all specifications, the coefficient on the seller's most recent win rate remains significantly positive. The Lucky Store Effect is thus robust to portfolio-specific or store-specific characteristics, including portfolio numerical similarity as represented by the Similarity Index.

We now turn to the Gambler's Fallacy result, which is given by the coefficient on the variable Similarity Index in each of the specifications, also in Table 4. Recall that the Gambler's Fallacy predicts that the more similar the current number portfolio being offered is to the previous winning ticket, the less popular it will be. Indeed, the coefficients on S_Index are consistently and significantly negative across all specifications, controlling for observable portfolio or store characteristics, including the previous Win Rate. Based on Table 4, the presence of a simultaneous Lucky Store Effect and Gambler's Fallacy in ticket buyers' purchase decisions is apparent and stable across all model specifications

5.2 Robustness Checks

While the presence of the Lucky Store Effect and Gambler's Fallacy are clear from our main results in Table 4, several robustness checks are in order. These include checks regarding alternative dependent variables, limited processing of the lottery customers, and alternative measures of numerical similarity. We discuss each of them in this section.

Alternative Dependent Variables

Our Tobit specification has the limitation that the identification of the Lucky Store Effect and Gambler's Fallacy is largely based on variation on unsuccessful portfolios' sales progress. To check the robustness of our result to the specification of the dependent variables intended to measure buyer popularity, we also consider some alternative measures.

First, we may want to merely consider whether a portfolio sold out or not, as a measure of its popularity among lottery buyers. Our Probit specification, shown in Table 5, focuses attention on the comparison between successfully sold out portfolios and unsuccessful portfolios, by using the sold out indicator as our dependent variable. Compared to the Tobit specification, the measured effects do not rely on the variation in sales progress among unsuccessful portfolios. The direction and significance of our coefficients of interest remain similar to those in the Tobit specification.

It may also be informative to provide a specification which focuses on the portfolios which do in fact successfully sell-out. To check this, we implement an ordinary least squares regression with the realized amount of time taken for the portfolio to sell-out, as the dependent variable. We include only those portfolios which were actually sold-out in the regression. We note here that compared to using sales progress or the sold-out indicator as the dependent variable of interest, the time until sold-out is a less well-defined measure of portfolio popularity. The reason is that it may not be clear whether a portfolio is slower to sell-out due to being less popular, or due to other factors such as customers' desire to delay their purchase of that portfolio, rather than avoiding purchasing it altogether. Furthermore, in the marketplace, speed of selling tickets is not monetarily rewarded from the perspective of the sellers.

Nevertheless, Table 6 shows that portfolios with higher win rates do take significantly less time to be sold-out, indicating that the Lucky Store effect is also robust to this measure of popularity. The coefficient on the Similarity Index is not significant in this specification, indicating that numerical differentiation from previous winners assists in the final success of portfolios, but not the speed with which the portfolio is sold out.

Table 4: Tobit Regression: Dependent Variable: Sales Progress

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WinRate	0.489***		0.503***	0.211***	0.210***	0.208***	0.208***	0.208***
	(0.0103)		(0.0103)	(0.00642)	(0.00630)	(0.00628)	(0.00628)	(0.00628)
Similarity Index		-0.594***	-0.812***	-0.175***	-0.170***	-0.161***	-0.161***	-0.160***
		(0.0450)	(0.0452)	(0.0315)	(0.0309)	(0.0309)	(0.0308)	(0.0308)
Self-Investment				3.471***	3.606***	3.581***	3.581***	3.580***
				(0.0131)	(0.0134)	(0.0135)	(0.0135)	(0.0135)
Commission					-0.0346***	-0.0348***	-0.0348***	-0.0348***
					(0.000587)	(0.000587)	(0.000587)	(0.000587)
Size						-0.00000528***	-0.00000607***	-0.00000560***
						(0.000000379)	(0.000000490)	(0.000000499)
Shares							0.000000354*	0.000000264
							(0.000000139)	(0.000000140)
Price								-0.0000874***
								(0.0000175)
Constant	1.771***	1.905***	1.860***	-0.0477***	0.0752***	0.0873***	0.0869***	0.0875***
	(0.0138)	(0.0149)	(0.0147)	(0.0103)	(0.0103)	(0.0103)	(0.0103)	(0.0103)
N	238317	238317	238317	238317	238317	238317	238317	238317

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5: Probit Model, Dependent Variable: Sold Out Indicator

	(1)	(2)	(3)	(4)	(5)	(6)
Big Win Indicator	0.504***	0.432***	0.438***	0.450***	0.451***	0.451***
	(0.0107)	(0.0124)	(0.0125)	(0.0126)	(0.0126)	(0.0126)
Similarity Index	-0.501***	-0.303***	-0.291***	-0.349***	-0.349***	-0.350***
	(0.0370)	(0.0500)	(0.0503)	(0.0507)	(0.0507)	(0.0507)
Self-Investment		4.507***	4.774***	4.612***	4.610***	4.609***
		(0.0177)	(0.0186)	(0.0189)	(0.0189)	(0.0189)
Commission			-0.0504***	-0.0479***	-0.0478***	-0.0476***
			(0.000937)	(0.000944)	(0.000944)	(0.000945)
Size				-0.00178***	-0.00161***	-0.00168***
				(0.0000465)	(0.0000637)	(0.0000650)
Shares (10^{-5})					-7.53***	-5.78*
					(1.97)	(1.98)
Price						0.00564***
						(0.00110)
N	242667	242667	242667	242667	242667	242667

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Ordinary Least Squares, Dependent Variable: Time until Portfolio Sold Out (Successful Portfolios Only)

	(1)	(2)	(3)	(4)
WinRate	-0.122*		-0.122*	-0.298***
	(0.0570)		(0.0570)	(0.0553)
Similarity Index		0.543	0.548	0.261
		(0.536)	(0.536)	(0.520)
Size				0.0329***
				(0.000974)
Self-Investment				-11.59***
				(0.122)
Commission				-0.00116
				(0.00579)
Price				-0.0984***
				(0.00561)
Constant	7.405***	7.346***	7.361***	14.32***
	(0.0939)	(0.103)	(0.103)	(0.130)
N	187179	187179	187179	187179
adj. R^2	0.013	0.013	0.013	0.072

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Limited Processing

One potential concern is that lottery ticket buyers in the online market face a very large number of "stores" to choose from. In this case, limited processing ability may lead the consumers to use stores' previous win rates and numerical similarity as heuristics for their purchase decisions, when faced with a number of stores too large to process. In addition to the possibility of limited processing by the decision-maker, there is also the physical issue of how many stores a consumer can plausibly observe on this computer screen when making a decision.

Taobao Lottery in fact has several buttons which consumers can use to sort the lottery stores by specific lottery portfolio features. These features are: winning rate, size of the portfolio, commission, share price and sales progress. By using this sorting feature, a consumer would see the top-most (or bottom-most) stores based on the chosen criteria in their search results. Although we cannot observe whether buyers actually used the sorting feature, to check the robustness of the Lucky Store Effect and Gambler's Fallacy against the possibility that consumers sort based on one of these features in making their purchase decisions, we can subdivide our sample of portfolios based on these sortable features and check whether the coefficients on win rate and similarity remain significant. If our results are robust to the sample restrictions, it implies that even among sorting results which may serve to restrict the consumer's choice set in the case of limited processing, buyers still exhibited the Lucky Store Effect and Gambler's Fallacy within those potential choices.

To begin, we consider whether the Lucky Store Effect and Gambler's Fallacy still hold even if the consumer had sorted the lottery stores on the website by more recent submissions added to the website. We consider the set of portfolios with less than 20 hours remaining until the lottery draw, in case these portfolios or buyer responses to them might be systematically different than for the entire sample of portfolios. As Table B1 in Appendix B shows, our coefficients on Win Rate and Similarity Index, while reduced slightly in magnitude compared to the baseline results, are still highly significant.

Consumers may find some particular portfolio features attractive, and may thus use the sorting feature to assist in making their decisions such as, low commissions (they will not have to pay the seller as high a fraction of winnings in the event of a win), low price per share (each share is more affordable), and small portfolio size (winnings do not need to be shared with as many other buyers). To check the robustness to these factors, we consider the subsample of portfolios with commissions less than 5%, price per share less than 1 yuan, and portfolio total size less than 200 yuan, which are shown in Tables B2, B3 and B4 of Appendix B, respectively. Just as in our baseline regressions, the other observable portfolio features are added in to the regression sequentially as control variables. The Lucky Store Effect and Gambler's Fallacy findings remain robust and of similar magnitude to the regression on the full sample.

Finally, we consider whether the tendency to gravitate towards lottery stores which have higher winning rates holds for the subset of lottery stores that do *not* have especially high return rates. This is to ensure that the effects we find are not only due to outlier portfolios with extraordinary eye-catching win rates. We restrict the sample to those stores whose win rate is less than 10%, which consists of about 213,000 portfolios, and the results are shown in Table B5 in Appendix B. The Lucky Store Effect is many times stronger when we restrict the regression to stores without any spectacular return rates, due to the presence of many zero return rates in this subset of stores. This implies that even among mediocre-performing stores, consumers still gravitate heavily towards stores with higher prior win rates.

Robustness of Lucky Store Effect to Similarity Measure

We may also be interested in knowing how sensitive the result is to the similarity levels of the lottery tickets sold on the Gambler's Fallacy side. In terms of the Lucky Store Effect, since the win rates of the great majority of sellers is zero, winning *any* positive amount in the previous round allows sellers to distinguish themselves in the minds of the buyers. However, the relationship could be more subtle for the similarity of numbers in the lottery ticket portfolios.

Tables B6 and B7 in the Appendix show the effects of the Similarity Index on sales progress by replacing the continuous variable Similarity Index from Table 4, with indicator variables for similarity levels passing particular thresholds. Table B6 considers an indicator variable for similarity index higher than 0.2, while Table B7 considers similarity levels higher than 0.5 (or in other words at least half of the numbers in the ticket are prior winning numbers). A comparison of the coefficients from Tables B6 and B7 implies that the Gambler's Fallacy effect is larger for more similar portfolios. That is, highly similar tickets are more drastically avoided on the margin than moderately similar tickets.

6. Conclusions

The Gambler's Fallacy and the Hot Hand Fallacy on first impression seem contradictory to one another, since one implies belief in negative correlation over outcomes while the other implies the opposite. Yet recent evidence suggests they can in fact be interpreted as being interrelated, as part of the same underlying psychological phenomenon of representativeness bias (also known as belief in the Law of Small Numbers). This approach was proposed in Rabin (2002) and Rabin and Vayanos (2010), which show that the Hot Hand Fallacy can be generated if there is uncertainty about true probabilities, especially over longer sequences of outcomes, when the decision-maker in fact believes in the Gambler's Fallacy over short sequences of outcomes. Existing empirical evidence on the Gambler's and Hot Hand Fallacies also focuses on this time dimension (Rao, 2009; Galbo-Jorgensen et al, 2012).

We extend the current evidence on representativeness biased beliefs by extrapolating insights and evidence to *cross-sectional* observations of outcomes.[8] By extending the model of Rabin (2002) to the case where a decision-maker with Law of Small Numbers beliefs must decide among a set of possible 'stores' to choose from, we show that for a sufficiently large psychological pot size, the information content that a decision-maker derives from a store's lottery win, exceeds the decreased perceived probability that the store will win again stemming from Gambler's Fallacy beliefs. In addition, we show that ask the perceived difference between stores is larger, the minimum psychological pot size to obtain the Lucky Store Effect in the Law of Small Numbers framework is decreasing. In other words, decision-makers can believe very heavily in the Law of Small Numbers, and still believe in Lucky Stores, if the perceived difference between stores is large. On the other hand, when there is very little perceived variation among stores, a decision-maker needs to be quite classically Bayesian to allow the "information" effect to dominate the belief in Gambler's Fallacy.

We then provide new evidence for such belief fallacies in the field, analyzing behavior in a popular collective betting market for the Chinese national lottery. In this marketplace, we find that in accordance with the Gambler's Fallacy and the Lucky Store Effect, lottery portfolios with dissimilar number picks compared to the previous winning ticket, and lottery portfolios being sold by sellers who have positive win rates from the previous lottery round respectively, are indeed more successful in their ticket sales. Our empirical result replicates the Lucky Store Effect of Guryan and Kearney (2008), while providing new evidence on buyers' simultaneous belief in the Gambler's Fallacy over lottery number picks. The empirical pattern also accords with the intuition in our theoretical model: while it is reasonable to believe that decision-makers may have diffuse priors over the winning rates of stores, it is arguably more plausible for decision-makers to believe individual lottery number digits are ex-ante very similar. This can explain why we continue to observe Gambler's Fallacy beliefs

[8] For an evaluation of representativeness bias over sets of outcomes in the lottery number setting, see Lien and Yuan (2015a).

over number picks, while observing the Lucky Store Effect over winning stores.

We would like to conclude by discussing the relationship between the Lucky Store Effect and the Hot Hand Fallacy. The Hot Hand Fallacy has traditionally referred to the belief in streaks of outcomes, although discussion has generally been limited to events occurring over time for a single random process, rather than in the cross-section over multiple possible processes. In Rabin (2002) and Rabin and Vayanos (2010), the Hot Hand Fallacy is generated from Gambler's Fallacy beliefs when true probabilities are uncertain to the decision-maker. The Lucky Store Effect is a phenomenon which occurs purely in the cross-section over several possible random processes (or 'stores') after just one previous observation.

Like the Hot Hand Fallacy, the Lucky Store Effect predicts that decision-makers will believe that "previous winners will win again". In this sense the two phenomena deliver behavior which is qualitatively similar. However, the mechanism for the belief in the framework we propose is in fact different. In the case of the traditional Hot Hand Fallacy, the Hot Hand belief arises when a sequence of positive outcomes convinces the decision-maker that a particular data generator (ex. a basketball player or a lottery store) is having a successful streak. The key driver of the Lucky Store Effect on the other hand, is the Bayesian updating feature of the model; however, in order for the Bayesian updating process to generate such a result when facing a reality of serially independent probabilities, a psychologically plausible structure for decision-makers' belief biases is needed. The Law of Small Numbers framework (Rabin, 2002; Tversky and Kahneman, 1981) provides exactly this structure. By allowing for concurrent comparisons, the Rabin (2002) model can indeed support a Lucky Store Effect, and the conditions which can support the effect are intuitive; the psychological 'pot size' must be large enough. This implies that *either* the odds must be miniscule enough (as in our lottery setting) for the pot to have a large lower bound on its size in a finite urn, *or* that decision-makers are not overly biased. In addition, holding these factors fixed, the variation in plausible success rates of stores in the decision-maker's beliefs, should be sufficiently large.

We can see several possible directions for future research. First, in addressing the Lucky Store effect theoretically and empirically, our study has focused primarily on the immediate effect of prior wins. Future work could add to the relatively sparse existing literature on the time dimensional Hot Hand Fallacy, as well as the possible interactions between the Hot Hand Fallacy and Lucky Store Effect. Our empirical analysis has also mostly focused on aggregate measures in the marketplace, rather than analysis at the individual level; future work could provide a richer picture of individual heterogeneity in belief biases in the collective lottery setting. Finally, although our empirical analysis focuses on the online lottery market, future work should explore the Lucky Store Effect and related belief fallacies in other contexts.

References:

Asparouhova, Elena, Michael Hertzel, and Michael Lemmon, "Inference from Streaks in Random Outcomes: Experimental Evidence on Beliefs in Regine Shifting and the Law of Small Numbers",*Management Science,* Vol. 55, p. 1766 – 1782.

Camerer, Colin, F., "Does the Basketball Market Believe in the 'Hot Hand'?" *American Economic Review,* Vol. 79 (1989), No. 5, p. 1257 – 1261.

Clotfelter, Charles, T., and Philip J. Cook, "The 'Gambler's Fallacy' in Lottery Play", *Management Science,* Vol 39 (1993), p. 1521 – 1525.

Croson, Rachel and James Sundali, "The Gambler's Fallacy and the Hot Hand: Empirical Data from Casinos", *The Journal of Risk and Uncertainty,* 30:3, p. 195 – 209, 2005.

Galbo-Jorgensen, Claus B., Sigrid Suetens, and Jean-Robert Tyran, "Predicting Lotto Numbers: A Natural Experiment on the Gambler's Fallacy and the Hot-Hand Fallacy", *Journal of the European Economic Association*, forthcoming (2015).

Gilovich, T., Vallone, R., and Amos Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences", *Cognitive Psychology,* Vol. 17 (1985), p. 295 – 314.

Guryan, Jonathan and Melissa S. Kearney, "Gambling at Lucky Stores: Empirical Evidence from State Lottery Sales," *The American Economic Review*, Vol. 98, No. 1 (March 2008), p. 458 – 473.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhouser, "Hot Hands in Mutual Funds: Short-Run Persistence of Performance", *Journal of Finance,* Vol. 64, p. 2289 – 2326.

Huber, Jurgen, Michael Kirchler, and Thomas Stockl, "The Hot Hand Belief and the Gambler's Fallacy in Investment Decisions under Risk", *Theory and Decision,* Vol. 68, p. 445 – 462.

Humphreys, Brad R., and Levi Perez, "Syndicated play in lottery games", *Journal of Socio-Economics,* Vol. 45 (2013), p. 124 – 131.

Jagannathan, Ravi, Alexey Malakhov, and Dimitry Novikov, "Do Hot Hands Exist Among Hedge Fund Managers? An Empirical Evaluation." *Journal of Finance*, Vol. 65, p. 215 – 255.

Lien, Jaimie W. and Jia Yuan (a), "The cross-sectional 'Gambler's Fallacy': Set representativeness in lottery number choices", *Journal of Economic Behavior and Organization*, Vol. 109 (2015), p. 163 – 172.

Lien, Jaimie W. and Jia Yuan (b), "Selling to Biased Believers: Strategies of Online Lottery Ticket Vendors", *Economic Inquiry,* Vol. 53 (2015), p. 1506 – 1521.

Miller, Joshua Benjamin and Adam Sanjurjo (a), "Is it a Fallacy to Believe in the Hot Hand in the NBA Three-Point Contest", *Working paper*, June 2015.

Miller, Joshua Benjamin and Adam Sanjurjo (b), "Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers", *Working paper,* September 2015.

Powdthavee, Nattvudh and Yohannes Riyanto, "Would You Pay for Transparently Useless Advice? A Test of Boundary Beliefs in the Folly of Predictions", *Review of Economics and Statistics*, Vol. 97 (2015), No. 2, p. 257 – 272.

Rabin, Matthew, "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics*, Vol. 117 (2002), p.775 - 816.

Rabin, Matthew and Dimitri Vayanos, "The Gambler's and Hot-Hand Fallacies: Theory and Applications," *The Review of Economic Studies*, Vol.77 (2010), p. 730 – 778.

Rao, Justin M., “When the Gambler’s Fallacy becomes the Hot Hand Fallacy: An Experiment with Experts and Novices,” *Working Paper*, October 2009.

Terrell, Dek, “A Test of the Gambler’s Fallacy: Evidence from Pari-mutuel Games”, *Journal of Risk and Uncertainty*, Vol. 8 (1994), p. 309 – 317.

Tversky, Amos, and Daniel Kahneman, “Belief in the Law of Small Numbers”, *Psychological Bulletin*, Vol. 76 (1971), No. 3, p. 105 – 110.

Yuan, Jia, Guang-Zhen Sun, and Ricardo Siu, “The Lure of Illusory Luck: How Much Are People Willing to Pay for Random Shocks?” *Journal of Economic Behavior and Organization*, Vol. 106 (2014), p. 269 – 280.

Appendix A:
The Lucky Store Effect: n *stores,* K *winning rates, and no restriction on prior distributions*

In this section, we show that the Lucky Store effect holds for the more general case of n stores, K winning rates, and no restriction on prior distributions. The possible outcomes for each store in a given lottery round are winning, a, and not winning, b. Again, we focus on the case where just one round of lottery outcomes has been observed, $m_i = m = 1$, since this is the scenario under which the Lucky Store effect could occur.

Returning to our original setup where stores are indexed by $i = 1,...,n$ and the possible winning rates are given by $\theta_i \in \{r_1,...,r_K\}$, for store i we have

$$P(\theta_i = r_k \mid y_{i,1} = a) = \frac{\pi_i(r_k)\cdot r_k}{\sum_{k'=1}^{K}\pi_i(r_{k'})\cdot r_{k'}} \qquad \forall k = 1,...,K$$

$$P(\theta_i = r_k \mid y_{i,1} = b) = \frac{\pi_i(r_k)\cdot(1-r_k)}{\sum_{k'=1}^{K}\pi_i(r_{k'})\cdot(1-r_{k'})} \qquad \forall k = 1,...,K$$

Recall that each store has its own pot, so when considering the probability of store i winning in the next lottery round, only store i's previous outcome matters. We have for the case where store i won,

$$\begin{aligned}
P(y_{i,2} = a \mid y_{i,1} = a) &= \sum_{k=1}^{K} P(y_{i,2} = a \mid \theta_i = r_k, y_{i,1} = a)\cdot P(\theta_i = r_k \mid y_{i,1} = a) \\
&= \sum_{k=1}^{K}\left(\frac{\pi_i(r_k)\cdot r_k}{\sum_{k'=1}^{K}\pi_i(r_{k'})\cdot r_{k'}}\cdot\frac{r_k M - 1}{M-1}\right) \\
&= \frac{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k^2\right)}{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}\cdot\frac{M}{(M-1)} - \frac{1}{(M-1)}.
\end{aligned}$$

For the case where store i did not win,

$$\begin{aligned}
P(y_{i,2} = a \mid y_{i,1} = b) &= \sum_{k=1}^{K} P(y_{i,2} = a \mid \theta_i = r_k, y_{i,1} = b)\cdot P(\theta_i = r_k \mid y_{i,1} = b) \\
&= \sum_{k=1}^{K}\left(\frac{\pi_i(r_k)\cdot(1-r_k)}{\sum_{k'=1}^{K}\pi_i(r_{k'})\cdot(1-r_{k'})}\cdot\frac{r_k M - 1}{M-1}\right) \\
&= \frac{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\cdot r_k\right)}{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)}\cdot\frac{M}{(M-1)}.
\end{aligned}$$

Since we assume that all stores are ex ante the same, we have $\forall i, j = 1,...,n$,

$P(y_{i,2}=a \mid y_{i,1}=b)=P(y_{j,2}=a \mid y_{j,1}=b)$. As in the two-store case, we compare the attractiveness of store i which won in the previous round to the attractiveness of store j which did not win in the previous round.

$$
\begin{aligned}
&P(y_{i,2}=a \mid y_{i,1}=a)-P(y_{j,2}=a \mid y_{j,1}=b)\\
&=\frac{\left[\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k^2\right)-\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\cdot r_k\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)\right]}{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}\cdot\frac{M}{(M-1)}-\frac{1}{(M-1)}\\
&=\frac{\left(\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot r_j^2(1-r_k)-\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot r_j r_k(1-r_k)\right)}{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}\cdot\frac{M}{(M-1)}-\frac{1}{(M-1)}\\
&=\frac{\left(\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot(r_j^2-r_j r_k)\right)}{\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}\cdot\frac{M}{(M-1)}-\frac{1}{(M-1)}\\
&=\frac{\left(\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot(r_j-r_k)^2\right)}{2\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}\cdot\frac{M}{(M-1)}-\frac{1}{(M-1)}.
\end{aligned}
$$

In order to obtain the Lucky Store Effect, the above expression must be positive. Solving for M yields the following inequality:

$$
M>\frac{2\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}{\left(\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot(r_j-r_k)^2\right)}
$$

When the above inequality is reversed, we obtain a belief in reversals of outcomes.

Denoting the cutoff value of M by M^* (that is $M^*=\frac{2\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot(1-r_k)\right)\left(\sum_{k=1}^{K}\pi_i(r_k)\cdot r_k\right)}{\left(\sum_{j=1}^{K}\sum_{k=1}^{K}\pi_i(r_j)\pi_i(r_k)\cdot(r_j-r_k)^2\right)}$), we immediately have Proposition 1, as shown below.

Proposition 1: $\forall i, j=1,\cdots,n$, when $M>M^*$, $P(y_{i,2}=a \mid y_{i,1}=a)>P(y_{j,2}=a \mid y_{j,1}=b)$; when $M<M^*$, $P(y_{i,2}=a \mid y_{i,1}=a)<P(y_{j,2}=a \mid y_{j,1}=b)$.

Appendix B: Empirical Robustness Checks

Table B1: Tobit Regression: Dependent Variable: Sales Progress, Hours_left < 20 (Late Submissions)

	(1)	(2)	(3)	(4)	(5)	(6)
Win Rate	0.415***	0.150***	0.151***	0.150***	0.150***	0.150***
	(0.0171)	(0.0102)	(0.0100)	(0.00999)	(0.00999)	(0.00999)
Similarity Index	-0.817***	-0.216***	-0.193***	-0.189***	-0.189***	-0.188***
	(0.0741)	(0.0512)	(0.0506)	(0.0505)	(0.0505)	(0.0505)
Self-Investment		3.396***	3.518***	3.501***	3.502***	3.500***
		(0.0210)	(0.0217)	(0.0218)	(0.0218)	(0.0218)
Commission			-0.0297***	-0.0299***	-0.0299***	-0.0298***
			(0.000996)	(0.000996)	(0.000996)	(0.000996)
Size				-0.00000287***	-0.00000336***	-0.00000303***
				(0.000000524)	(0.000000697)	(0.000000710)
Shares					0.000000232	0.000000158
					(0.000000222)	(0.000000224)
Price						-0.0000748*
						(0.0000298)
Constant	2.041***	-0.0905***	0.0135	0.0230	0.0227	0.0231
	(0.0278)	(0.0193)	(0.0194)	(0.0194)	(0.0194)	(0.0194)
sigma						
_cons	1.387***	0.818***	0.808***	0.807***	0.807***	0.807***
	(0.00852)	(0.00476)	(0.00469)	(0.00469)	(0.00469)	(0.00469)
N	93054	93054	93054	93054	93054	93054

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B2: Tobit Regression: Dependent Variable: Sales Progress, Commission < 0.05 (Low Commission)

	(1)	(2)	(3)	(4)	(5)	(6)
Win Rate	0.486***	0.203***	0.200***	0.197***	0.197***	0.196***
	(0.0145)	(0.00922)	(0.00915)	(0.00912)	(0.00912)	(0.00911)
Similarity Index	-1.015***	-0.225***	-0.225***	-0.215***	-0.215***	-0.215***
	(0.0642)	(0.0439)	(0.0435)	(0.0435)	(0.0435)	(0.0434)
Self-Investment		3.460***	3.410***	3.378***	3.379***	3.375***
		(0.0192)	(0.0189)	(0.0190)	(0.0190)	(0.0190)
Commission			-0.0897***	-0.0892***	-0.0894***	-0.0901***
			(0.00327)	(0.00326)	(0.00326)	(0.00326)
Size				-0.00000402***	-0.00000472***	-0.00000422***
				(0.000000373)	(0.000000480)	(0.000000486)
Shares					0.000000311*	0.000000216
					(0.000000132)	(0.000000133)
Price						-0.000124***
						(0.0000193)
Constant	1.535***	0.0570***	0.124***	0.135***	0.135***	0.136***
	(0.0202)	(0.0139)	(0.0140)	(0.0140)	(0.0140)	(0.0140)
sigma						
_cons	1.205***	0.725***	0.719***	0.717***	0.717***	0.717***
	(0.00633)	(0.00362)	(0.00358)	(0.00357)	(0.00357)	(0.00357)
N	89071	89071	89071	89071	89071	89071

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B3: Tobit Regression: Dependent Variable: Sales Progress, SharePrice < 1 (Lower Share Price)

	(1)	(2)	(3)	(4)	(5)	(6)
Win Rate	0.571***	0.276***	0.268***	0.267***	0.267***	0.265***
	(0.0160)	(0.0106)	(0.0104)	(0.0104)	(0.0104)	(0.0103)
Similarity Index	-0.746***	-0.0990*	-0.0722	-0.0712	-0.0710	-0.0920
	(0.0668)	(0.0498)	(0.0488)	(0.0488)	(0.0488)	(0.0486)
Self-Investment		3.831***	4.011***	4.007***	4.007***	3.976***
		(0.0208)	(0.0215)	(0.0215)	(0.0215)	(0.0214)
Commission			-0.0415***	-0.0415***	-0.0415***	-0.0407***
			(0.000899)	(0.000899)	(0.000899)	(0.000895)
Size				-0.00000339**	-0.00000107	0.00000218
				(0.00000117)	(0.00000357)	(0.00000355)
Shares					-0.000000288	-0.000000765
					(0.000000419)	(0.000000417)
Price						-0.566***
						(0.0235)
Constant	2.260***	-0.0120	0.120***	0.122***	0.123***	0.307***
	(0.0217)	(0.0159)	(0.0159)	(0.0159)	(0.0159)	(0.0175)
sigma						
_cons	1.477***	0.930***	0.911***	0.911***	0.911***	0.906***
	(0.00759)	(0.00454)	(0.00445)	(0.00445)	(0.00445)	(0.00442)
N	166221	166221	166221	166221	166221	166221

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B4: Tobit Regression: Dependent Variable: Sales Progress, Size < 200 (Small Packages)

	(1)	(2)	(3)	(4)	(5)	(6)
Win Rate	0.632***	0.319***	0.311***	0.313***	0.312***	0.313***
	(0.0149)	(0.0103)	(0.0100)	(0.00996)	(0.00995)	(0.00995)
Similarity Index	-0.780***	-0.197***	-0.168***	-0.197***	-0.197***	-0.203***
	(0.0577)	(0.0442)	(0.0434)	(0.0431)	(0.0431)	(0.0431)
Self-Investment		3.877***	4.028***	3.925***	3.925***	3.917***
		(0.0191)	(0.0195)	(0.0194)	(0.0194)	(0.0194)
Commission			-0.0422***	-0.0404***	-0.0405***	-0.0399***
			(0.000801)	(0.000799)	(0.000799)	(0.000799)
Size				-0.00425***	-0.00491***	-0.00585***
				(0.000163)	(0.000204)	(0.000217)
Shares					0.000209***	0.000375***
					(0.0000390)	(0.0000414)
Price						0.0293***
						(0.00254)
Constant	2.259***	0.00165	0.151***	0.254***	0.249***	0.231***
	(0.0191)	(0.0145)	(0.0145)	(0.0149)	(0.0150)	(0.0150)
sigma						
_cons	1.474***	0.965***	0.945***	0.940***	0.940***	0.939***
	(0.00666)	(0.00416)	(0.00407)	(0.00405)	(0.00404)	(0.00404)
N	213789	213789	213789	213789	213789	213789

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B5: Tobit Regression: Dependent Variable: Sales Progress, Win Rate < 0.1 (Lower Return Rate Sellers)

	(1)	(2)	(3)	(4)	(5)	(6)
Win Rate	12.14***	0.0781***	7.139***	7.088***	7.085***	7.080***
	(0.405)	(0.00558)	(0.279)	(0.278)	(0.278)	(0.278)
Similarity Index	-1.149***	-0.0294	-0.298***	-0.290***	-0.290***	-0.289***
	(0.0466)	(0.0310)	(0.0308)	(0.0308)	(0.0308)	(0.0308)
Self-Investment		1.042***	3.433***	3.413***	3.413***	3.412***
		(0.0996)	(0.0131)	(0.0131)	(0.0131)	(0.0131)
Commission			-0.0329***	-0.0331***	-0.0331***	-0.0331***
			(0.000585)	(0.000585)	(0.000585)	(0.000585)
Size				-0.00000410***	-0.00000484***	-0.00000443***
				(0.000000355)	(0.000000459)	(0.000000467)
Shares					0.000000332*	0.000000254
					(0.000000129)	(0.000000131)
Price						-0.0000755***
						(0.0000162)
Constant	1.717***	0.0897***	0.0331**	0.0432***	0.0428***	0.0434***
	(0.0149)	(0.0103)	(0.0102)	(0.0102)	(0.0102)	(0.0102)
sigma						
_cons	1.271***	0.352***	0.725***	0.724***	0.724***	0.724***
	(0.00484)	(0.00200)	(0.00261)	(0.00260)	(0.00260)	(0.00260)
N	188919	19105	188919	188919	188919	188919

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B6: Tobit Regression: Dependent Variable: Sales Progress, $\textit{Similarity Indicator}$ ($S\_Index > 0.2$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Win Rate	0.489***		0.495***	0.210***	0.210***	0.208***	0.208***	0.207***
	(0.0103)		(0.0103)	(0.00642)	(0.00630)	(0.00628)	(0.00628)	(0.00628)
Similarity Indicator		-0.0366**	-0.0970***	-0.0315***	-0.0350***	-0.0345***	-0.0347***	-0.0350***
		(0.0131)	(0.0131)	(0.00901)	(0.00887)	(0.00885)	(0.00885)	(0.00885)
Self-Investment				3.473***	3.608***	3.583***	3.583***	3.582***
				(0.0131)	(0.0134)	(0.0135)	(0.0135)	(0.0135)
Commission					-0.0346***	-0.0349***	-0.0349***	-0.0349***
					(0.000587)	(0.000587)	(0.000587)	(0.000587)
Size						-0.00000531***	-0.00000611***	-0.00000564***
						(0.000000379)	(0.000000490)	(0.000000499)
Shares							0.000000358*	0.000000267
							(0.000000139)	(0.000000140)
Price								-0.0000881***
								(0.0000175)
Constant	1.771***	1.841***	1.779***	-0.0656***	0.0583***	0.0715***	0.0711***	0.0718***
	(0.0138)	(0.0139)	(0.0138)	(0.00962)	(0.00967)	(0.00970)	(0.00970)	(0.00970)
N	238317	238317	238317	238317	238317	238317	238317	238317

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B7: Tobit Regression: Dependent Variable: Sales Progress, $\boldsymbol{Similarity\ Indicator}(S\_Index > 0.5)$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Win Rate	0.489***		0.493***	0.210***	0.209***	0.207***	0.207***	0.207***
	(0.0103)		(0.0103)	(0.00642)	(0.00629)	(0.00627)	(0.00627)	(0.00627)
Similarity Indicator		-0.151***	-0.264***	-0.0894***	-0.0691**	-0.0704**	-0.0702**	-0.0704**
		(0.0354)	(0.0357)	(0.0248)	(0.0244)	(0.0243)	(0.0243)	(0.0243)
Self-Investment				3.473***	3.608***	3.583***	3.583***	3.582***
				(0.0131)	(0.0134)	(0.0135)	(0.0135)	(0.0135)
Commission					-0.0346***	-0.0348***	-0.0348***	-0.0348***
					(0.000587)	(0.000587)	(0.000587)	(0.000587)
Size						-0.00000532***	-0.00000611***	-0.00000564***
						(0.000000379)	(0.000000490)	(0.000000499)
Shares							0.000000352*	0.000000262
							(0.000000139)	(0.000000140)
Price								-0.0000877***
								(0.0000175)
Constant	1.771***	1.840***	1.774***	-0.0671***	0.0562***	0.0694***	0.0690***	0.0697***
	(0.0138)	(0.0139)	(0.0138)	(0.00960)	(0.00965)	(0.00967)	(0.00967)	(0.00967)
N	238317	238317	238317	238317	238317	238317	238317	238317

Standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$