Predicting Lotto Numbers:
A natural experiment on the gambler's fallacy and the hot-hand fallacy

Abstract

We investigate the 'law of small numbers' using a data set on lotto gambling that allows us to measure players' reactions to draws. While most players pick the same set of numbers week after week, we find that those who do change react on average as predicted by the law of small numbers as formalized in recent behavioral theory. In particular, players tend to bet less on numbers that were drawn in the last week, as suggested by the 'gambler's fallacy', and bet more on a number if it was frequently drawn in the recent past, consistent with the 'hot-hand fallacy'.

Keywords: behavioral economics, decision biases, gambler's fallacy, hot-hand fallacy, lotto gambling

1 Introduction

Drawing the right inference from observing noisy data is difficult. People even tend to see patterns in data when, in fact, there are absolutely none. This paper focuses on two — apparently contradictory — fallacies in inference from noisy data that have been found to be pervasive in the literature (Tversky and Kahneman, 1971). One is the 'gambler's fallacy' (GF) according to which people tend to believe in frequent reversals, i.e. that a particular event is less likely to occur because it occurred recently. In state lotteries, for example, after a number has been drawn, the amount bet on that number falls sharply (Clotfelder and Cook, 1993; Terrell, 1994), and roulette players expect that a black number is 'due' after observing a sequence of red numbers (Croson and Sundali, 2005).[1] The other is the 'hot-hand fallacy' (HHF) according to which people believe in the continuation of a trend, i.e. that a particular event is more likely to occur because a recent streak of such events occurred. A much-cited case in point is that bets on basketball players who scored unusually well in the recent past tend to be too high (Gilovich et al., 1985; Camerer, 1989). And Guryan and Kearney (2008) provide evidence for a 'lucky store effect', the observation that lotto vendors sell many more tickets if they have sold a ticket that won a large prize a week earlier.[2]

A theoretical model by Rabin and Vayanos (2010) (henceforth RV) shows that the two apparently contradictory fallacies can be reconciled with reference to the 'law of small numbers' (Tversky and Kahneman, 1971).[3,4] According to this 'law', people hold the fallacious belief that small samples should be representative of the population and should therefore 'look like' large samples. RV show that a person who falls prey to the GF in the short run is also prone to develop the HHF in the long run. We provide supportive evidence for such

[1]Frequent reversals have also been observed in various laboratory experiments where participants are asked to generate a random sequence as in a coin toss (e.g., Bar-Hillel and Wagenaar, 1991; Rapoport and Budescu, 1997).

[2]False hot-hand beliefs also arise in laboratory experiments with an ambiguous data-generating process (Offerman and Sonnemans, 2004) or among participants who let 'experts' bet for them on outcomes of random draws (Huber et al., 2010). Powdthavee and Riyanto (2012) show that lab participants are even willing to pay for the (transparently useless) advice of 'experts'.

[3]See also Rabin (2002), who assumes that individuals draw balls from an urn with replacement, but think that replacement occurs only every odd period.

[4]In fact, the Marquis de Laplace (1812) had already suggested that both fallacies (he calls them illusions) are likely to occur in lotto gambling: "When a number in the lottery of France has not been drawn for a long time the crowd is eager to cover it with stakes. They judge since the number has not been drawn for a long time that it ought at the next drawing to be drawn in preference to others. So common an error appears to me to rest upon an illusion (...). By an illusion contrary to the preceding ones one seeks in the past drawings of the lottery of France the numbers most often drawn, in order to form combinations upon which one thinks to place the stake to advantage. But (...) the past ought to have no influence upon the future." Cited after Truscott and Emory, 1902: 161f.

'fallacy reversal' using data from a state lottery.

The intuition behind fallacy reversal is as follows.[5] A person who believes that small samples should 'look like' large samples expects frequent reversals in short random sequences, and thus falls prey to the GF in the short run. If such a person is uncertain about the true probability underlying a sequence of events, she starts to doubt about the true probability when observing a long streak because this observation does not correspond to what she believes a random sequence should look like. As a consequence, this person revises her estimate of the true probability, starts to believe in the continuation of the streak, and thus develops the HHF. Uncertainty about the true probability is key for a fallacy reversal to occur according to this theory, and it also seems to be key in practice. Asparouhova et al. (2009) ask participants in a lab experiment to predict the next observation in a random-walk process. The authors show that the HHF becomes more prevalent (compared to the GF) as subjects perceive the data-generating process to be less random.[6]

The law of small numbers as modeled by RV not only reconciles gambling behavior and betting in sports markets, but may also provide a unifying framework to account for several anomalies observed in financial markets. One of these anomalies is that asset prices underreact to news in the short run and overreact over long time horizons.[7] In a nutshell, the reasoning is as follows. If many investors are prone to the GF, and believe that short sequences of unexpectedly high earnings will quickly reverse in the future, stock prices will

[5]See Appendix A.1 for an interpretation of the model by RV in terms of lotto play.

[6]Lab experiments in psychology show that the GF is mostly observed when events are believed to be totally random while the HHF arises when events are perceived to be at least partly driven by a systematic factor, involving, for example, human skill (see Oskarsson et al., 2009, for a review). Guryan and Kearney (2008) provide an explanation along these lines for their 'lucky store effect'.

[7]De Bondt and Thaler (1985, 1987) and Jegadeesh and Titman (1993) are seminal papers documenting long-run overreaction and short-run momentum, respectively. More recent evidence for underreaction to earnings announcements in the short run comes from Cohen and Frazzini (2008), DellaVigna and Pollet (2009), and Hirshleifer et al. (2009).

underreact to news about earnings. The same investors, if uncertain about the process behind earnings sequences, attribute long streaks of unexpectedly high earnings to an underlying fundamental and expect such streaks to continue, leading to overreaction of stock prices in the long run.[8] Among investors who are confident that earnings are iid, underreaction persists and may even become stronger, the longer the streak.[9]

We study the GF and the HHF among lotto players by inferring players' beliefs about winning numbers from the numbers they choose on their lotto tickets, and by relating these choices to recent histories of lotto draws. Using this lotto data set has a number of advantages over other data sources as is explained next.

Identifying such biases and how they may relate using financial or sports market data is fraught with difficulties. A particularly important problem is that the data-generating process in these markets is not known to the researcher but has to be somehow estimated. They do not provide a clean test environment because these data are not only driven by randomness but also by ability, skill, or other systematic factors that are often difficult to measure. For example, long streaks of unexpectedly high earnings in a firm can be due to chance but they can also have a more fundamental cause, so it is not clear whether investing in stocks with a long history of high earnings should be seen as a hot-hand *fallacy* or as a hot-hand *reality*.[10] In contrast, lotto gambling is an ideal test environment to study these fallacies and how they relate. It provides a natural experiment because lotto draws follow a known,

[8]These aggregate patterns may additionally be driven by biases in the processing of private information among investors (e.g. Daniel et al., 1998; Hong and Stein, 1999). Such biases are unlikely in lotto gambling given that all information about outcomes of lotto draws is public.

[9]For example, Loh and Warachka (2012) show that stock returns underreact to 'news' (earnings surprises) in the short and in the long run. They suggest that overreaction to streaks of high earnings is absent because investors in their sample are confident that the long-term distribution of earnings surprises is iid.

[10]For example, persistence has been observed in the performance of mutual funds (Hendricks et al., 1993; Carhart, 1997) and superior hedge funds (Jagannathan et al., 2010). And in sports, a belief in a hot hand is not necessarily fallacious but may be 'real'. See Bar-Eli et al. (2006) for an overview and Yaari and David (2012) and Miller and Sanjurjo (2014) for recent studies.

truly random process (with a fixed probability for each number), and are thus independent and truly exogenous. In fact, the true randomness of the game is tightly controlled (often by government regulation) and made transparent to players (e.g., by drawing balls from an urn and by airing the draws on TV).

A key novelty of our data is that we can track players' choices (the numbers they bet on) over time. The data come from lotto played over the Internet in Denmark where the law requires gamblers to be uniquely identified. This identification allows us to measure how individual players react with their number choices to recent draws and provides three advantages. First, in contrast to aggregate lottery data (see Clotfelder and Cook, 1993; Terrell, 1994), our data allows us to identify whether changes in the aggregate amount bet on recently drawn numbers are caused by the GF, or by players who stop playing after they have won. Second, our data allows us to test whether players prone to the GF are also prone to the HHF as numbers get 'hotter'. Such a test was not possible with data used in the previous lottery studies because no streaks of 'hot' numbers were observed.[11] Third, the main advantage as compared to Sundali and Croson (2006), who use data from videotaped roulette players, is that players in our data set have a unique ID and can thus be carefully identified. In addition, our dataset has many more players, and not only gamblers in casinos.

We find evidence of statistically significant fallacy reversal. Consistent with the GF, players place on average 1.6% fewer bets on numbers drawn in the previous week compared to numbers not drawn, as long as those numbers are not 'hot'. Consistent with the HHF, players bet on average more money on numbers, as they get 'hotter', i.e. as they have won more often in the recent past. In particular, players bet about 1% more on numbers drawn in the previous week for each additional week they have been drawn before. While many players choose the same numbers week after week, we find that these results are driven by

[11] Both Clotfelder and Cook (1993), and Terrell (1994) study three-digit numbers (ranging from 000 to 999) which means that the occurrence of (even short) streaks is highly unlikely. For example, the chance of observing a streak of length two is 1/1'000'000.

frequent players who change their number choices, rather than by players who stop betting after their numbers have won. To illustrate, if we focus on the sample of players who change their number choices at least once, we find that the effects are larger in size: they place 2% fewer bets on numbers drawn in the previous week than on numbers not drawn, and they bet about 1.4% more on numbers drawn in the previous week for each unit increase in 'hotness'.

In an individual-level analysis we show that the two fallacies are systematically related as predicted by Rabin and Vayanos (2010). The same players who exhibit the GF, also tend to be prone to the HHF. To illustrate, out of those players who react significantly to the recent drawing history, 55% exhibit both fallacies. This number is more than twice the percentage one would expect if players choose numbers randomly (25%).

The data analyzed here has been used in Suetens and Tyran (2012) with a particular focus on gender issues, showing that male players are prone to the GF while female players are not. The current paper provides a much broader analysis and studies various issues that were not touched upon previously. In particular, we study both fallacies and how they relate here. To the best of our knowledge, this paper is the first to study fallacy reversal in the aggregate and at the individual level using lotto data.

The paper is organized as follows. Section 2 describes the lotto data and defines the main variables of interest. Section 3 analyzes the aggregate reaction of lotto players to the recent drawing history of lotto numbers. Section 4 uses individual-level data to investigate the relation between the two biases. Section 5 concludes.

2 The data

We analyze data from lotto played on Saturdays in Denmark over the Internet (henceforth lotto for short) covering 28 weeks in 2005. Lotto is organized by a state monopoly (Danske Spil). Every Saturday, 7 balls are drawn from an urn containing 36 balls numbered from 1 to 36, which is aired on state TV. Information about the winning numbers is also posted on the website of Danske Spil, that is, the same website where players buy tickets. However, we do not know whether the players have looked at this information.[12] The price of a lotto ticket is about EUR 0.40 (DKK 3).[13]

The payout rate is set to 45% by law and the remainder of the revenues is earmarked for 'good causes' or goes to the general government budget. Lotto has a parimutuel structure as the payout rates are fixed per prize category and the prize money per category is shared among the winners in that category. One quarter of all payouts are reserved for the jackpot (7 correct numbers), and there are four graded prizes for having selected fewer correct numbers. If no-one wins the jackpot, it is rolled-over to the next week. In our data set, the average jackpot was about EUR 534'000 (4 million DKK), and the highest jackpot, net of taxes, was 1.4 million EUR (10.2 million DKK).[14] Given that a lotto ticket costs DKK 3, and the chances to win are about 1: 8 million (see footnote 17), there was thus no week where the jackpot was even close to being large enough to make gambling profitable in expectation.

Lotto is normally played in Denmark by purchasing hard-copy lotto tickets in vending booths like drugstores and supermarkets. Since 2002, lotto can also be played over the Internet. Lotto numbers can be picked in various ways in Denmark. Traditionally, players manually select 7 out of 36 numbers on each ticket they buy. However, we analyze numbers picked in 'Systemlotto'. Here, players select between 8 and 31 numbers manually and let the lotto agency choose combinations of 7 out of these numbers.[15] Our data has been provided directly by the state lottery agency and is unlikely to contain any error. All players in our

[12]There are a number of other websites, unrelated to Danske Spil, that give information about the history of the number drawings (for example, which numbers are 'due', or which numbers are 'hot'. However, we do not know whether players have used these websites.

[13]The following describes the rules of lotto at the time the data was collected. The prize structure has been modified since to yield higher jackpots.

[14]Prizes above DKK 200 are subject to a special tax of 15% but are otherwise exempt from income tax.

[15]Other ways to play are 'Quicklotto' where all numbers are selected randomly by the lotto agency and 'Lucky-lotto' where players select up to 6 numbers manually and let the lotto agency choose the remaining numbers.

dataset are identified by a unique ID-number which allows the lottery agency to track the choices of players over time.[16] We also received information about the gender and age of (most) players. In total, 189'531 persons have played lotto over the Internet at least once in the second half of 2005. More than half of these (100'386) manually select their numbers in the traditional way, and 25'807 select numbers using Systemlotto.

We focus our analysis on Systemlotto rather than the traditional manual selection. The main reason is that in Systemlotto players choose *numbers*, rather than *combinations* of numbers, as in the traditional manual selection. They choose fewer unique numbers than players who select in the traditional way, which suggests that they are more likely to believe that a particular number is going to win. To illustrate, Systemlotto players pick less than half among the 36 available numbers (14 numbers in an average week, 8 in a modal week). In manual selection, players pick most available numbers (29 in an average week, 32 in a modal week). Also, the selection screens look different for traditional manual selection and Systemlotto. In Systemlotto the choice of particular numbers is emphasized because players simply enter between 8 and 31 numbers. With traditional manual selection they enter their choice of combinations of numbers on a grid which may give rise to particular visual patterns like crosses, see (Simon, 1999). Using Systemlotto thus avoids confound between choosing numbers randomly and non-randomly according to visual patterns versus the GF/HHF (??xx). Moreover, in our sample we observe particular *numbers* winning repeatedly but not particular *combinations* of numbers winning repeatedly.[17] Given that a key focus of our study is on how players react to a particular drawing outcome being more or less frequent in 'the recent

[16]We do not have information about the exact birth dates of the players so that, unfortunately, in our analyses we are not able to test whether players based their number selection on their birth date.

[17]While there are only 36 numbers, there are about 8 million ways to combine 7 out of 36 numbers ($36!/(7!(36-7)!) = 8'347'680$). The probability that the same combination occurs twice in a row is therefore about 1 in 70 trillion (7 x 10^{13}) in Danish lotto. Curiously enough, the same six numbers were drawn twice in a row in the Bulgarian lotto in September 2009. This event was considered to be so unlikely that it prompted the Bulgarian government to initiate an investigation for manipulation of the game.

past, it is natural to focus on data where we can believe that players choose numbers rather than combinations of numbers. (xx unclear xx)

An advantage of our data set compared to laboratory data is that it reflects behavior of a heterogeneous pool of people and behavior is observed in a 'natural' situation. In fact, lotto is quite popular in Denmark. For example, according to the lotto agency, about 75% of the adult Danish population have played lotto at least once. Yet, Systemlotto players are clearly not representative for the Danish population or even for the pool of internet lotto players. People playing Systemlotto buy on average about twice as many tickets as other internet players (29 vs. 14 tickets per week; the medians are 19 and 10, respectively). Systemlotto is also especially popular with male players: 82% of the players are male compared to 73% for other selection devices.

2.1 Dependent variables

Our empirical strategy is to make inferences about the (unobservable) belief in one's ability to predict winning lotto numbers more accurately than pure chance from observable reactions to previous drawings. That is, we infer that players think recently drawn numbers are more likely to win if they systematically prefer them and *vice versa* if they avoid them. A player is said to be more confident that a particular number is going to win if the player is more likely to pick the number, and/or if he or she places more bets on it (i.e. buys more tickets including this number).

We use two proxies for the confidence of a player that a particular number is going to win. The first measure is called NumberBet. It is very straightforward and simply indicates whether or not a player has bet in a particular week on a particular number. While the NumberBet does not take into account how much money is bet in total, the second measure for confidence, called MoneyBet does. This measure reflects how the total amount of money a player bets in a particular week is distributed over the lotto numbers. More specifically, the two measures (which serve as our independent variables below) are defined as follows.

Our first dependent variable measures whether a player picks a number in week t:

$$\text{NumberBet}_{ijt} = \begin{cases} 1 & \text{if} \quad \text{if player } i \text{ picks number } j \text{ in week } t, \\ 0 & \text{if} \quad \text{otherwise .} \end{cases} \tag{1}$$

Our second dependent variable measures how much money a player bets on a particular number relative to other numbers. Recall that in Systemlotto, players pick one or more sets containing 8 to 31 numbers. For each chosen set, the lotto agency generates at least 8 tickets with different combinations of subsets of 7 out of the chosen numbers.[18] The relative weight put on a number thus depends on how many numbers are chosen in total. Therefore, we define our second dependent variable as the total amount of money bet in t by player i multiplied by the relative weight put on the number by the player in t:

$$\text{MoneyBet}_{ijt} = \text{Total amount bet in DKK}_{it} \times \text{Weight}_{ijt}, \tag{2}$$

$$\text{with Weight}_{ijt} = \frac{\#\text{ of times number } j \text{ is picked in week } t \text{ by player } i}{\#\text{ of numbers picked in week } t \text{ by player } i}.$$

The following three examples illustrate the interpretation of the variable MoneyBet.

Example 1 *Player A chooses a set of 10 numbers and Player B chooses a set of 24 numbers. Both A and B buy 120 tickets (so each spend DKK 360) generated out of their chosen sets.*

Example 2 *Player C chooses a set of 10 numbers from 1 to 10 and a set of 10 numbers from 5 to 14. For each set, 8 tickets are generated. Player C thus buys 16 tickets in total (spends DKK 48).*

[18] The exact relation between the total number of tickets/combinations generated by Systemlotto out of a set of chosen numbers and the total number of lotto numbers a player chooses in the set depends on which of three 'systems' players use to generate tickets/combinations. See Appendix A.2 for details. We do not have information about the subsets of 7 numbers that eventually end up on the tickets.

Table 1: Summary statistics of dependent variables

	Full sample		**Active players**		**Changers**	
Dep. var.:	NumberBet	MoneyBet	NumberBet	MoneyBet	NumberBet	MoneyBet
Minimum	0	0	0	0	0	0
Maximum	1	1665.21	1	1665.21	1	1665.21
Mean	.15	1.10	.37	2.76	.42	3.54
Standard Deviation	.36	4.29	.48	6.45	.49	8.06
Median	0	0	0	0	0	0
75% Percentile	0	0	1	3.00	1	4.80
90% Percentile	1	3.00	1	9.00	1	11.70
95% Percentile	1	6.00	1	12.00	1	13.80
99% Percentile	1	15.23	1	27.60	1	36.00
Nr. of data points	26'013'456		9'148'464		4'419'612	
Nr. of players	25'807		17'318		8'224	

Notes: The table reports summary statistics of the dependent variables as defined in eqs. 1 and 2 for the three samples we use in our analysis. In the full sample, data points from players who do not participate in the lottery game in a particular week are coded as zero (for all numbers in that week). In the sample of active players, which is a subset of the full sample, these data points are coded as missing so are not part of the sample. The changers are a subset of the sample of active players. This sample is obtained by excluding players who always bet the same amount on the same numbers across the 28 weeks of study. The unit of observation is a player (i) who bets in a given week (t) on a given lotto number (j).

Example 3 *Player D chooses a set of 12 numbers from 1 to 12 in week 1, from which 12 tickets are generated (spends DKK 36). In week 2 player D chooses the same set of numbers, but now opts for 48 tickets (DKK 144).*

In example 1, both players buy the same number (120) of tickets. Yet, it is plausible to assume that player A is more confident that (some of) his 10 numbers are going to win than player B who picks 24 numbers. The variable MoneyBet gives each of the 10 numbers chosen by player A a larger weight (of 1/10) than each of the 24 numbers chosen by player B (1/24). In example 2, player C chooses two sets which partly overlap since the numbers 5 to 10 are elements of both sets. It seems plausible to assume that player C is more confident that one of the numbers contained in both sets (5 to 10) is going to win than one of the numbers contained in only one of the sets (1 to 4 and 11 to 14). The variable MoneyBet gives more weight to numbers occurring in overlapping sets than to numbers occurring in only one set. Finally, in example 3, where the amount bet by player D is multiplied by four, the variable MoneyBet is multiplied by four as well.

Table 1 provides summary statistics of the dependent variables for the three samples we use in our analysis. The first is what we call the 'full sample' which includes at any time t all (N = 25'807) subjects who play at least once over the entire observation period (28 weeks). The second is what we call the sample of 'active players'. This sample contains at any point in time only the players who are active in the sense that they bet in this particular period. Technically speaking, the difference between the two samples is that data points from players who do not participate in the lottery in a given period are coded as zeroes in the full sample while they are coded as missing (i.e. are not part of the sample) in the active sample. Thus, the active sample is a subsample of the full sample, and varies in size from period to period. One way to think about the two samples is that inclusion in the 'full sample' of player i in period t is unconditional on betting in t, while inclusion in the 'active' sample is conditional on betting in this period.

The third sample is a subsample of the active sample. It contains only players who change their bets (as defined in eq. 2) at least once over the entire observation period. Technically speaking, this 'sample of changers' is obtained from the sample of active players by excluding players who pick the same numbers week after week. For the samples of active players and of the changers (i.e. excluding players who do not participate in a particular week), including a lagged dependent variable in the regressions (see below) implies that only data points are taken into account where players play in at least two consecutive weeks. This has the additional advantage that these data come from players for whom we can be sure that they visit the Danske Spil website two weeks in a row, and hence presumably know whether one of their chosen numbers has won in the previous week.

To illustrate, in the full sample the mean amount bet on a number by a player per week is DKK 1.10. Since there are 36 lotto numbers in total, this corresponds to a total amount bet by player per week of DKK 39.68 (about EUR 5.3). The mean amount bet in the active sample is more than twice as large, namely DKK 2.76, corresponding to a total of DKK 99.44 (about EUR 13.25). Finally, 'changers' bet even more with an average of DKK

3.54 per week per number, so DKK 127 (about EUR 17) in total per week. The table also provides information about the distribution of bets. In all samples, the meidan NumberBet is 0, meaning that fewer than half of the numbers are selected for each player and week. In about 5 percent of the cases, a player bets almost one Euro (DKK 6.00) on a given number in a given week for the full sample, and the corresponding numbers are about double that size in the sample of active players (DKK 12.00) and the sample of changers (DKK 13.80), respectively.

2.2 Main independent variables

Our aim is to study whether players systematically react to drawings in previous weeks and, if so, whether players who fall prey to the GF also develop the HHF. To test for the presence of GF, i.e. whether players bet less on numbers drawn in the previous week, we define the variable Drawn_{jt-1} as follows:

$$\text{Drawn}_{jt-1} = \begin{cases} 0 & \text{if number } j \text{ has not been drawn in week } t-1, \\ 1 & \text{if number } j \text{ has been drawn in week } t-1. \end{cases} \tag{3}$$

To study whether the HHF is present, we test whether players bet more on 'hot' numbers. More specifically, we test whether players bet more on numbers that have been drawn frequently in x weeks preceding week $t-1$. We use this measure of 'hotness' rather than a literal 'streak', i.e., the number of consecutive weeks a number has been drawn, because long streaks are rare in lotto by the nature of randomness. In our sample, the maximum length of weeks with consecutive draws of a particular number is 4, and such a streak occurs only once.[19] Clearly, a trade-off is involved in the choice of x. On the one hand, in order to obtain sufficient power, x can not be too small. On the other hand, in order to capture

[19]The expected probability that a streak of length k occurs is $29/36 * (7/36)^k$. In an earlier version of this paper, we use 'streaks' rather than the indicator of 'hotness' in our analysis. The results are qualitatively the same but statistically less powerful.

recent drawing history, x must not be too large. Therefore, we chose $x = 5$.[20] Thus, our second independent variable Hotness_{jt-1} measures how often number j has been drawn in weeks $t-2$ to $t-6$:

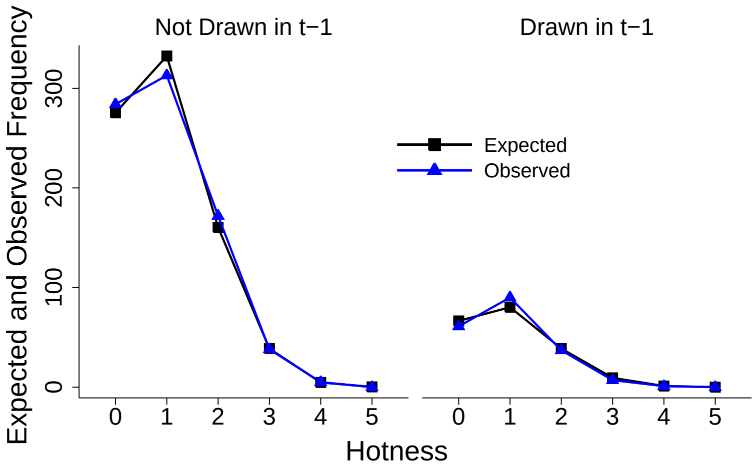
$$\text{Hotness}_{jt-1} = k \text{ if number } j \text{ has been drawn } k \text{ times in weeks } t-2 \text{ to } t-6. \quad (4)$$

Figure 1 illustrates the properties of the variable Hotness_{jt-1}. The figure shows expected and observed frequencies of hotness in our sample of numbers drawn (right panel) and numbers not drawn (left panel) in the previous week. One can see in the figure that the probability a number is drawn k times in 5 weeks quickly decays as k increases from $k = 1$ onwards. More crucially, given that it is much less likely that a number is drawn in week $t-1$ than that it is not drawn (7/36 vs. 29/36), the hotness of numbers drawn in week $t-1$ has a different interpretation than the hotness of numbers not drawn. For example, the probability that a number is drawn twice in weeks $t-2$ to $t-6$ and drawn in $t-1$ is just 3.8% compared to 15.9% if not drawn in $t-1$. So the 'hotness perception' of a number with hotness equal to 2 will be much stronger if it is drawn in week $t-1$ than if not drawn.

To test whether the lotto drawings are truly random, we ran a number of Chi-square tests that compare the observed and expected distributions of variables. Reassuringly, these tests show that drawn numbers in Danish lotto are random. For example, we cannot reject the null that the observed distribution of drawings over all 28 weeks is different from the uniform distribution ($\chi^2 = 22.57$; critical value of 53.20 with 35 degrees of freedom and $\alpha = .05$). We also test whether the hotness variable is distributed as expected. Figure 1 shows that observed hotness in our sample (see triangles) matches expected hotness (see squares) closely. Chi-square tests confirm this impression ($\chi^2 = 2.50$ and $\chi^2 = 2.37$ for hotness given Drawn = 1 and Drawn = 0, respectively; the critical value is 11.14 with 4 degrees of freedom and

[20] Our results do not hinge on this choice. We get very similar results with $x = 4 or 6$, see Appendix tables A1 and A2.

Figure 1: Expected and observed frequencies of hotness



Notes: The chart shows expected and observed frequencies of hotness for 196 number draws (7 lotto numbers times 28 weeks). The expected frequency a number is drawn k times in weeks $t-2$ to $t-6$ (i.e. in 5 weeks) is equal to $\frac{k!}{5!(5-k)!}\left(\frac{7}{36}\right)^{k}\left(\frac{29}{36}\right)^{5-k}$. The expected frequency a number is drawn k times in $t-2$ to $t-6$ and in week $t-1$ is equal to $\frac{k!}{5!(5-k)!}\left(\frac{7}{36}\right)^{k+1}\left(\frac{29}{36}\right)^{5-k}$. The expected frequency a number is drawn k times in $t-2$ to $t-6$ and not in week $t-1$ is equal to $\frac{k!}{5!(5-k)!}\left(\frac{7}{36}\right)^{k}\left(\frac{29}{36}\right)^{4-k}$.

$\alpha = .05$).

2.3 Control variables

Our regressions control for various factors that plausibly affect betting behavior. First, betting is known to go up in weeks when the jackpot is 'rolled over' from the previous week because this increases the expected payout. To illustrate, in an average rollover week, the total amount bet in our sample is about 23% higher than in a non-rollover week (DKK 1.16 mio. vs. DKK 0.95 mio., or about EUR 155'100 vs. EUR 126'200). We control for a rollover effect by including a binary variable indicating whether week t is a rollover week. Second, it is well-known that low numbers are generally more popular than high numbers, and this is also the case in our sample. For example, the lowest 5 numbers (1 to 5) are picked more than 30% more often than the highest 5 numbers (32 to 36).[21] We control for such effects by including fixed number effects.

Finally, to control for the fact that some (perhaps idiosyncratically 'lucky') numbers are more popular with particular players than others, we include a lagged dependent variable in the regressions. For example, consider a player who always chooses lotto number 22, say. Suppose number 22 happens to be one of the hot numbers, i.e. has been drawn frequently in the recent past. If we do not correct for the player's idiosyncratic preference, we would wrongly conclude that the player exhibits the HHF. Such cases are qualitatively important given that out of the 17'318 players who have at least two consecutive observations, 9'094 players do not change how they bet on numbers at all over the period of study, and given that those who change, typically keep the majority of numbers the same. In the regressions with NumberBet as the dependent variable, we additionally control for the total number of numbers chosen by a player in week t, and for the total number of tickets bought by a player in week t. These two variables are expected to be positively associated with NumberBet because it is more probable that a player picks a particular number in week t, the more numbers she picks and the larger the number of tickets she buys in t.

3 Aggregate data analysis

This section analyzes how players bet as a function of recent drawings by reporting the results from pooled regressions. We estimate the following regression models separately for the three samples explained in Table 1:

$$\text{DV}_{ijt} = \beta_0 + \beta_1 \text{Drawn}_{jt-1} + \beta_2 \text{Hotness}_{jt-1} + \Gamma \text{Controls}_{ijt-1} + \epsilon_{ijt}, \text{ and} \tag{5}$$

$$\text{DV}_{ijt} = \beta_0 + \beta_1 \text{Drawn}_{jt-1} + \beta_2 \text{Hotness}_{jt-1} + \beta_3 \text{Drawn}_{jt-1} \text{Hotness}_{jt-1} + \Gamma \text{Controls}_{ijt-1} + \epsilon_{ijt}, \tag{6}$$

[21] The popularity of the lotto numbers is shown in Appendix A.3.

$$\text{with } i = 1, ..., N; j = 1, ..., 36; t = 1, ..., T_i.$$

The dependent variable (DV) is either NumberBet$_{ijt}$, as defined in (1) or MoneyBet$_{ijt}$, as defined in (2). Drawn$_{jt-1}$ is defined as in eq. 3 and Hotness$_{jt-1}$ as in eq. 4. Note that the only difference between the two specifications is the inclusion of an interaction effect between Drawn and Hotness. This interaction effect is supposed to measure the differential effects of Hotness given that a number was drawn in the previous week or not (Cf. the discussion of figure 1).

Table 2 shows the estimation results for the three samples. We find robust evidence for the GF. With one exception to be discussed below, the effect of Drawn is negative across all specifications and samples. This finding shows that the players' tendency to avoid numbers drawn in the previous week is rather robust. We also find robust evidence for the HHF at the aggregate level across all samples. The effect of Hotness in specification (5) is positive, which suggests that players tend to bet more on numbers, the hotter they are. Also, the effect of Drawn x Hotness in specification (6) is always positive, which suggests that players bet more on numbers drawn in the previous week, the hotter they are.

The control variables all have the expected sign: players tend to pick the same numbers in subsequent weeks (see the coefficient on the lagged dep. var.), bets are higher in rollover weeks, and the probability of betting on a certain number is higher the more tickets one buys and the more numbers one chooses.

Let us look more closely at the results for the different samples, and let us first focus on the GF. A first observation, based on specification (5), is that for the full sample the negative effect of Drawn is either not (with NumberBet) or marginally (with MoneyBet) significant if the interaction with Hotness is not included in the regression, whereas for the active players and the changers it is highly significant. A second observation is that the negative effect of Drawn is significant in all cases if the interaction with Hotness is included, but the size of the effect is relatively larger for active players and changers compared to the full sample. These findings indicate that the GF is mainly driven by variation at the intensive margin (how

Table 2: Pooled regression results

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
		(a) Full sample		
Constant	.0103 (.0046)***	.0104 (.0046)***	.4355 (.0348)***	.4367 (.0348)***
Drawn	-.0001 (.0001)	-.0010 (.0002)***	-.0055 (.0029)*	-.0185 (.0039)***
Hotness	.0001 (.0000)***	-.0000 (.0000)	.0011 (.0006)**	-.0012 (.0007)*
Drawn x Hotness		.0010 (.0001)***		.0133 (.0020)***
Rollover	.0117 (.0010)***	.0117 (.0010)***	.2455 (.0077)***	.2458 (.0077)***
# Tickets	.0059 (.0002)***	.0059 (.0002)***		
# Numbers	.0006 (.0002)***	.0006 (.0002)***		
Lagged dep. var.	.6838 (.0148)***	.6838 (.0148)***	.6090 (.0266)***	.6090 (.0266)***
R^2	.639	.639	.368	.368
# Data points		25'084'404		
# Players		25'807		
		(b) Active players		
Constant	.0177 (.0040)***	.0614 (.0084)***	.4326 (.0362)***	.4341 (.0362)***
Drawn	-.0020 (.0003)***	-.0028 (.0004)***	-.0279 (.0071)***	-.0433 (.0092)***
Hotness	.0004 (.0001)***	.0003 (.0001)***	.0040 (.0013)***	.0012 (.0014)
Drawn x Hotness		.0008 (.0002)***		.0159 (.0038)***
Rollover	.0048 (.0003)***	.0048 (.0003)***	.1350 (.0111)***	.1353 (.0111)***
# Tickets	.0001 (.0000)***	.0001 (.0000)***		
# Numbers	.0011 (.0004)***	.0011 (.0004)***		
Lagged dep. var.	.9141 (.0026)***	.9141 (.0026) ***	.8497 (.0120)***	.8497 (.0120)***
R^2	.855	.855	.668	.668
# Data points		8'525'016		
# Players		17'318		
		(c) Sample of changers		
Constant	.0612 (.0084)***	.0776 (.0098)***	.8477 (.0706)***	.8510 (.0706)***
Drawn	-.0047 (.0007)***	-.0065 (.0009)***	-.0684 (.0178)***	-.1054 (.0231)***
Hotness	.0010 (.0002)***	.0006 (.0002)***	.0076 (.0032)***	.0009 (.0036)
Drawn x Hotness		.0019 (.0005)***		.0383 (.0093)***
Rollover	.0115 (.0009)***	.0115 (.0009)***	.3499 (.0278)***	.3507 (.0278)***
# Tickets	.0002 (.0000)***	0002 (.0000)**		
# Numbers	.0015 (.0007)***	.0015 (.0007)***		
Lagged dep. var.	.7974 (.0061)***	.7974 (.0061)***	.7636 (.0184)***	.7636 (.0184)***
R^2	.672	.672	.518	.518
# Data points		3'380'508		
# Players		8'224		

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 based on the full sample (panel (a)), the sample of active players (panel (b)), and the sample of changers (panel (c)). Estimated coefficients for the fixed number effects (jointly statistically significant) are not reported. Standard errors are robust to within-player dependency. The stars ***, ** or * indicate that the effect of the variable is statistically significant at the 1%, 5% or 10% level, respectively.

players bet) rather than by variation at the extensive margin (whether players bet at all and how much they bet in total). This is so because in the full sample we potentially observe both reactions while in the sample with active players only, we only observe the variation at the intensive margin by definition.[22]

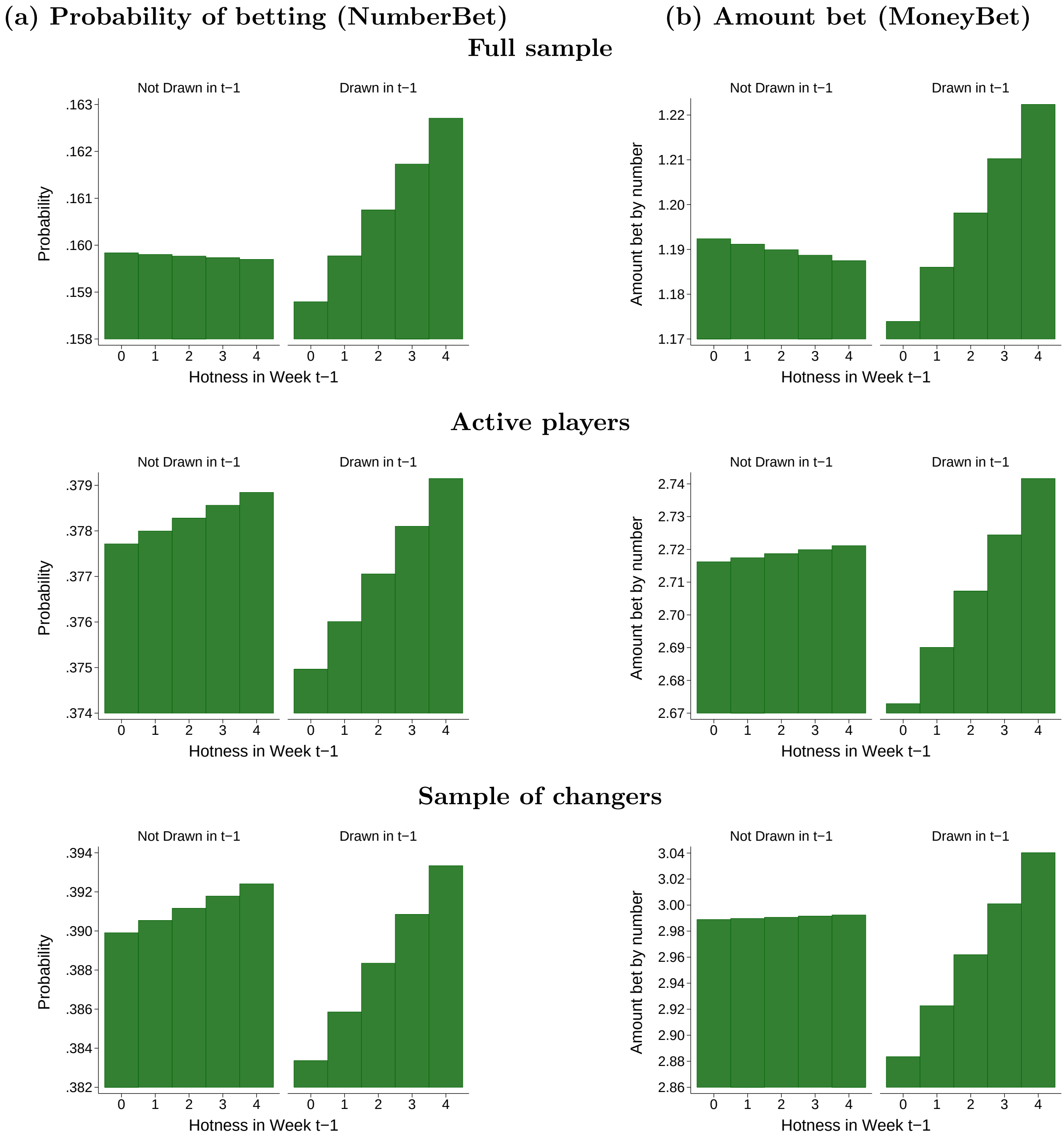
We now turn to the results on the HHF. Table 2 shows that the positive effect of Hotness in specification (5) is significant in all cases. Results based on specification (6) show that the positive effect of Hotness is almost entirely driven by Drawn x Hotness, meaning that players primarily tend to move to hot numbers drawn in the previous week. This tendency is much less pronounced (or even absent) if a previously hot number has not been drawn in the previous week. This is not surprising given that numbers *not* drawn in the previous week are in a sense 'less hot' than numbers with the same value of our Hotness variable but drawn in the previous week: that drawn numbers are hot is much less likely than than not drawn numbers are hot (cf. Figure 1 and discussion).

Figure 2 illustrates the estimated effects of the recent drawing history on our two indicators of betting. The panels in the left column (a) show the effect on NumberBet, i.e. the probability to pick a particular number, and the panels in the right column (b) show the effect on MoneyBet, i.e. how much money a typical player bets on a particular number, relative to other numbers. The figure is based on specification (6), see also Table 2.

The panels illustrate the GF and the HHF differentiated by whether a number has been drawn in the previous week (right part in each panel) or has not been drawn (left part) as follows. The GF-effect is illustrated by comparing two bars at a given level of Hotness within a panel. For example, comparing the two bars at Hotness = 0 in the upper left panel shows that the probability that a particular number is picked is lower when it has been drawn in the previous week than when it has not been drawn (just below .159 vs. just below .16, i.e.

[22]Recall that in the full sample, a data point of zero for the dependent variable may not only mean that the player does not bet on that particular number in that period (as in the other two samples) but also that the player does not bet at all in that period.

Figure 2: Estimated betting behavior as a function of the recent drawing history



Notes: The panels on the left show the estimated probability to pick a particular number in week t depending on whether the number was drawn in week $t-1$ or not, as specified in eq. 6 with NumberBet as dependent variable. The probabilities are calculated at the mean of the independent variables (apart from the fixed number effects). The panels on the right show the estimated amount bet (in DKK) on a number in week t depending on whether the number was drawn in week $t-1$ or not, as specified in equation 6 with MoneyBet as dependent variable. The estimated amounts are calculated at the mean of the independent variables (apart from the fixed number effects). The first row refers to estimations for the full sample, the second and third row to the sample of active players, and the sample of changers, respectively.

a drop of about 0.6%). The HHF-effect is illustrated by the sequence of bars within a part of a panel. For example, in the right part of the upper left panel, the probability to bet on a particular number increases from just below .159 when the number is not hot at all (Hotness = 0) to just below .163 (equivalent to an increase of 2.5%) when the number is very hot (Hotness = 4). These estimates reflect the case where a particular lotto number has been drawn in the previous week and we study the full sample. The middle and bottom rows show the corresponding cases for the sample of active players and the sample of changers, respectively. The right column shows the corresponding cases for the MoneyBet.

Inspection of the figure reveals that the GF effects are weakest in the full sample (upper panels) — which includes many instances where players do not bet at all — but are much stronger for the sample of active players and changers. For example, the GF effect for the simple indicator of betting (variable NumberBet) increases from a meager 0.6% in the full sample to a difference of 2% for the sample of changers. The increases for MoneyBet are similar (from about 1.6% in the full sample to 3.3% in the sample of changers).

Another result that can easily be seen in the figure is that the effect of Hotness is much more pronounced when a number has been drawn in the previous week vs. when it has not been drawn. This holds for both indicators of betting (compare steepness of the sequence in the right vs. left half of each panel).

The finding that effects are much smaller in the full sample than in the other two samples shows that the GF and the HHF are mainly driven by variation at the intensive margin (how betting changes in response to drawings in previous weeks) rather than variation at the extensive margin (players entering or exiting the lottery, or changing their total bets).[23] In fact, regressions where the samples are split into frequent and infrequent players (playing in

[23]This is so because the full sample contains many data points with people who do not participate in a particular week, and their inclusion in the sample tends to water down the effects. This is confirmed in a regression where the decision to play is regressed on Drawn and Hotness: these variables have no significant effect on the decision to play.

Table 3: The effect of arbitrage

Dep. var.: MoneyBet	(1) Est. (s.e.)	(2) Est. (s.e.)
Constant	.4352 (.0348)***	.4368 (.0348)***
Drawn	-.0049 (.0028)*	-.0209 (.0040)***
Hotness	.0012 (.0007)*	-.0018 (.0008)**
Rollover	.2460 (.0008)***	.2443 (.0078)***
Drawn x Rollover	-.0020 (.0032)	.0073 (.0052)
Hotness x Rollover	-.0001 (.0012)	.0017 (.0014)
Drawn x Hotness		.0160 (.0021)***
Drawn x Hotness x Rollover		-.0088 (.0044)**
Lagged dep. var.	.6090 (.0266)***	.6090 (.0266)***
R^2	.368	.368
# Data points	25'084'404	
# Players	25'807	

Notes: The table reports estimations for the full sample of the regressions specified in eqs. 5 and 6, including interactions of Drawn and Hotness (and its interaction in specification (2)) with the rollover dummy. The dependent variable is MoneyBet. Estimated coefficients for the fixed number effects (jointly statistically significant) are not reported. Standard errors are robust to within-player dependency. The stars ***, ** or * indicate that the effect of the variable is statistically significant at the 1%, 5% or 10% level, respectively.

more than half versus less than half of the weeks), or into relatively 'heavy' and 'light' players (betting more or less money than the median player), show that the effects are strongest for frequent, heavy players (see Tables A3 to A6 in the Appendix).[24] Moreover, the GF results are driven by male players (see Tables A8 and A9 in the Appendix, or Suetens and Tyran, 2012).

Our results are robust to alternative specifications of the Hotness variable (see Tables A1 and A2 in the Appendix), and to the exclusion of 'lucky' numbers that are typically below 10 (see Table A7).[25]

Table 3 addresses the question of whether GF and HHF patterns are different in weeks where the main prize is rolled-over from the previous weeks. Because demand for lotto tickets does normally not increase in proportion to the increase in prize money, the expected payoff

[24] As shown in Tables A4 and A6, infrequent or light players in the full sample tend to pick numbers drawn in the previous week rather than avoid them (positive effect of Drawn in specification (5)), in particular, numbers that are hot as well (positive effect of interaction term in specification (6)).

[25] By excluding birthdays (i.e. numbers for 1 to 31), we exclude 86% of the data points, so it is not clear we should expect the same effects *a priori*.

in these weeks is higher (see Perez, 2013). It could be that this stimulates 'arbitrageurs' with a taste for gambling to enter, and that the behavior of these new entrants cancels out the patterns observed in their absence. For example, arbitrageurs may bet 'against the odds' or choose numbers randomly, without taking into account the recent drawing history. We tested for this possibility by running regressions for the full sample including interactions between the recent drawing history variables and the rollover variable, and using MoneyBet as a dependent variable (because NumberBet does not measure the amount bet). Table 3 shows that while the interaction of Drawn and Hotness is positive in non-rollover weeks, it is significantly negative in rollover weeks (see coefficient on Drawn x Hotness x Rollover in specification (2) in the table). This may be due to arbitrageurs entering the lottery, and avoiding hot numbers that are also drawn in the previous week. In specification (1) we do not find an interaction between the drawing history variables and rollover.[26]

To summarize, our results for the aggregate level indicate that there is a systematic relation between betting behavior and the recent drawing history of numbers. On average, players bet less on numbers drawn in the previous week than on numbers not drawn, as long as these numbers are not hot (between 1.6 and 3.8 percent). Players also bet more on drawn numbers, the hotter they are (marginal effect between 0.9 and 1.4 percent). The effects are driven by male, frequent, heavy players, and tend to be less strong in rollover weeks with.

4 Individual-level analyses

This section studies whether the two fallacies simply coexist in the aggregate because some players are prone to the GF and others to the HHF, or whether instead, the same players who are prone to the GF are also the ones who tend to be prone to the HHF if a number is sufficiently 'hot' (as predicted by Rabin and Vayanos, 2010). Our empirical approach in

[26]As one would expect, such an effect is not observed for the sample of active players, where entry and exit of arbitrageurs does not shape effects by definition.

this section involves estimating regressions for each individual. We focus on the variable MoneyBet as defined by eq. 2 since this is the richest proxy for a player's betting behavior. Given that we do not have enough data points at the player level to correct for fixed number effects, we focus on the change in MoneyBet. For each player, we thus estimate how much he changes his bet on a particular number as a function of the recent drawing history. The GF is operationalized as a decrease in bets on numbers drawn in the previous week compared to numbers not drawn. We define HHF-players as those who bet more on numbers drawn in the previous week, the hotter they are. Also, we focus on the active players, since the participation choice of players does not depend on the recent drawing history. In particular, we estimate the following regression for each player i who bets in week t:

$$\begin{aligned} \Delta \text{MoneyBet}_{ijt} = \beta_{0j} + \beta_{1i}\text{Drawn}_{jt-1} + \beta_{2i}\text{Hotness}_{jt-1} + \beta_{3i}\text{Rollover}_{t-1} + \epsilon_{ijt}, \\ \text{with } i = 1, ..., N; j = 1, ..., 36; t = 1, ..., T_i. \end{aligned} \tag{7}$$

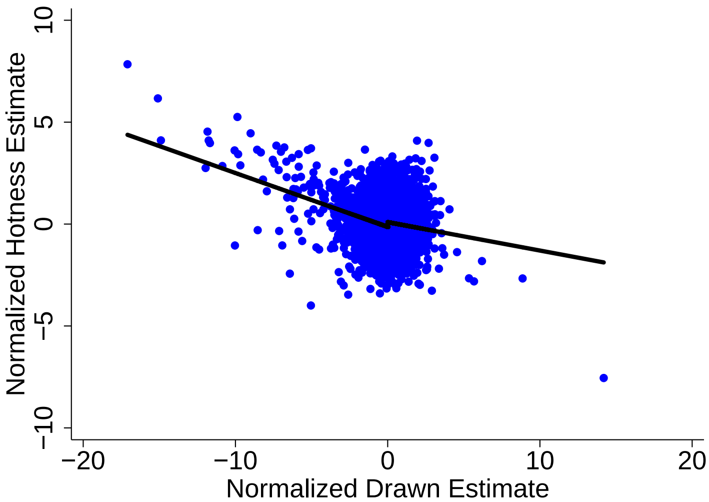
According to the RV-hypothesis, for players prone to the GF in the 'short' run and to the HHF in the 'long' run, the effect of Drawn is negative, and the effect of Hotness is positive in eq.(7). We control for rollover weeks by including a rollover dummy. We summarize the results from the regressions in two ways: (1) by plotting the estimated effects of Hotness as a function of those of Drawn, and (2) by classifying players according to the sign of the estimated effect of Drawn and Hotness.

Figure 3 plots the normalized individual-level Hotness estimates as a function of the normalized individual-level Drawn estimates. The figure includes a regression line that allows for a structural break at the point where the Drawn estimate is zero. The structural break is included in order to allow for a different relation between the Hotness and Drawn estimates, depending on whether the Drawn estimate is negative or not.[27] The figure shows that, overall, there is a negative relation between the Hotness and Drawn estimates. This relation

[27]Because none of the estimates is exactly equal to 0, the structural break in the regression line does not appear at exactly 0.

is consistent with the hypothesized behavioral pattern: players who decrease bets on numbers drawn in week $t-1$ also tend to increase bets on hot numbers. Importantly, the relation between the Hotness and Drawn estimates is significantly stronger — the regression line in Figure 3 is steeper — for the range where the Drawn estimates are negative. The tendency that players decrease bets on numbers drawn in week $t-1$ as well as increase bets on hot numbers is thus stronger than the tendency that players increase bets on numbers drawn in week $t-1$ as well as decrease bets on hot numbers. We take this as evidence supportive of the theoretical prediction that players who are prone to the GF also tend to be prone to the HHF.

Figure 3: Hotness estimates as a function of Drawn estimates



Notes: The figure shows a scatter plot of the normalized estimates of Hotness in eq. 7 as a function of the normalized estimates of Drawn from the same regression equation. The plot is based on 7'323 observations (covering the same number of players). The black line is a regression line that runs through the observations, with a structural break at the point where the estimate of Drawn is zero. It shows the predicted normalized Hotness estimates in the following regression: $\hat{\beta}_{2i} = \gamma_0 + \gamma_1 z\hat{\beta}_{1i} + \gamma_2 D_i + \gamma_3 D_i z\hat{\beta}_{1i}$ where $z\hat{\beta}_{1i}$ and $z\hat{\beta}_{2i}$ are the normalized individual-level Drawn and Hotness estimates from eq. 7, respectively, and $D_i = 1$ if $\hat{\beta}_{1i} > 0$. We used robust regression (command rreg in Stata). The regression results are $\hat{\gamma}_0 = -.135$, $\hat{\gamma}_1 = -.280$, $\hat{\gamma}_2 = .184$, and $\hat{\gamma}_3 = .161$ ($p < .001$ in all cases).

Table 4 provides a classification of players based on the sign of the estimated effect of Drawn and Hotness. Panel (a) uses many but relatively noisy observations, while panel (b) uses few but highly informative ones. More specifically, panel (a) includes results using all players for whom eq. 7 can be estimated, and panel (b) uses only those players who are

Table 4: Classification of players

(a) All players

Effect of Drawn	Effect of Hotness Negative	Positive	Total	Fisher p-value
Negative	1687 (23.04%)	2005 (27.38%)	3692 (50.42%)	< .001
Positive	1911 (26.10%)	1720 (23.49%)	3631 (49.58%)	.026
Total	3598 (49.13%)	3725 (50.87%)	7323	.294
Fisher p-value	.009	.001	.614	

(b) Significantly biased players

Effect of Drawn	Effect of Hotness Negative	Positive	Total	Fisher p-value
Negative	17	71	88	< .001
Positive	21	20	41	.913
Total	38	91	129	.001
Fisher p-value	.646	< .001	.004	

Notes: The table reports numbers of players where $\hat{\beta}_{1i} < 0$ and $\hat{\beta}_{1i} > 0$ and $\hat{\beta}_{2i} < 0$ and $\hat{\beta}_{2i} > 0$ in eq. 7. The p-values come from Fisher exact tests that compare the observed distributions with the uniform distribution. Panel (a) includes results from all players and panel (b) from players for whom both coefficients are significantly different from zero at the 10% level.

significantly biased at the 10% level.[28]

The row variable in Table 4 refers to the effect of Drawn and the column variable to the effect of Hotness. The first row in panel (a) of Table 4 shows that 3'692 players out of 7'323 (50.4%) decrease their bets on numbers drawn in the previous week. Out of the 7'323 players, 3'725 players (50.9%) increase their bets on numbers drawn in the previous week, the more frequently they have been drawn in the previous six weeks. Overall, 2'005 players both decrease bets on numbers drawn in the previous week and increase bets on numbers the hotter they are (27.4% compared to 25% with random picking).

Panel (b) of the table shows how GF and HHF relate if we consider players whose estimated reactions are statistically significant at the 10% level.[29] The classification follows

[28]While the absolute number of players for whom *both* effects (Drawn and Hotness in eq. 7) are significant at the 10% level is rather low (129), it is much higher than the number that would have been observed had all players chosen numbers randomly. Indeed, 129 is alsmot twice the number that would have been observed had all players chosen numbers randomly (10% times 10% of 7'323 is about 73).

[29]The qualitative nature of this classification is the same if we use the 1% or 5% level instead.

the same logic as Table 4, yielding a somewhat different picture. We now find that 68.2% (88 out of 129) decrease their bets on last week's winners and that 70.5% (91 out of 129) of players increase their bets on numbers on 'hot' numbers. As a result, the 'overlap' between both fallacies is 55% (71 out of 129), which is more than twice the percentage when all players would randomize (25%). Thus, the majority of the significantly biased players are prone to both fallacies.

5 Discussion

xx to be polished Given that lotto drawings are truly random, it seems absurd to believe that anyone can predict next week's numbers. Yet, our data suggests that the lotto players studied here, on average hold such beliefs and, curiously enough, the lotto agency itself describes the lotto as: 'a number game which is about predicting the correct numbers drawn' (translated from danskespil.dk, see 'rules of the game'). In particular, in line with recent economic theory, we find that players bet less on numbers drawn in the previous week (consistent with the gambler's fallacy), and bet more on these numbers, the 'hotter' they are (consistent with the hot-hand fallacy).

The magnitude in the gambler's fallacy in our data is substantially smaller than the magnitudes found by Clotfelder and Cook (1993) and Terrell (1994). They find that bets drop up to 36% and 18%, respectively, after a number was drawn. We can think of two reasons for why the magnitudes in our study are considerably smaller than in these studies. One explanation is that these studies use *aggregate* data on lottery play. So it may be that part of the observed drop in bets on winning numbers is caused by players who stop playing after their number has won, instead of by players who stop choosing these numbers. A second possible explanation is that, in our data, there is an indirect relation between winning numbers and winning amounts: a match between a number chosen by a player and a number drawn does not necessarily mean that the player wins a prize. He only does so if this chosen number is part of a winning *combination.* In contrast, the lotteries used in the above-mentioned studies are characterized by a one-to-one relation between winning numbers and (large) winning prizes. This, in fact, may have made previous week's winning numbers more salient to the players. In addition, and perhaps more importantly, it may have lead to more players dropping out after a win than would otherwise be the case.

Our results resonate well with Asparouhova et al. (2009), who ran a lab experiment where subjects are asked to predict the next observation in a random-walk process. The magnitudes of our effects are in the same ball-park as theirs. They find that players reduce their probability estimate of continuation by 1.8-2.6% after 'short' streaks, and increase it by 1.75% for each unit increase in streak length for 'long' streaks.[30] We find that players bet 1.6-3% less on numbers drawn in the previous week, and increase their bets by 0.9-1.4% for each unit increase in 'hotness'.[31]

The evidence we provide for the existence of biased inference from noisy data is potentially relevant for a number of anomalies that seem to be common in financial decision-making and financial markets. For example, it may explain why investors have a willingness to pay for 'expert' predictions of investment performance (Powdthavee and Riyanto, 2012), and if sufficiently prevalent, it may also explain why stock prices underreact to news, particularly if driven by small investors (e.g. Hvidkjaer, 2006). Indeed, it does not seem entirely implausible that such biases may manifest themselves at least in some financial markets given that people

[30]Asparouhova et al. (2009) find that players reduce their probability estimate of continuation by 0.9% for each unit increase in streak length for short streaks. In order to translate this number into our context, taking into account the difference in lottery (in their lottery, the probability of winning is 50%, making streaks or the occurrence of hot numbers much more likely), we calculated the effect for streak lengths that occur with a probability close to 7/36 (which is the probability a number is drawn in our lottery). In the 50% lottery, 7/36 lies between the probability of observing a streak of 2 (which occurs with probability .25) and a streak of 3 (which occurs with probability .125).

[31]The results reported by Croson and Sundali (2005) and Sundali and Croson (2006) do not allow for a direct comparison of marginal effects.

who play lotto also seem to be likely to invest in lotto-type stocks, such as low-priced stocks with high idiosyncratic volatility and skewness (Kumar, 2009). Clearly, the extent to which biases observed in lotto gambling extrapolate to financial markets, depends on differences in possibilities to engage in arbitrage — rational investors may compensate the behavior of irrational ones such that no effect is observed in the aggregate (e.g. Fehr and Tyran, 2005). Although the scope for arbitrage is rather limited in lotto gambling, we find evidence suggestive of arbitrage. In particular, possibly due to the entry of arbitrageurs, some of the biases are less strong in weeks with large jackpots.[32]

Our setting has many advantages over field data: it is a natural experiment in the sense that the data generation process is tightly controlled, it is transparent and truly random, and substantial amounts are at stake. Our lotto data come from a natural setting, and it is therefore also less controlled than a laboratory setting. For example, we do not know to which extent players have consulted information about past drawings. Neither do we know whether some players (arbitrageurs) are aware of the common fallacies among lottery players, or can we pin down the causal effect of the pari-mutuel nature of the lottery on betting behavior (see Camerer, 2011, for a general discussion on 'lab' versus 'field' data). The data at hand do not allow us to study or model why people play the lottery, how much they bet, why they use a particular selection format, or why they switch between different selection formats. We hope that future laboratory studies will be able to address these issues We therefore think that our study nicely complements previous evidence from lotto play and also complements evidence coming from the laboratory.

Earlier research (Cook et al. 1998) has shown that frequent lottery players (in the US) have lower intellectual efficiency), and oechssler et al. (xx) have found in laboratory experiments that proneness to various anomalies is related to low cognitive ability.

[32]The difference in magnitude of the gambler's fallacy between Clotfelder and Cook's study ('traditional' lottery) and Terrell's study (pari-mutuel lottery), can also be attributed to differences in scope for arbitrage.

References

Asparouhova, E., Hertzel, M., and Lemmon, M. (2009). Inference from streaks in random outcomes: experimental evidence on beliefs in regime shifting and the law of small numbers. *Management Science*, 55:1766–1782.

Bar-Eli, M., Avugos, S., and Raab, M. (2006). Twenty years of “hot hand” research: Review and critique. *Psychology of Sport and Exercise*, 7:525–553.

Bar-Hillel, M. and Wagenaar, W. (1991). The perception of randomness. *Advances in Applied Mathematics*, 12:428–454.

Camerer, C. (1989). Does the basketball market believe in the hot hand? *American Economic Review*, 79:1257–1261.

Camerer, C. (2011). The promise and success of lab-field generalizability in experimental economics: A critical reply to Levitt and List. Technical report, California Institute of Technology.

Carhart, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52:57–82.

Clotfelder, C. and Cook, P. (1993). The “gambler’s fallacy” in lottery play. *Management Science*, 39:1521–1525.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *Journal of Finance*, 63:1977–2011.

Croson, R. and Sundali, J. (2005). The gambler’s fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty*, 30:195–209.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under- and overreactions. *Journal of Finance*, 53:1839–1885.

De Bondt, W. and Thaler, R. (1985). Does the stock market overreact? *Journal of Finance*, 40:793–805.

De Bondt, W. and Thaler, R. (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*, 42:557–581.

DellaVigna, S. and Pollet, J. (2009). Investor inattention and Friday earnings announcements. *Journal of Finance*, 64:709–749.

Fehr, E. and Tyran, J.-R. (2005). Individual irrationality and aggregate outcomes. *Journal of Economic Perspectives*, 19:43–66.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314.

Guryan, J. and Kearney, M. (2008). Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review*, 98:458–473.

Hendricks, D., Patel, J., and Zeckhauser, R. (1993). Hot hands in mutual funds: Short-run persistence of performance. *Journal of Finance*, 48:93–130.

Hirshleifer, D., Lim, S., and Teoh, S. (2009). Drive to distraction: Extraneous events and underreaction to earnings news. *Journal of Finance*, 64:2289–2326.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. *Journal of Finance*, 54:2143–2184.

Huber, J., Kirchler, M., and Stöckl, T. (2010). The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and Decision*, 68:445–462.

Hvidkjaer, S. (2006). A trade-based analysis of momentum. *Journal of Financial Studies*, 19:457–491.

Jagannathan, R., Malakhov, A., and Novikov, D. (2010). Do hot hands exist among hedge fund managers? An empirical evaluation. *Journal of Finance*, 65:215–255.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: implications for stock market efficiency. *Journal of Finance*, 48:65–91.

Kumar, A. (2009). Who gambles in the stock market? *Journal of Finance*, 64:1889–1933.

Laplace, P. (1812). *A philosophical essay on probabilities.* Wiley and Sons, New York, Translated by F.W. Truscott and F.L. Emory from *Essai philosophique sur les probabilités, 1902*.

Loh, R. and Warachka, M. (2012). Streaks in earnings surprises and the cross-section of stock returns. *Management Science*, 58:1305–1321.

Miller, J. and Sanjurjo, A. (2014). The hot hand: A cognitive illusion? Mimeo.

Offerman, T. and Sonnemans, J. (2004). What's causing overreaction? An experimental investigation of recency and the hot-hand effect. *Scandinavian Journal of Economics*, 106:533–553.

Oskarsson, A., van Boven, L., McClelland, G., and Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychological Bulletin*, 135:262–285.

Perez, L. (2013). The Who and Why of lottery: Empirical highlights from the seminal economic literature. *Journal of Economic Surveys*, 27:915–940.

Powdthavee, N. and Riyanto, Y. (2012). Why do people pay for useless advice? Implications of gamblers and hot-hand fallacies in a false-expert setting. Discussion Paper 6557, IZA.

Rabin, M. (2002). Inference by believers in the law of small numbers. *Quarterly Journal of Economics*, 117:775–816.

Rabin, M. and Vayanos, D. (2010). The gambler's and hot-hand fallacies: Theory and applications. *Review of Economic Studies*, 77:730–778.

Rapoport, A. and Budescu, D. (1997). Randomization in individual choice behavior. *Psychological Review*, 104:603–617.

Simon, J. (1999). An analysis of the distribution of combinations chosen by UK national lottery players. *Journal of Risk and Uncertainty*, 17:243–276.

Suetens, S. and Tyran, J.-R. (2012). The gambler's fallacy and gender. *Journal of Economic Behavior & Organization*, 83:118–124.

Sundali, J. and Croson, R. (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making*, 1:1–12.

Terrell, D. (1994). A test of the gambler's fallacy: Evidence from pari-mutuel games. *Journal of Risk and Uncertainty*, 8:309–317.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76:105–110.

Yaari, G. and David, G. (2012). “Hot hand” on strike: Bowling data indicates correlation to recent past results, not causality. *PLoS ONE*, 7:e30112.

A Supplementary Appendix

A.1 Application of Rabin and Vayanos to Lotto

This section applies the model Rabin and Vayanos (2010) (RV) to the context of lotto play in Denmark. We assume that an individual observes a sequence of lotto drawings that depends on the probability that a number is drawn and an *i.i.d.* normal shock. The signal s_t in week $t = 1, 2, ...$ is

$$s_t = \theta_t + \epsilon_t, \tag{8}$$

where θ_t is the probability that a number is drawn and ϵ_t the normal shock with mean zero and variance $\sigma_\epsilon^2 > 0$. The probability that a number is drawn is assumed to evolve according to the following auto-regressive process:

$$\theta_t = \mu + \rho(\theta_{t-1} - \mu) + \eta_t, \tag{9}$$

where μ is the long-run mean, $\rho \in [0, 1]$ the persistence parameter, and η_t an *i.i.d.* normal shock with mean zero, variance σ_η^2, and independent of ϵ_t. Given that the drawing machines and the sets of lotto balls used are replaced from time to time, the parameter ρ can be seen as measuring the persistence of the drawing machines and lotto balls: a ρ close to one implies a high probability of using the same machine and the same set of balls and a ρ close to zero implies a high switching or replacing probability.[33] The variance σ_η^2 measures the variability in drawing outcomes between different drawing machines and sets of balls: a high σ_η^2 would imply that different machines or sets of balls generate different drawing probabilities. Since the lotto agency does everything it can to generate fair drawings, we assume that $\sigma_\eta^2 = 0$. Therefore, θ_t is constant and equal to the probability that a number is drawn.

[33] The Danish lotto agency operates with three sets of lotto balls and two identical drawing machines. The machines are switched every six months and the balls are replaced when they have been used fifty-two times. The public is not informed about the timing of switching the machine nor replacing the balls.

The GF is modeled as the mistaken belief that ϵ_t is not *i.i.d.* but exhibits reversals in the following sense:

$$\epsilon_t = \omega_t - \alpha\rho \sum_{k=0}^{\infty} (\delta\rho)^k \epsilon_{t-1-k}, \tag{10}$$

where ω_t is an *i.i.d.* normal shock with mean zero and variance σ_ω^2, and $\alpha, \delta \in [0, 1)$. Whether a lotto player who falls prey to the GF starts to develop hot hand beliefs after observing long streaks depends on whether he is certain about the constancy of the probability that a number is drawn (equal to 7/36).

Consider first the case where a lotto player is absolutely certain about the drawing probability: he believes that the probability that a number is drawn exhibits no variability ($\tilde{\sigma}_\eta^2 = 0$) and is equal to 7/36. In this case $\theta_t = \mu$ and the player observes signals according to $s_t = \mu + \epsilon_t$, where ϵ_t refers to the GF beliefs modeled in eq. 10.

When a lotto player is uncertain about the drawing probability, for example, because he does not fully trust the drawing mechanism and believes in variability between different sets of lotto balls, RV show that the player will develop a belief in the hot hand. In particular, Proposition 5 in RV shows that under mild assumptions[34] an uncertain player will develop the wrong belief about ρ, namely $\tilde{\rho} = \underline{\rho}$, and the additional wrong belief that the drawing probability varies over time: $\tilde{\sigma}_\eta^2 > 0$. The intuition is that in order to explain the absence of reversals, an uncertain lotto player will underestimate the persistence of the drawing machines and lotto balls (i.e. underestimate ρ) and overestimate the variability in drawing outcomes between these machines or sets of balls. The consequence is that an uncertain lotto player who believes in the law of small numbers will expect reversals after short streaks and continuation after long streaks (see Proposition 6 in RV).

Another model that predicts fallacy reversal is Rabin (2002). This model assumes individuals draw balls from an urn with replacement, but they think that replacement occurs

[34] The player should be confident that there is some persistence of the drawing mechanism: $\tilde{\rho} \in [\underline{\rho}, 1]$ with $\underline{\rho} > 0$. In other words, the player should believe that the drawing machine and the lotto balls are not replaced every week.

only every odd period. The underlying state in this model does not change over time (so that $\rho = 1$).

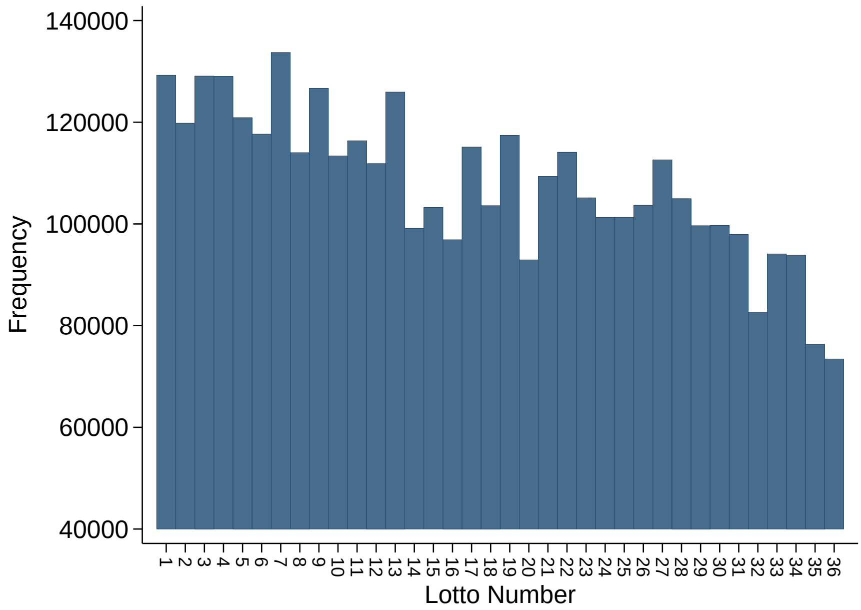
A.2 Overview of the different systems in Systemlotto

Option	Type of system	# chosen numbers in a set	# tickets/combinations generated
1	M	8	8
2	M	9	36
3	M	10	120
4	M	11	330
5	M	12	792
6	R	10	8
7	R	10	30
8	R	11	20
9	R	11	34
10	R	12	12
11	R	12	24
12	R	12	48
13	R	13	18
14	R	13	66
15	R	14	48
16	R	14	132
17	R	15	24
18	R	15	69
19	R	16	32
20	R	16	109
21	R	16	240
22	R	17	272
23	R	18	82
24	R	19	338
25	R	20	450
26	R	20	1040
27	R	21	198
28	R	23	345
29	R	24	455
30	R	25	600
31	C	17	17
32	C	18	33
33	C	19	52
34	C	20	20
35	C	20	80
36	C	22	60
37	C	24	24
38	C	24	120
39	C	25	100
40	C	25	200
41	C	28	194
42	C	30	268
43	C	31	155

Notes: Under system 'M' all potential combinations of the chosen numbers in a set are generated. Systems 'R' and 'C' are both reduced systems, in which only some of the potential combinations are generated. Systems 'C' are more heavily reduced than systems 'R'. That is, a smaller share of the potential combinations are generated in systems 'C' than in systems 'R'.

A.3 Popularity of lotto numbers

Figure A1: Frequency distribution of lotto numbers



Notes: The chart shows for each lotto number the total number of times the number is picked in the period of study (weeks 25 to 52 of 2005) using the Systemlotto selection device over the Internet. Note that a player is not counted more than once in a week.

A.4 Robustness checks

This section reports the results from various robustness checks. In all tables reported in this section, panel (a) reports results from pooled estimations based on the full sample of players, panel (b) reports results based on the sample of players (i.e. those who bet in week t), and panel (c) reports results based on the active players who change their numbers at least once over the entire in the period of 28 weeks. The regressions include control variables but in the interest of readability, we only report the estimated coefficients for the main independent variables of interest (i.e. Drawn and Hotness). Standard errors are robust to within-player dependency. The stars ***, ** or * indicate that the effect of the variable is statistically significant at the 1%, 5% or 10% level, respectively.

Table A1: Pooled regression results with hotness defined over a shorter history

Dep. var.:	**NumberBet**		**MoneyBet**	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
		(a) Full sample		
Drawn	-.0001 (.0001)	-.0010 (.0002)***	-.0056 (.0029)*	-.0178 (.0038)***
Hotness	.0002 (.0000)***	-.0001 (.0000)**	.0009 (.0006)	-.0023 (.0008)*
Drawn x Hotness		.0013 (.0001)***		.0155 (.0023)***
R^2	.639	.639	.368	.368
# Data points		25'084'404		
# Players		25'807		
		(b) Active players		
Drawn	-.0020 (.0003)***	-.0026 (.0004)***	-.0279 (.0071)***	-.0437 (.0093)***
Hotness	.0005 (.0001)***	.0004 (.0001)***	.0047 (.0014)***	.0007 (.0015)
Drawn x Hotness		.0007 (.0002)***		.0201 (.0042)***
R^2	.855	.855	.668	.668
# Data points		8'525'016		
# Players		17'318		
		(c) Changers		
Drawn	-.0047 (.0007)***	-.0061 (.0009)***	-.0685 (.0178)***	-.1060 (.0232)***
Hotness	.0012 (.0002)***	.0008 (.0002)***	.0087 (.0033)***	-.0009 (.0036)
Drawn x Hotness		.0018 (.0005)***		.0481 (.0104)***
R^2	.672	.672	.518	.518
# Data points		3'380'508		
# Players		8'224		

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 where hotness is defined as the number of times a lotto number has been drawn in weeks $t-2$ to $t-5$.

Table A2: Pooled regression results with hotness defined over a longer history

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	-.0001 (.0001)	-.0010 (.0002)***	-.0055 (.0029)*	-.0197 (.0040)***
Hotness	.0001 (.0000)***	-.0000 (.0000)	.0013 (.0005)**	-.0007 (.0006)*
Drawn x Hotness		.0011 (.0001)***		.0122 (.0020)***
R^2	.639	.639	.368	.368
# Data points	25'084'404			
# Players	25'807			
(b) Active players				
Drawn	-.0020 (.0003)***	-.0030 (.0004)***	-.0279 (.0071)***	-.0430 (.0092)***
Hotness	.0004 (.0001)***	.0002 (.0001)***	.0034 (.0011)***	.0013 (.0012)
Drawn x Hotness		.0008 (.0002)***		.0130 (.0034)***
R^2	.855	.855	.668	.668
# Data points	8'525'016			
# Players	17'318			
(c) Changers				
Drawn	-.0047 (.0007)***	-.0070 (.0010)***	-.0684 (.0178)***	-.1031 (.0231)***
Hotness	.0008 (.0002)***	.0005 (.0002)***	.0065 (.0028)***	.0017 (.0031)
Drawn x Hotness		.0019 (.0005)***		.0301 (.0083)***
R^2	.672	.672	.518	.518
# Data points	3'380'508			
# Players	8'224			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 where hotness is defined as the number of times a lotto number has been drawn in weeks $t-2$ to $t-7$.

Table A3: Pooled regression results with relatively frequent players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	-.0008 (.0003)***	-.0017 (.0004)***	-.0167 (.0064)***	-.0309 (.0082)***
Hotness	.0002 (.0001)***	.0001 (.0001)	.0032 (.0010)***	.0006 (.0012)
Drawn x Hotness		.0009 (.0002)***		.0146 (.0036)***
R^2	.763	.763	.633	.633
# Data points	8'896'716			
# Players	9'153			
(b) Active players				
Drawn	-.0015 (.0003)***	-.0021 (.0004)***	-.0234 (.0069)***	-.0374 (.0086)***
Hotness	.0003 (.0001)***	.0002 (.0001)***	.0043 (.0012)***	.0017 (.0013)
Drawn x Hotness		.0006 (.0002)***		.0145 (.0034)***
R^2	.887	.887	.759	.759
# Data points	7'284'168			
# Players	9'153			
(c) Changers				
Drawn	-.0040 (.0008)***	-.0055 (.0010)***	-.0609 (.0180)***	-.0959 (.0222)***
Hotness	.0008 (.0002)***	.0005 (.0002)***	.0086 (.0030)***	.0022 (.0035)
Drawn x Hotness		.0016 (.0005)***		.0362 (.0085)***
R^2	.735	.735	.619	.619
# Data points	2'795'256			
# Players	3'654			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 based on players who play in more than half of the weeks in the period of study (at least 15 weeks).

Table A4: Pooled regression results with relatively infrequent players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	.0002 (.0001)*	-.0004 (.0002)**	.0023 (.0017)	-.0052 (.0027)*
Hotness	.0001 (.0001)***	.0000 (.0001)	.0008 (.0007)	-.0006 (.0008)
Drawn x Hotness		.0006 (.0001)***		.0078 (.0023)***
R^2	.327	.327	.096	.096
# Data points	15'892'200			
# Players	16'350			
(b) Active players				
Drawn	-.0041 (.0008)***	-.0058 (.0012)***	-.0238 (.0130)*	-.0357 (.0190)*
Hotness	.0012 (.0003)***	.0009 (.0003)**	.0018 (.0058)	-.0004 (.0064)
Drawn x Hotness		.0018 (.0008)**		.0123 (.0138)
R^2	.674	.674	.448	.48
# Data points	1'134'648			
# Players	7'866			
(c) Changers				
Drawn	-.0055 (.0016)***	-.0086 (.0023)***	-.0437 (.0278)	-.0648 (.0406)
Hotness	.0025 (.0007)***	.0019 (.0007)***	.0003 (.0122)	-.0035 (.0138)
Drawn x Hotness		.0032 (.0016)**		.0220 (.0288)
R^2	.406	.406	.328	.328
# Data points	531'756			
# Players	4'407			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 based on players who play in less than half of the weeks in the period of study (14 weeks or less).

Table A5: Pooled regression results with relatively heavy players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	-.0006 (.0002)***	-.0018 (.0003)***	-.0141 (.0062)**	-.0390 (.0083)***
Hotness	.0002 (.0001)***	-.0000 (.0001)	.0018 (.0012)	-.0027 (.0015)*
Drawn x Hotness		.0013 (.0002)***		.0257 (.0043)***
R^2	.591	.591	.345	.345
# Data points	11'771'892			
# Players	12'111			
(b) Active players				
Drawn	-.0043 (.0006)***	-.0055 (.0010)***	-.0677 (.0172)***	-.1017 (.0223)***
Hotness	.0007 (.0001)***	.0005 (.0002)***	.0084 (.0031)***	.0022 (.0034)
Drawn x Hotness		.0013 (.0004)***		.0352 (.0090)***
R^2	.781	.781	.626	.759
# Data points	3'504'708			
# Players	7'796			
(c) Changers				
Drawn	-.0075 (.0012)***	-.0098 (.0015)***	-.1240 (.0323)***	-.1848 (.0419)***
Hotness	.0012 (.0004)***	.0008 (.0003)***	.0115 (.0057)**	.0005 (.0065)
Drawn x Hotness		.0023 (.0007)***		.0631 (.0166)***
R^2	.613	.613	.485	.485
# Data points	1'856'448			
# Players	4'785			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 based on players who buy more tickets than the median player in the sample.

Table A6: Pooled regression results with relatively light players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	.0004 (.0001)***	-.0003 (.0002)	.0023 (.0005)***	-.0005 (.0009)
Hotness	.0001 (.0000)*	-.0001 (.0000)	.0003 (.0002)*	-.0002 (.0002)
Drawn x Hotness		.0008 (.0001)***		.0028 (.0007)***
R^2	.714	.714	.598	.598
# Data points	13'312'512			
# Players	13'696			
(b) Active players				
Drawn	-.0004 (.0002)*	-.0008 (.0003)***	.0003 (.0008)	-.0022 (.0013)*
Hotness	.0002 (.0001)***	.0001 (.0001)	.0007 (.0003)**	.0002 (.0003)
Drawn x Hotness		.0004 (.0002)**		.0026 (.0010)**
R^2	.912	.912	.736	.736
# Data points	5'020'308			
# Players	9'522			
(c) Changers				
Drawn	-.0011 (.0006)*	-.0024 (.0010)***	.0007 (.0026)	-.0069 (.0041)*
Hotness	.0006 (.0002)***	.0004 (.0002)	.0019 (.0009)**	.0005 (.0011)
Drawn x Hotness		.0013 (.0006)**		.0079 (.0031)**
R^2	.736	.736	.548	.548
# Data points	1'524'060			
# Players	3'439			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 based on players who buy fewer tickets than the median player in the sample.

Table A7: Pooled regression results excluding lotto numbers below 10

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	-.0005 (.0001)***	-.0016 (.0002)***	-.0122 (.0031)***	-.0243 (.0042)***
Hotness	.0002 (.0000)***	-.0000 (.0001)	.0028 (.0006)***	.0004 (.0008)
Drawn x Hotness		.0011 (.0001)***		.0119 (.0021)***
R^2	.632	.632	.361	.361
# Data points	18'813'303			
# Players	25'807			
(b) Active players				
Drawn	-.0023 (.0003)***	-.0030 (.0004)***	-.0314 (.0076)***	-.0485 (.0098)***
Hotness	.0005 (.0001)***	.0003 (.0001)***	.0053 (.0015)***	.0020 (.0016)
Drawn x Hotness		.0008 (.0002)***		.0170 (.0039)***
R^2	.850	.850	.663	.663
# Data points	6'393'762			
# Players	17'318			
(c) Changers				
Drawn	-.0053 (.0007)***	-.0071 (.0010)***	-.0772 (.0191)***	-.1169 (.0244)***
Hotness	.0011 (.0002)***	.0007 (.0002)***	.0115 (.0036)***	.0039 (.0040)
Drawn x Hotness		.0018 (.0005)***		.0394 (.0096)***
R^2	.666	.666	.515	.515
# Data points	2'535'381			
# Players	8'224			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 including lotto numbers above 9 (from 10 to 36).

Table A8: Pooled regression results of male players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	-.0003 (.0001)**	-.0013 (.0002)***	-.0084 (.0035)**	-.0228 (.0047)***
Hotness	.0002 (.0000)***	-.0000 (.0000)	.0012 (.0007)*	-.0014 (.0008)*
Drawn x Hotness		.0010 (.0001)***		.0149 (.0024)***
R^2	.638	.638	.348	.348
# Data points	20'613'204			
# Players	21'207			
(b) Active players				
Drawn	-.0025 (.0003)***	-.0032 (.0005)***	-.0346 (.0087)***	-.0511 (.0113)***
Hotness	.0005 (.0001)***	.0004 (.0001)***	.0051 (.0016)***	.0021 (.0017)
Drawn x Hotness		.0008 (.0002)***		.0171 (.0045)***
R^2	.844	.844	.653	.653
# Data points	6'858'612			
# Players	14'099			
(c) Changers				
Drawn	-.0057 (.0008)***	-.0074 (.0011)***	-.0820 (.0210)***	-.1198 (.0272)***
Hotness	.0011 (.0002)***	.0007 (.0002)***	.0090 (.0037)**	.0021 (.0041)
Drawn x Hotness		.0018 (.0005)***		.0392 (.0107)***
R^2	.662	.662	.502	.502
# Data points	2'825'172			
# Players	6'943			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 for male players.

Table A9: Pooled regression results of female players

Dep. var.:	NumberBet		MoneyBet	
	Eq. 5	Eq. 6	Eq. 5	Eq. 6
	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)	Est. (s.e.)
(a) Full sample				
Drawn	.0010 (.0002)***	.0002 (.0004)	.0082 (.0027)***	.0031 (.0047)
Hotness	.0001 (.0001)	-.0008 (.0001)	.0003 (.0007)	-.0006 (.0009)
Drawn x Hotness		.0008 (.0003)***		.0052 (.0035)
R^2	.683	.683	.541	.541
# Data points	4'471'200			
# Players	4'600			
(b) Active players				
Drawn	.0003 (.0004)	-.0007 (.0006)	-.0002 (.0053)	-.0106 (.0088)
Hotness	.0001 (.0001)	-.0001 (.0001)	-.0005 (.0015)	-.0024 (.0019)
Drawn x Hotness		.0008 (.0004)**		.0107 (.0055)*
R^2	.903	.903	.766	.766
# Data points	1'666'404			
# Players	3'219			
(c) Changers				
Drawn	.0002 (.0012)	-.0020 (.0016)	.0013 (.0158)	-.0298 (.0260)
Hotness	.0003 (.0003)	-.0001 (.0004)	-.0008 (.0044)	-.0065 (.0058)
Drawn x Hotness		.0023 (.0010)**		.0321 (.0160)**
R^2	.739	.739	.634	.634
# Data points	555'336			
# Players	1'281			

Notes: The table reports estimations of the regressions specified in eqs. 5 and 6 for female players.