Proceedings of CONF-MPCS 2026 Symposium: Theoretic Physics and Plasma Physics
DOI: 10.54254/2753-8818/2026.DL34356

Beyond the Jackpot: Modeling Expected Returns and the Birthday Effect in Lotteries

Yihui Shao

YK Pao School, Shanghai, China
s25502@ykpaoschool.cn

Abstract. This paper focuses on the equations of the expected returns of lotteries as well as their limitations. It begins by introducing the background of the lottery market, highlighting its rapid growth and increasing participation due to easier accessibility, further emphasizing the importance of analysis to ensure a higher rate of return and better odds. The basic information, such as the definition and the concepts of expected rate of return as a framework and the lottery mechanism, are then examined and explained, including jackpot dynamics and pari-mutuel prizes. Finally, the paper explains and compares two prominent models for calculating the expected rate of return: the Mecklin-Donnelly model and the Abrams-Garibaldi model, covering their methods and referenced data. Through quantitative analysis, this paper concludes that the Mecklin-Donnelly formula significantly overestimates the expected rate of return in practical lottery systems, especially under real-world conditions involving shared jackpots, taxes, and varying ticket sales volumes.

Keywords: Lotteries, expected rate of return, birthday effect

1. Introduction

People are often attracted to purchasing lottery tickets with the expectation of winning. The global lottery market itself is enormous, and the lottery industry generates plenty of revenue for states each year. In the city of New York alone, lottery sales have increased to over $658 million in 2025 [1]. Millions of people participate in the games, ranging from regional brands, such as Lotto Texas, to large ones such as the Powerball. Although lotteries are actually promoted by stories of pure luck and life-changing jackpots, the vast majority of buyers lose money, according to statistics [2].

In recent decades, ticket buying has increased due to greater accessibility and widespread marketing. The digital lottery market is expected to experience increasing growth, with its value increasing from $12.58 billion in 2026 to $19.45 billion by 2031 according to projections [3]. Key contributors are the adoption of internet technology, advancements in mobile payment methods, such as Alipay etc. [3]. In addition, lotteries are often promoted as low-risk entertainment because the cost of a single ticket is small. Although this is true, frequent participation can lead to financial losses over time, especially among lower-income populations, who spend a large share of their income on tickets, hoping they will hit the jackpot.

Because of these social and economic implications, analyzing the expected returns of lottery games is important not only from a mathematical perspective but also from a societal point of view.

DOI: 10.54254/2753-8818/2026.DL34356

According to research in 2016, lottery buyers mostly aren't aware of the extremely small probabilities in lotteries and tend to overweigh these small numbers, which incentivizes them to buy more [4].

Therefore, by calculating the expected rates of return for different lotteries, consumers can make a rational and informed choice of purchases, as well as find the most suitable lottery that has the highest rate of return.

This paper will examine the expected rate of returns for lottery games using probability and expected value. Firstly, this paper will introduce the concept of expected rate of return, as well as the lottery mechanism. Secondly, the review will focus mainly on two formulas- the Mecklin-Donnelly formula and the Abrams Garibaldi Formula. In addition, comparisons and limitations of the formulas will also be observed and quantified such as the Birthday Effect.

2. Preliminaries

2.1. The expected rate of return

The expected rate of return is the expected value of all probable returns from an investment, weighted by its probabilities. Its logic originates from the simplest arithmetic average, where outcomes are very uncertain, such as in stock analysis.

In everyday life, when outcomes are equally probable in theory, an average is found. For example, in the case of coin flipping, assume 1 dollar is distributed if a coin lands on its heads, and 2 dollars for a coin landing on tails. If two experiments are conducted, theoretically, the coin will land on heads once and tails once, in theory, 3 dollars are distributed in total, and for every experiment, an average of $\frac{1+2}{2}=1.5$ dollars are earned. However, for stock analysis and lotteries, things get much more complicated than equally likely situations. This requires a stronger formula that can accurately represent the return. This is where the formula of the expected rate of return. The main formula is E(R)= $\sum_{i=1}^{n} p_i a_i$ [5].

p_i is the probability of the situation happening and a_i is the return in the situation. If there is a 30% chance to win 30 dollars, while there is a 70% chance of losing 20 dollars, according to the formula $30\% \times 30 - 70\% \times 20$ =-5, the expected value is $-5. This represents that, on average, $5 will be lost when playing this game in the long run [6]. A negative expected value shows that this game is not favorable in the long run, as the expected loss outweighs the game [7].

This formula is derived from the arithmetic average of all possible returns of the game. In the previous example, assume a person plays 100 times the same game. In theory, the person wins 30 dollars 30 times and loses 20 dollars 70 times. The final return is $\frac{30\times 30-20\times 70}{100}$ =-5. This value exactly aligns with the calculation using the expected value formula. In other words, the expected rate of return is just an arithmetic average of all likely outcomes, weighed according to their probabilities. It shows the long-run average gain or loss per game if the game is played multiple times.

EV turns the game into a number that reflects its average outcome: for each possible circumstance , take the probability and multiply it by the payout received, then add them together and subtract the price of a ticket [8].

DOI: 10.54254/2753-8818/2026.DL34356

2.2. Lottery mechanism

In a typical American Lottery, the prizes are usually categorized into three different sets(although it may be even more complicated in reality). Each group functions very differently from one another, and this section will explain this from the largest prize to the smallest prize.

Firstly, it's the jackpot, which is well-known for even non-lottery players. The exact jackpot amount is usually not fully known in advance, but an estimated starting jackpot is announced by lottery organizations before ticket sales. However, lottery jackpots typically fluctuate depending on ticket sales since, as more people buy the lottery, the jackpot amount gets larger and larger. In addition, the jackpot is not guaranteed to be hit.

If no one wins the jackpot, the prize rolls over to the next draw, increasing the starting jackpot for the next round. The payoff for a winning ticket fluctuates each week, depending on the size of the prize pool. The pool for Powerball usually initiates at $40 million but increases from time to time, until someone hits the jackpot. On January 13, 2016, the jackpot reached $1.586 billion, the highest of all records. This is a key characteristic that will appear in the discussion of formulas [9].

Moreover, the jackpot is also split if multiple people hit the same combinations of numbers. This often leads to the expected rate of return being overestimated when it is not taken into account. The second category is the pari-mutuel prize for the second tier. Organizations typically decide on a fixed total amount and split it into multiple parts if many people hit the same numbers. Finally, it's the fixed price that doesn't fluctuate at all and stays the same no matter how many people buy the ticket.

3. Theoretical comparison

This project will relate to the paper by Abrams and Garibaldi who are very dedicated to quantifying the expected rate of return for lotteries. In the paper, they developed an important formula, named after them-the Abrams-Garibaldi Formula-which this paper will use to compare with the Mecklin-Donnelly Formula. This model has two formulas.

3.1. The Abrams-Garibaldi model

3.1.1. Jackpot cutoff C_0

The first formula assumes a ticket price of $1 and categorizes the prizes into fixed prizes and parimutuel prizes. The key variables that are taken into consideration here are f, which is the proportion of each dollar that excludes the fixed prices, F as the proportion of each $1 dollar entering the jackpot and operator profit, and t as the total number of possible ticket combinations. The jackpot cut-off is F*t, representing the minimum jackpot size required for a ticket to generate a positive expected value. As shown in Table 1, this threshold varies significantly across different lottery systems due to differences in t, f, and F.

If the actual jackpot C is less than the expected return, it is guaranteed to be negative, since even buying every combination loses money.

Table 1. Some statistics of lotteries ranging from 2009-2011 about t, f, F, and C_0

Lottery	t	f	F	C_0
Mega Millions	175,711,536	0.838	0.838	147m
Powerball	146,107,962	0.821	0.821	120m
Lotto Texas	25,827,165	0.957	0.910	23.5m
New Jersey Pick 6	13,983,816	0.947	0.855	11.9m

3.1.2. Expected rate of return formula

Second, the expected rate of return formula appears to be much more complicated, and is shown below: $eRoR$ =-f+ $\sum r_i\left(1-(1-p_i)^N\right)+Js\left(\frac{1}{t},N\right)$ [5].

The term -f accounts for fixed prizes and is the money lost without a choice. The second term symbols pari-mutuel prizes where r_i is the prize pool proportion and p_i the probability, and the final term captures jackpot expectation with splitting among different people who simultaneously win.

3.2. The Mecklin- Donnelly Model

The Mecklin- Donnelly Model is a paper published in 2005 by Mecklin and Donnelly. It consists of a lot of formulas, and the report will only include the one for the expected rate of return.

In the formula, the expected rate of return is calculated by $\sum p_i * x_i$ - ticket cost. By modifying the Powerball 2005 data into the equation, the final EoR equation is equal to: E(X)= $\frac{J}{120{,}526{,}770}$ $-$ 0.827 [10]

Just based on data of the 2005 Powerball, the cut-off jackpot is approximately $100 million, and only when the Jackpot exceeds $100 million, the expected rate of return is guaranteed to be positive.

3.3. Comparisons between models

The Mecklin-Donnelly formula has fundamental methodological and mathematical differences that this paper will address in this report.

3.3.1. Treatment of Jackpot Sharing

The Mecklin-Donnelly Formula doesn't consider Jackpot Sharing and nor does it consider pari-mutuel prizes in the formula. It automatically assumes that the lottery is not split, and every amount of money the buyer receives is fixed. However, this is generally not true for most lotteries, which include the parimutuel prize mentioned above. This leads to severe overestimation, which can be very harmful.

The complicating factor, however, is that as the jackpot grows, the number of ticket buyers increase as well, although not proportionally. Though the jackpot increases, thus leading to a higher expected rate of return, the increased number of ticket buyers increase the probability that the winning number will be shared by two or more people. Therefore, the increase in the expected rate of return is relatively mild, as players have to share this large sum of money [11].

DOI: 10.54254/2753-8818/2026.DL34356

This paper quantified the overestimation using Powerball 2007 data. As shown in Table 2, the key parameters for Mega Millions, including jackpot size, total combinations, and ticket sales, are used for the calculation.

Table 2. Statistics of Mega Millions in 2005

Mega Millions	
J (Jackpot)	$390,000,000 (before tax)
t (possibilities)	175,711,536
N (number of tickets sold)	170,000,000

Mecklin Donnelly Formula: EoR= $\frac{J}{t} = \frac{390\ million}{175{,}711{,}536} = 2.219$.

Abrams- Garibaldi Formula: EoR= J* $\frac{1-\left(1-\frac{1}{t}\right)^{N}}{N} = 1.422$.

After calculation, this is a Mecklin-Donnelly Formula has overestimated the value of about 0.797. This really matters because using this model, players are buying more lotteries because of a high expected rate of return which is only an illusion.

3.3.2. Role of N (sales of the ticket)

The Mecklin- Donnelly formula alarmingly idealizes the situation because it guarantees that one person will win the jackpot. However, this is not true, and in real life situations, if no one wins the jackpot, the prize rolls over to the next draw, increasing the starting jackpot for the next round. This means that N plays a role in this. In contrast, the Abrams- Garibaldi Formula takes this into serious account. Both the second term and the third term, accounting for the pari-mutuel prizes and the jackpot, consider N as a variable that plays a role in the expected rate of return.

Instinctively speaking, a bigger N (sales of tickets), the bigger probability that someone hits the jackpot, or at least the pari-mutuel prizes. This may influence the expected rate of return in three ways. The first way is that more N means an increased probability of people wining.

This can be mathematically proved by the Abrams- Garibaldi formula. One term representing the pari-mutuel prizes include $\sum r_i\left(1-(1-p_i)^N\right)$. As N gets larger, $\left(1-(1-p_i)^N\right)$ approaches 1, and there is a bigger probability of someone winning.

The second way is that when more people enter the lottery market, even if the jackpot is hit, a great number of people will split the prize, which decreases the money received by the participant, and thirdly, a bigger N means a bigger Jackpot, because the jackpot is adjusted using N.

4. Flaws for the formulas

4.1. The birthday effect

Even for the Abrams- Garibaldi Formula, problems surface. The formula assumes that there is an equal chance of winning with all combinations of numbers. However, this is very flawed. It is true that the probability of hitting the prize will be the same, but choosing hot numbers may very well increase the chances of splitting the prize with someone else, thus the money received decreases.

Reviews on lotteries show that players don't just select numbers randomly. As a matter of fact, players prefer to pick smaller numbers, particularly less than 30, either because they prefer small numbers as a lucky number, or because they are victims of the birthday effect. This is because

DOI: 10.54254/2753-8818/2026.DL34356

people who buy lotteries believe in lucky numbers, such as the date of their anniversary or their birthday, and usually pick these numbers as a combination. Because lotto games are parimutuel, such preferences result in poor ticket choices in terms of achieving below average returns [12]. This paper quantified the birthday effect using data based on assumption:

t (total number of distinct combinations) =25.8 million, $t_h = number\ of\ hot\ combinations$,a (the proportion of players who choose numbers randomly) =0.75, N (the sale of tickets) = 4.2 million, W (total number of people who hits the prize)

$$W_{cold} = 1 + \left(\mathrm{aN} - 1\right) * \frac{1}{t} \ = \ 1.22 \tag{1}$$

$$W_{hot} = 1 + \frac{aN}{t} + \frac{(1-a)N}{t_h} \ = \ 2.549 \tag{2}$$

$$Share_c = \frac{1}{1.22} = 0.891\,(89.1\%) \tag{3}$$

$$Share_h = \frac{1}{2.549} = 0.392\ \left(39.2\%\right) \tag{4}$$

$$Advantage = \frac{0.891}{0.392} = 127\% \tag{5}$$

This shows a huge advantage received if a cold number is chosen. For example, in the 2025 Powerball, the 720 million jackpots was split into 15.3 million per share before distributed. Among 47 ticket holders who won the jackpot, 42 chose a number below 31.

4.2. The prediction problem

Even though the Abrams- Garibaldi model sounds very promising, considering variables like N-sales of tickets and F, the problem lies in N itself. No organizations are keen to update the number of tickets sold in the market until the jackpot is hit. As a result, buyers can't use the Garibaldi Formula at all to predict their expected rate of return. Therefore, this formula is only used for past lotteries that already have completed results.

5. Conclusion

In conclusion, this paper highlights the importance of the expected rate of return as a framework for evaluating uncertain outcomes in complex situations, using the arithmetic average.

The comparison between the Meckin-Donnelly formula and the Abrams-Garibaldi formula reveals the different models of quantifying the expected rate of return for lotteries, incorporating real-world features such as jackpot sharing, pari-mutuel prizes, and sales of tickets, which significantly alter the values of expected rate of returns.

Furthermore, elements that relate to psychological behavior, such as the birthday effect, underline the main flaws in the formulas-assuming every number has the same probability of being chosen. This may impact the expected rate of return since frequently picked numbers have a lower rate of return. In addition, the advantage of choosing numbers above 30 may result in a 127% advantage.

Ultimately, although advanced models such as the Abrams-Garibaldi model may improve the accuracy and decrease the number of errors, significant limitations such as the birthday effect may restrict its usefulness, reinforcing the unpredictability of lotteries.

DOI: 10.54254/2753-8818/2026.DL34356

References

[1] User, S. (2025). New York lottery sales surpass $658 million in September. Public Gaming Research Institute. http: //www.publicgamingresearchinstitute.com/news-categories/lottery/14974-new-york-lottery-sales-surpass-658-million-in-september
[2] Brown, S., & Ziemba, W. T. (2022). Lotto as options: Measuring the silver lining effect. Multinational Finance Journal, 26, 27–59.
[3] Research and Markets. (2026). Online lottery - Market share analysis, industry trends and statistics, growth forecasts (2026–2031). https: //www.researchandmarkets.com/r/nyls9e
[4] Steiner, J., & Stewart, C. (2016). Perceiving prospects properly. American Economic Review, 106(7), 1601–1631.
[5] Abrams, A., & Garibaldi, S. (2025). Finding good bets in the lottery, and why you shouldn't take them. arXiv.
[6] Bodie, Z., Kane, A., & Marcus, A. J. (2011). Investments (9th ed.). McGraw-Hill.
[7] Anonymous. (2021). Maximizing the expected value of a lottery ticket: How to sell and when to buy. arXiv.
[8] Chen, L., Novy-Marx, R., & Zhang, L. (2006). The expected value premium (NBER Working Paper No. 12183). National Bureau of Economic Research. https: //doi.org/10.3386/w12183
[9] Lucky Picks. (2025). Lottery odds & expected value — How they work, why they matter. https: //luckypicks.io/lottery-odds/
[10] Mecklin, C., & Donnelly, R. (2005). The lottery paradox: Expected value and decision making. Journal of Statistics Education. http: //jse.amstat.org/v13n2/mecklin.html
[11] Matheson, V. A., Grote, K. R., & others. (2003). In search of a fair bet in the lottery.
[12] D'Hondt, C., De Winne, R., & others. (2024). Is there a gender gap in the birthday-number effect? The case of lotto players and the role of sequential choice. Journal of Gambling Studies, 40(3), 1439–1463.