

# The Lottery Gap: Unraveling Income-Driven Differences in Lottery Play Choices and Earnings

Paul Parker\*    Paulo Albuquerque†    Yakov Bart‡

September 26, 2023

## Abstract

The authors examine how play timing, game choice, and number selection affect consumer and retailer lottery earnings. While prior research indicates that poorer consumers spend a higher proportion of their income on lottery tickets, the authors quantify income-driven differences on when and how lottery games are played. Using market and transactional lottery data, they find that consumers in low-income areas prefer games with lower payouts and are less likely to switch between games with jackpot size changes. Furthermore, players in low-income areas are more likely to choose numbers manually and select popular number combinations. These preferences result in lower expected earnings: Players in low-income areas lose 10% more per ticket than those in high-income areas. Due to existing incentive structures, retailer earnings are also affected by these differences in earnings. The authors conclude by discussing public policy implications. To reduce the inequality between consumers in low- and high-income areas, lottery advertising and communications should include expected earnings information, retailer incentive structures should be closer aligned with consumer earnings, and lottery game formats should remove pari-mutuel prize pools.

## Keywords:

Lottery, Inequality, Retailers, Number choice, Financial decision-making

---

\* Assistant Professor of Marketing, The University of Kansas.

† Professor of Marketing, INSEAD.

‡ Associate Professor of Marketing, Northeastern University.

“Poor people should pick better lottery numbers” (*Don’t Look Up*)

Lottery play captures the attention not only of the public but also of policy makers for many reasons. For many consumers, the possibility of winning an instant fortune has enough appeal to generate repeat purchases of lottery tickets. At the same time, policy makers view lotteries as a source of additional revenue for funding public education and social programs (Grote and Matheson 2011; Matheson and Grote 2008). However, mounting evidence reveals the negative externalities created by lotteries (Coughlin and Garrett 2009; Feehan and Forrest 2007; Griffiths and Wood 2001; Laitner 1999; Layton and Worthington 1999; Livernois 1987; Lorenz 1990; Stranahan and Borg 2004). Notably, several prior studies have demonstrated that lotteries are a regressive tax: Poorer populations spend a relatively higher percentage of their income on lottery products (Borg and Mason 1988; Clotfelter and Cook 1987; Farrell and Walker 1999; Suits 1977). Moreover, lottery agencies do not know the full extent to which their customers are negatively affected, and further systematic study of the lottery and of the demographics of players is necessary (Miller 2019).

This research shows that consumers in low-income areas not only spend disproportionately more on lottery tickets, but also that they lose 10% more per ticket than those in high-income areas because of when and how they choose to play. We focus on identifying differences in lottery play decisions across income levels and evaluating how these differences affect expected returns. Our analysis is based on systematic characterization of differences in, and consequences of, multiple consumer decisions involved in lottery gambling. Lottery agencies tend to offer products with different odds and prize amounts that vary over time, and some game formats, particularly games with large jackpots, allow players to choose numbers manually or have them chosen randomly. Along these dimensions, heterogeneity in consumer choices of games and numbers may affect expected earnings. For instance, consumers choosing to play games with higher odds have, in expectation, a higher return than those choosing games with lower odds, given the same prize

distribution. Moreover, many large jackpot games are pari-mutuel, with prize pools split between the winners. This means that picking more commonly-chosen number combinations results in smaller shares of prizes and lower expected earnings (Cook and Clotfelter 1991).

Furthermore, these consumer decisions may also affect the retailer revenues generated from selling lottery products. This connection between consumer decisions and retailer revenues stems from the structure of lottery-based bonus payments. In addition to earning a base commission on ticket sales, retailers receive bonuses proportional to the prizes won by their customers. Consequently, players' choices and resulting expected earnings have a direct effect on retailers' bottom lines.

Our contribution is threefold. First, we examine income-based heterogeneity in game choice decisions (when to play and which lottery game to play), the tendency to use tools for random number selection, and the frequency of manually chosen number combinations. Prior research has mainly explored how much people spend on lotteries across different income levels (finding that lottery play is regressive), but few studies have examined how demographics interact with other, subsequent playing decisions (Guryan and Kearney 2008; Oster 2004). To our knowledge, we are the first to examine the resulting differences in expected earnings, per ticket purchased, across income levels. Second, we measure the impact of income-based heterogeneity in lottery play decisions on retailer earnings, which has not been considered in previous literature. Third, informed by our analysis, we discuss how marketing and advertising activities, as well as retailer commissions and bonuses, may affect lottery playing decisions and how changes in incentive structures can help reduce the inequality of lottery play outcomes across different income levels.

Using both aggregate-level and transaction-level lottery data, we quantify the impact of income on earnings for players and revenues for retailers by separately modeling each step in the consumer decision-making process: (1) when to play, (2) which game to play, (3) whether numbers are generated with a random selection tool, and (4) manual number choice. We find that players in low-income areas are negatively affected at every stage compared to those in high-income areas. Specifically, with regards to the choice of the lottery game, we show that people in low-income areas are less likely to switch between games when the jackpot increases. As a result, they are less

likely to choose games with higher expected payouts. They also tend to use the random number selection tool less frequently and are more likely to choose common combinations of numbers when selecting numbers manually. Given the pari-mutuel structure of the lottery games, both of these tendencies lead to lower expected earnings. Overall, our results show that players in low-income areas lose more than 10% more per lottery ticket than players in high-income areas.

Consumer decisions on when and how to play, along with the lottery's incentive system, also influence ticket retailer earnings. We find that retailers in low-income areas make significantly fewer sales and, in aggregate, earn 45% less than those in higher-income areas. In addition, because low-income customers choose games with lower expected payouts, retailers in low-income areas earn 0.2% less on average per ticket sold than retailers in higher-income areas. Finally, we find diminishing returns to increasing the number of retailers selling lottery tickets in a given location. In all, these results suggest that the inequality of lottery outcomes between high- and low-income areas, for both consumers and retailers, is significantly greater than previously identified in studies focusing solely on the relationship between income and consumer lottery play incidence.

Our findings have several public policy implications. First, lottery operators and retailers can decrease the inequality of lottery returns across income areas by modifying how they advertise and communicate game information to lottery customers. Although the odds of winning and jackpot size are readily available, the expected payout is not. This information has been shown to affect lottery consumers' purchasing decisions (Stearns and Borna 1995) and, as such, would help offset potential differences in expected earnings between consumers in low- and high-income areas. Second, retailers need to be aware of the consequences – for themselves and their customers – of manual versus random number selection, since retailers are directly affected by the prize amounts earned by consumers. Moreover, we recommend modifying retailer incentive structures to include bonuses across all prize tiers and uncapped jackpot prize bonuses, so as to motivate retailers to push consumers to play in a mutually beneficial manner. More broadly, if lottery operators were to switch from a pari-mutuel structure to one with fixed prizes, differences in expected earnings across combinations would be eliminated because all combinations would receive the same payout,

irrespective of the number of winners.

## **Literature Review**

Three distinct literature streams are relevant to our study on income-driven differences in lottery play decisions: measuring the impact of lotteries on society as a potentially regressive tax, choosing numbers, and lottery marketing and advertising. Table 1 compares the scope of our analysis with that of previous studies.

### ***Lottery Incidence and Game Choice Decisions***

Research has long examined socioeconomic characteristics such as levels of income, poverty, and education to understand their impact on lotteries. For example, research that has evaluated whether poorer consumers spend disproportionately more on lottery products than people who are well-off implies that lotteries function as a regressive tax (see the first set of studies listed in Table 1). Indeed, research almost unanimously concludes that lottery sales are regressive, finding income elasticities of demand less than 1 (Borg and Mason 1988; Carpenter, Perlman, and Norris 2010; Clotfelter and Cook 1987; Farrell and Walker 1999; Suits 1977). Moreover, some studies find evidence that lottery products can be inferior goods, with negative elasticities of demand, especially for those with the lowest range of income (Blalock, Just, and Simon 2007; Clotfelter 1979; Hansen, Miyazaki, and Sprott 2000). Extending to other types of lottery products, other studies have found that instant games, such as scratch cards, have lower income elasticities, while games with jackpot prizes attract richer consumers (Garrett and Coughlin 2009; Jackson 1994; Mikesell 1989), so much so as to render the lottery game progressive (Oster 2004). More broadly, research has found that the odds of winning a jackpot and jackpot size, rather than the expected value of a ticket, drive sales (Forrest, Simmons, and Chesters 2002; Garrett and Sobel 2004).

In examining lottery sales, studies have scrutinized demographic variables other than income. For example, level of education typically has an inverse relationship to lottery sales (Beckert and Lutter 2013; Clotfelter and Cook 1987; Kitchen and Powells 1991); men are more likely to

Table 1. Literature on lottery playing behavior

	Demographics		Randomness		Retailer Analysis
	Single Game	Across Games	Single Game	Across Games	
Suits (1977); Clotfelter (1979); Clotfelter and Cook (1987); Farrell and Walker (1999)					
Borg and Mason (1988); Hansen, Miyazaki, and Sprott (2000); Blalock, Just, and Simon (2007)	✓				
Clotfelter and Cook (1990); Kitchen and Powells (1991)					
Hersch and McDougall (1989); Oster (2004); Guryan and Kearney (2008)					
Mikesell (1989); Jackson (1994); Garrett and Coughlin (2009)	✓	✓			
Cox, Daniell, and Nicole (1998); Farrell et al. (2000); Baker and McHale (2009)			✓		
Clotfelter and Cook (1993); Wang et al. (2016)			✓		
Baker and McHale (2011)				✓	
Present study	✓	✓	✓	✓	✓

purchase tickets than women (Clotfelter and Cook 1987, 1990; Farrell and Walker 1999; Kitchen and Powells 1991); African Americans and those of Hispanic descent are more likely to play the lottery than white individuals (Jackson 1994; Scott and Garen 1994); residents' proximity to lottery retailers increases the numbers of tickets purchased (Hersch and McDougall 1989; Kitchen and Powells 1991); and residents in zip codes characterized by lower education rates, higher poverty levels, and older populations are significantly more likely to purchase lottery tickets from retailers that recently sold a winning ticket (Guryan and Kearney 2008).

Several reasons have been proposed for why individuals with lower incomes play lotteries. First, studies have shown that family members' and friends' play has a significant effect on lottery expenditures (Ariyabuddhiphongs 2006; Beckert and Lutter 2013; Coups, Haddock, and Webley 1998). Money consciousness also plays a role in lottery playing decisions (Ariyabuddhiphongs 2006). Wohl, Branscombe, and Lister (2014) find that participants reminded of poor economic prospects are more likely to gamble and with a greater amount. Low-income consumers could also have access to relatively fewer investment opportunities and could see lotteries as a viable means to improve their financial circumstances (Blalock, Just, and Simon 2007; Lee and Chang 2008). Finally, it has been suggested that the lottery could simply be a form of entertainment (Clotfelter and Cook 1991; Hirshleifer 1966; McCaffery 1994; Thaler and Ziemba 1988; Wu 1979). Moreover, Burger et al. (2020) find that lottery participation before the draw has a positive effect on happiness.

Although our aim is to identify any potential differences in expected earnings across income groups, not the reasons behind how people play the lottery, prior research provides a basis for what we expect to observe. Guryan and Kearney (2008) suggest that biases in lottery play that stem from income or education are consistent with the literature claiming that cognitive biases are mitigated by cognitive resources. Frederick (2005) and Benjamin, Brown, and Shapiro (2013) find that low cognitive ability is associated with impatience and risk aversion for small gambles. Specifically, individuals with higher cognitive ability are more likely to take higher expected value gambles versus lower certain gains than those with lower cognitive ability (Frederick 2005). Consistent with

this stream of literature, we expect consumers in higher education and income areas to make lottery purchasing and playing decisions that yield higher expected returns. We expect these consumers to purchase proportionally more tickets for draws with higher jackpots, to choose games with better expected payouts, and to exhibit less bias when choosing number combinations.

### ***Biased Number Selection***

A second stream of related research investigates the various biases exhibited in playing lotteries (Rogers 1998). In games with pari-mutuel formats (i.e., prizes are divided among multiple winners), playing lottery games in a biased way affects how much one stands to win. For example, selecting a number combination that other players are disproportionately likely to choose would lead to a large number of winners, should it be drawn as the winning combination. Prizes would then be split among a larger than average number of winners, which, when considering all possible combinations to choose from, would lead to lower expected earnings.

Research shows that the representativeness bias (i.e., the tendency to estimate the likelihood of an event on the basis of the essential features of its parent population) affects the selection of numbers in a draw game (Tversky and Kahneman 1974). In general, players perceive random sequences as more likely to win (Hardoon et al. 2001), and random-looking sequences appear more representative of winning combinations than numbers with repeat digits (Holtgraves and Skeel 1992). The representativeness heuristic also extends to the negative recency effect, also known as the “gambler’s fallacy,” or the belief that the probability of an event occurring is low after that event has recently occurred, even though this probability is independent from one period to the next. In a lottery, the negative recency effect can be observed after a particular number is drawn: the subsequent amount bet on that particular number falls sharply (Clotfelter and Cook 1993). This fallacy appears to be common: For example, the lottery.net website showing frequencies of numbers drawn in the past and recommending to pick numbers that “go months without being chosen” has been visited over half million times in July 2023 alone (Semrush 2023).

Lottery players exhibit other biases when selecting numbers. For example, they often choose

numbers toward the center of the lottery ticket and tend to spread numbers evenly across the possible range (Wang et al. 2016). They also often choose personally meaningful numbers such as birthdays, age, and postal code (Wang et al. 2016). In addition, players consider situationally available numbers from the environment - even remembering numbers to which they are briefly exposed (Kubovy 1977) , such as numbers inside fortune cookies (Lee 2005). They also have a tendency to choose numbers such as the date of play (day, month, or year), date of the draw, jackpot size, and remaining time until the draw (Wang et al. 2016). Such biases often lead to multiple jackpot winners. For example, 433 players won a lottery drawing in the Philippines in 2022, because they all picked the same winning combination of six numbers — 09-45-36-27-18-54 (multiples of nine) (Elemia and Ives 2022). Consistent with the literature on cognitive ability and the lottery (Benjamin, Brown, and Shapiro 2013; Frederick 2005; Guryan and Kearney 2008), we expect consumers with higher incomes to choose number combinations in a less biased manner than those with lower incomes, leading to higher expected earnings.

In general, most methodologies used to estimate the degree of bias and non-randomness in lottery number selection (see the second section in Table 1) rely on the observation that if players have a preference for specific individual numbers or number combinations, the number of winners will vary more than random number selection would predict. For example, if the number 7 is disproportionately chosen (over other numbers), there will be more winners when 7 is drawn as a winning number and, conversely, fewer winners if the number 7 is not drawn. Therefore, greater variance in the number of winners conditional on the number of tickets sold is an indicator of higher bias in number selection.

Several studies have proposed methods to estimate these number-selection biases. Cox, Daniell, and Nicole (1998) use a maximum entropy estimator to identify preferred combinations and find that players choose numbers toward the center of the ticket and that they can double their expected winnings by choosing unpopular number combinations. Farrell et al. (2000) model the preference for selecting the individual numbers themselves and find that prize rollovers occur more often than expected because of a nonuniform number choice. Baker and McHale (2009) construct an

empirical model of the joint distribution of prize winners to calculate the moments of ticket value as a function of sales. They confirm that lottery players can increase the expected value of their ticket by strategically choosing less popular numbers. Finally, Baker and McHale (2011) present a theoretical model in which players preferentially choose clusters of number combinations. Using sales data in combination with the number of winners across draws, they estimate a three-parameter model that better represents the correlations between the number of winners across prize tiers. We adapt this methodology to examine the differences in preferences for combination choices across areas of different income levels.

### ***Lottery Advertising***

Multiple studies have found that lottery-type mechanisms can be an effective tool in increasing consumer response rates. As incentives, single prize giveaways (vs. multiple cash prizes) increase consumer response rates and decrease the costs per completed survey (Balakrishnan et al. 1992), while promotions offered as low probability gambles (vs. a price discount) can lead to higher sales (Dhar, Gonzalez-Vallejo, and Soman 1995). Moreover, the literature hints at why lottery-related promotions have an incentive not to disclose full information to prospective consumers. Yan and Muthukrishnan (2014) find that presenting information about consolation prizes in promotional lotteries draws attention away from the large prize, resulting in lower consumer valuations of the promotion, and Goldsmith and Amir (2010) show that uncertain promotions are most effective for intuitive, low-cost decisions but may be negated by probability considerations and contexts that generate careful thinking. These effects are of particular concern because lottery-linked products are especially popular with low-income customers (Guillen and Tschoegl 2002). These customers find lotteries appealing in general because they regard them as an opportunity for recovery after sudden adversity, when no other options are available (Brenner 1986).

Relatedly, several studies have argued that lottery product advertising is deceptive and unethical. Landman and Petty (2000) find that lottery purchases induce imagining “what might have been,” something that lottery advertising exploits. Lottery advertising also conveys an “ethos of winning,”

by celebrating top prize winners, huge jackpots, and easy wins, that suggests certainty in winning despite the actual overwhelming chances of loss (McMullan and Miller 2009). This type of advertising is also particularly problematic for at-risk gamblers. The response to advertising is close to twice as high for addicted gamblers (Narayanan and Manchanda 2012), and lottery advertising has a greater impact on consumers with lower income and lower education. Specifically, lower-income consumers respond more favorably to lottery advertising and perceive more positive influences, such as entertainment value. Conversely, higher-income consumers are more likely to advocate for social responsibility in lottery advertising (Lee and Chang 2008). Finally, Stearns and Borna (1995) argue that lottery advertising is deceptive in that it promotes the lottery as a solution to financial problems and through the omission of objective information, such as the actual value of lottery tickets. They find that consumers adjust their lottery purchases accordingly when presented with expected value information before ticket purchase. Taken together, these findings are especially concerning since the reminder effect of lottery advertising has a large impact on boosting total sales (He and Klein 2022).

The current work builds on this literature by exploring how much different income groups stand to lose through the expected value of ticket purchases across lottery draws. Because lower-income consumers are more likely to respond to messages from lottery advertising and value the entertainment side of the lottery more than higher-income consumers (Lee and Chang 2008), we expect players in low-income areas to continue to play even when the jackpot is low and to select numbers manually (vs. using a machine that offers random selection of numbers), respectively. In addition, we expect to find a strong effect of the number of retailers in the area on lottery ticket sales, as retailers serve as local advertisements for the lottery and its jackpot.

## **Lottery Setting**

### ***Industry***

Our empirical application uses data from the lottery industry in California. As is typical for lotteries, California operates games as a monopoly through a state agency that determines the

Table 2. Draw Games Overview (2004-2006)

Game	Fantasy 5	Mega Millions	SuperLotto Plus
Mean jackpot size	\$108,000	\$66,200,000	\$25,700,000
Mean sales per draw	\$424,000	\$3,880,000	\$8,400,000
Total numbers	39	56	47
Number to choose	5	5	5
Total Mega numbers	—	46	27
Total combinations	575,575	175,711,536	41,416,353
Overall odds (any prize)	1 in 9	1 in 39	1 in 23
Draw frequency	Daily	Tuesdays and Fridays	Wednesdays and Saturdays

lottery products and their formats and sales rules. Independent retailers sell lottery tickets after they complete an application process, undergo an evaluation by the lottery agency, and install lottery hardware. Lottery retailers include a wide range of businesses, the most common being gas stations and convenience stores.

Retailers are incentivized to sell lottery tickets given three payment forms: commission, sales bonus, and cashing bonus. The flat-rate commission on each lottery ticket sold ranges from 4.5% to 6%. A sales bonus is earned if a retailer sells a jackpot-winning ticket for a draw, typically 0.5% of the value of the winning ticket. A cashing bonus is earned on winning tickets associated with smaller non-jackpot prize tiers, whereby, in an effort to distribute prizes quickly, retailers pay the winning ticket holders directly and are reimbursed plus a bonus of 3% (for prizes between \$99 and \$599) by the lottery agency. Figure 1 describes the incentive structure for lottery retailers.

We focus on draw games, as these give players the most freedom of choice, have the greatest potential for heterogeneity in playing behavior, and account for more than 90% of sales for non-scratch-card lottery products. Draw games operate as follows: A customer purchases a ticket and chooses a combination of  $n$  out of  $N$  numbers. At a scheduled time,  $n$  balls are drawn from the  $N$  possible numbers, with the numbers on those balls designating the winning combination. Matching all  $n$  balls correctly wins the jackpot (highest prize). Subsets of the winning numbers win smaller prizes (matching two balls wins the minimum prize).

Figure 1. Retailer incentives



## California Lottery Retailer Incentives

**COMMISSIONS**

*Draw Games Commissions:*

Altura Terminal .....	6.0%
Quick Terminal .....	6.0%
Self-Serve Terminal .....	4-5%
GamePoint .....	4-5%
Gemini .....	4-5%
Fantasy 5 Replays Handling Fee .....	6.0%

*Scratchers\* Commissions:*

Scratchers* .....	6.0%
Free Scratchers* Handling Fee .....	6.0%

**DRAW GAMES SALES BONUSES**

Retailers earn one half of one percent for selling winning tickets for the following Draw game prizes:

- Powerball\* ..... \$1 Million Prize or more (regardless of the payment option selected, up to a Maximum bonus of \$1 Million)
- Mega Millions\* ..... \$1 Million Prize or more (regardless of the payment option selected, up to a Maximum bonus of \$1 Million)
- SuperLotto Plus\* ..... Jackpot winner
- Fantasy 5 ..... 5/5 Top Prize winner
- Daily Derby\* ..... Grand Prize winner

*As an example, if you sell a \$12 Million SuperLotto Plus\* or Mega Millions\* jackpot winning ticket, you will earn a selling bonus of \$60,000!*

**SCRATCHERS\* SALES BONUSES**

Instant Scratchers\* top prizes of \$1 Million or more earn a bonus of one half of one percent.

Additionally, Scratchers\*, in connection with a promotion\* resulting in a promotional award of \$1 Million or more, also earn a one half of one percent bonus.

\* See Official Promotion Rules for eligibility details.

**CASHING BONUSES**

*Draw Games Cashing Bonus:*

**\$99-\$599 Winners .....** 3.0%

Cash winning draw game plays worth \$99 to \$599 and you earn a three percent cashing bonus!

*As an example, cash a \$500 winning ticket for Daily 3 and earn a cashing bonus of \$15!*

*Cashing draw games can be profitable and allows for multiple-line prize payments of more than \$599. For example, cash a Fantasy 5 ticket where two lines win a prize. Line A wins \$200 and Line B wins \$500. This adds up to \$700 in "payable at retail" prizes because it is two separate prizes. The terminal will display the total prize amount and ask if you can pay. To cash the ticket, press "SEND" and pay the player \$700. You'll earn a bonus of \$21.*

**Note: If one of the winning combinations on a multiple-line winning ticket exceeds \$599, the terminal will display "Exceeds Cash Limit, File Claim Form" and a claim receipt will automatically be printed out.**

*Scratchers\* Cashing Bonus:*

**\$1-\$599 Winners .....** 1.0%

Cash any winning Scratchers\* ticket from \$1 to \$599 and you earn one percent of the prize value!

*As an example, cash a \$500 winner and earn a cashing bonus of \$5!*

(Updated December 2015)

Electronic copy available at: <https://ssrn.com/abstract=3660130>

All prizes for draw games in California are pari-mutuel (i.e. prize pools are equally divided if there are multiple winners). During the data period (2004-2006), there were three such games, all priced at \$1 per ticket: Fantasy 5, Mega Millions (introduced to California in mid-2005), and SuperLotto Plus. The lottery market is relatively stable, with only Mega Millions changing its pricing and odds structure after our study period; both Fantasy 5 and SuperLotto Plus have remained the same. The purchase process for lottery tickets has also remained similar. California lottery products can only be purchased through authorized retailers, with app-based and online lottery ticket resellers being expressly forbidden in the state (California Lottery 2023).

All three offer a “Quick Pick” combination generator, a feature that selects numbers at random. The latter two games require a “mega” number to be selected in addition to  $n$  regular numbers. As Table 2 shows, these games require the same number of numbers to be chosen (five) but have varying odds and prize pools. The games with lower odds typically have larger prize pools. Note that Fantasy 5 and SuperLotto Plus are only available in California, while Mega Millions is played across multiple states simultaneously. Table 3 breaks down the odds, prize amounts, and number of winners across all the prize tiers for each of the three game formats.

## ***Data***

We use four different data sets, three of which were provided by the California Lottery Agency, in our analysis: an aggregate market-level data set, a transactional level data set, a validations data set, and lottery game statistics. The market-level data consist of daily ticket sales for all lottery retailers in California (25,134) across 12 weeks, one week per quarter, with selection timing being consistent across three years (2004-2006).

We break down sales per retailer in Table 4. In this data set, we use a trade code (eg., gas station, supermarket, bar) to categorize retailers, with latitudinal and longitudinal information. We merged this geolocation with census tract-level<sup>1</sup> information such as income per capita, education

---

<sup>1</sup>Census tracts are small geographic subdivisions of a county, determined by the U.S. Census, designed to represent relatively homogeneous groups of people in terms of population characteristics. In our data set, the average census tract covers 5,600 people.

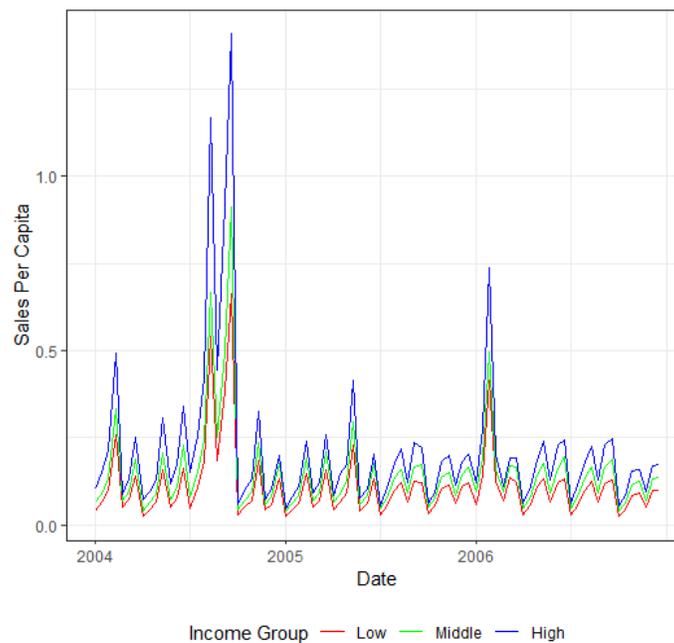
Table 3. Prize Tier Statistics (2004-2006)

	Fantasy 5			Mega Millions			SuperLotto Plus		
	Odds 1 in	Mean Prize	# Winners	Odds 1 in	Mean Prize	# Winners	Odds 1 in	Mean Prize	# Winners
Match 5 of 5 + Mega	575,757	\$91,950	.840	175,711,536	\$66.17M	0.025	41,416,353	\$25.29M	0.165
Match 5 of 5				3,904,701	\$306,700	0.925	1,592,937	\$29,960	5.505
Match 4 of 5 + Mega				689,065	\$13,820	5.688	197,221	\$1,544	42.677
Match 4 of 5	3,387	\$430.63	138.8	15,312	\$159.26	247.8	7,585	\$95.74	1,112
Match 3 of 5 + Mega				13,781	\$163.11	279.7	4,810	\$53.84	1,759
Match 3 of 5	103	\$15.88	4,560	306	\$7.08	12,570	185	\$9.72	45,620
Match 2 of 5 + Mega				844	\$10.05	4,541	361	\$10.32	23,430
Match 2 of 5	10	Free Play	48,630						
Match 1 of 5 + Mega				141	\$2.87	27,350	74	\$1.82	114,200
Match Mega				75	\$1.83	51,620	49	\$1	173,500

Table 4. Lottery Ticket Sales Across Retailers

Game	Aggregate Data (25,134 Retailers)	Transactional Data (65 Retailers)	
	Mean Sales Per Draw Per Retailer	Mean Sales Per Draw Per Retailer	Mean Transactions Per Draw Per Retailer
Fantasy 5	\$23.614	\$17.384	5.203
Mega Millions	\$63.548	\$144.497	45.865
SuperLotto Plus	\$160.970	\$313.893	91.317

Figure 2. Average Lottery Sales Per Capita (25,134 Retailers)



level, and other demographic variables. Retailers fall within 5,807 census tracts.

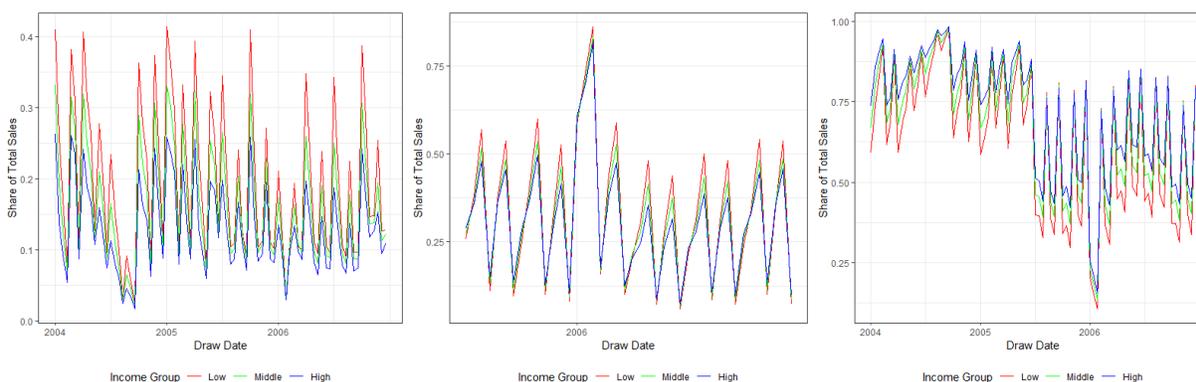
Figure 2 shows average daily sales per capita across three income groups<sup>2</sup> formed from the 5,807 census tracts. The data show that lottery ticket sales generally increase with income. Figure 3 shows the relative share of sales for each game across the same three income groups. The low-income group spends proportionately more on Fantasy 5, while the high-income group spends proportionately more on SuperLotto Plus.<sup>3</sup>

The transaction-level data contain sales data for 65 California retailers for three full years

<sup>2</sup>We create income groups using income per capita. The first cutoff, \$11,150, puts 12.8% (the proportion of Californians under the poverty line) of locations in the first group. The second cutoff, \$22,100, is the median income per capita across locations.

<sup>3</sup>We assume that players live in the census tract where the lottery transactions occur.

Figure 3. Average Share of Game Choice (25,134 Retailers)



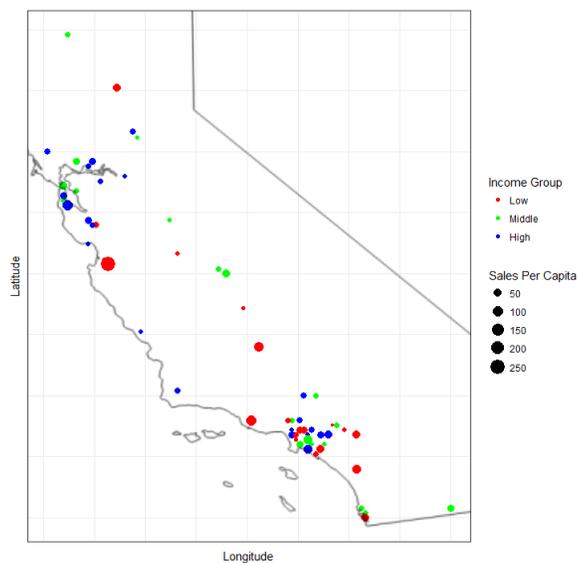
From left to right: Fantasy 5, Mega Millions, and SuperLotto Plus

(2004, 2005, and 2006), as summarized in Table 4. For the average retailer, SuperLotto Plus is the most popular game in terms of sales, followed by Mega Millions and then Fantasy 5. Per transaction, for each of the three games, a customer purchases an average of three lottery tickets. Figure 4 shows the location of the 65 retailers California present in the transaction-level data, along with each location's corresponding sales, over the three year span of the data, and income per capita. No two retailers are located in the same tract code, making demographic information unique.

Each entry in the transactional data set corresponds to an individual purchase and contains the following information: the time stamp of the purchase, the number of tickets purchased, the game type, and whether or not the player used the random selection feature (the Quick-Pick option). Figure 5 shows the differences in the choice of Quick-Pick between income groups. The high-income group of locations utilizes the Quick-Pick option the most frequently.

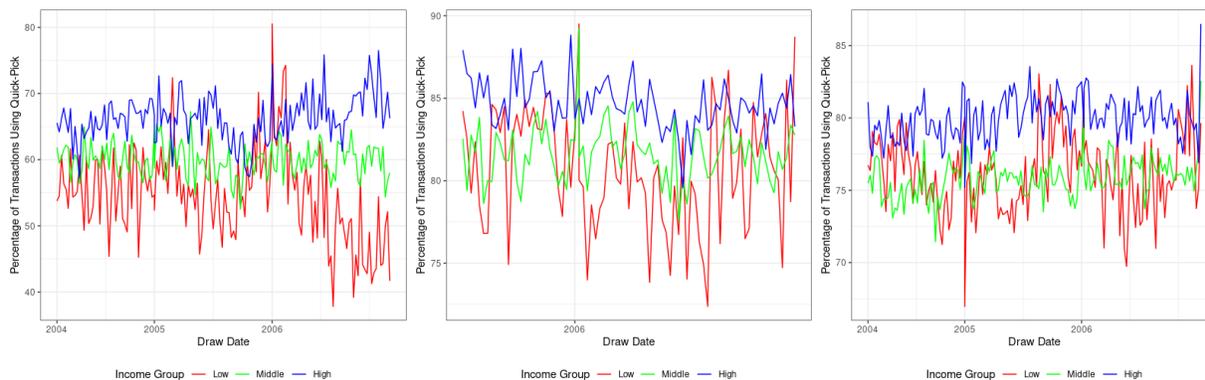
For smaller prizes, when customers win, they claim the prize from the retailer where they purchased the ticket. The retailer is then reimbursed by the lottery and given a bonus based on the prize amount. The validations data include the time stamp of the customer submitting a winning ticket to the retailer and the prize amount of the ticket. These data give us the number of prize winners per location, per draw, across the 65 locations. The variation in the number of winners included in this data set, across the draw dates, provides descriptive evidence for heterogeneity in

Figure 4. Retailer Locations and Sales (65 Retailers)



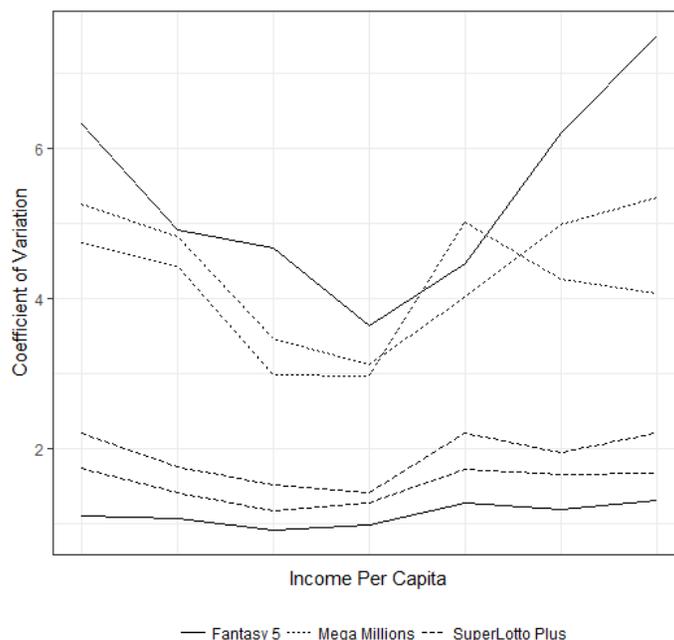
Note: We split locations into groups based on income per capita.

Figure 5. Average Quick-Pick Choice Proportion (65 Retailers)



From left to right: Fantasy 5, Mega Millions, and SuperLotto Plus

Figure 6. Coefficient of Variation for the Number of Winners (65 Retailers)



For a given lottery game, each line represents a prize tier from the validations data. The number of winners is conditional on the number of tickets sold.

number selection. Greater variance in the number of winners, conditional on the number of tickets sold, would indicate that customers prefer choosing particular combinations. This, in turn, would lead to lower expected earnings. Figure 6 shows, across groups of income per capita, the coefficient of variation of the number of winners for each of the prize tiers in the data for the three draw games. The validations data contain two prize tiers for each game.<sup>4</sup> The pattern across income groups is mostly U shaped, indicating that players in low- and high-income areas are more likely to choose popular combinations when manually selecting numbers. This information is necessary for the analysis on nonrandom number choices.

Finally, we collected lottery game summary statistics made public by the California Lottery Agency. For each lottery game, this includes the total number of winners, prize amounts, and odds by prize tier aggregated across all California lottery retailers. Overall, it is the combination of these different data sets enables us to shed greater light on the impact of lottery sales on both consumers

<sup>4</sup>Although larger prizes are not included in the validations data, the results with their inclusion are not substantively different, given how infrequently they are won.

and retailers.

## **Modeling Consumer Choices**

In this section, we propose an approach to understand the series of consumer decisions involved in playing the lottery: (1) whether to play and how many tickets to purchase; (2) which game to play; (3) whether to select numbers randomly or nonrandomly; and (4) if the latter, the degree of nonrandomness in the numbers chosen. Lottery play decisions are based on factors such as format and jackpot size, and we examine the roles of income and other demographic characteristics at each decision stage.

We model these different stages independently because of the structure of our data,<sup>5</sup> though some explanatory variables will be the same across stages.<sup>6</sup> We examine the first two stages, total ticket purchases and game choice, using the aggregate market-level data, and we use the transactional data set to examine the last two stages, random selection and manual selection. While the transactional data cover all draws over three years, the aggregate market-level data only contain 12 weeks of draws during that time span. Moreover, the transactional data only cover 65 lottery retailers, while the market-level data include all retailers (25,134). A combined data set, therefore, would only cover 12 weeks of draws for 65 retailers. Consequently, a more parsimonious model, across all stages, using both data sets simultaneously would be restricted by a substantially reduced sample size, leading to identification issues.

### ***Decision to Play Lottery Games***

We first examine the decision to play lottery draw games. As previously mentioned, related work suggests that lottery purchases are regressive: As income increases, lottery purchases increase, but lower-income consumers spend a higher proportion of their total budget on lottery tickets. This

---

<sup>5</sup>To assess correlations between stages, we computed the correlation in the errors obtained between the first and second stages. The resulting correlation is not significantly different from zero.

<sup>6</sup>The decisions in the later stages are sequential, so we include the proportion of players who randomly chose their numbers as part of the final stage that evaluates the bias of combination choices. This process adds correlation to our formulation.

means that though poorer people purchase fewer tickets, they are disproportionately affected by the negative effects of the lottery, given that the expected winnings from a lottery ticket are almost always less than the ticket price.

We model the decision to play lotteries using the following specification:

$$\log(S_{zt}) = \beta_0 + \beta_1 \log(K_z) + \beta_2 \log(J_t) + \beta_3 [\log(K_z) \times \log(J_t)] + \beta_4 X_z + \theta_t + \vartheta_t + \pi_t + \varepsilon_{zt}, \quad (1)$$

where  $z$  indexes location<sup>7</sup> and  $t$  time in days. The term  $S_{zt}$  denotes sales per capita at location  $z$  and time  $t$ ,  $K_z$  represents income per capita at location  $z$ , and  $J_t$  represents the log maximum daily jackpot prize across all available draw games. We use the maximum jackpot prize as a summary statistic of the attractiveness of playing the lottery on each day  $t$ .<sup>8</sup> The coefficient  $\beta_1$  is the main parameter of interest to determine the impact of income on the decision to play lottery and, together with the log-log relationship between sales and income per capita, represents income elasticity of demand. The coefficient  $\beta_2$  captures the demand generated by the changes in the size of the jackpot across days. Finally,  $\beta_3$  represents the differential effects of jackpot size on sales across income groups; a positive value for  $\beta_3$  would indicate that, as the jackpot goes up, sales per capita in high-income areas increase more than sales in low-income areas.

The term  $X_z$  represents a vector of other variables for each location  $z$ : the log number of lottery retailers per capita in the census tract, the percentage of the population over 65 years of age, the percentage with at least a college degree, and the fraction of African Americans and Hispanics in the population. We include the number of retailers at location  $z$  to account for variation in exposure to lottery products created by a greater number of stores. Moreover, we include an interaction term between the number of retailers in a location and income per capita to examine whether lower income areas are disproportionately affected by the presence of more lottery outlets. The chosen demographic variables are based on previous literature (Miyazaki, Hansen, and Spratt 1998; Oster 2004) and allow us to rule out possible alternative explanations for lottery regressivity – namely,

---

<sup>7</sup>Locations correspond to census tracts.

<sup>8</sup>Using alternative summary statistics, such as the sum of all jackpots, yields substantively similar results.

that it is driven by specific groups of people. Finally,  $\beta_0$  is an intercept,  $\theta_t$  are fixed effects for day of the week and year, and  $\vartheta_t$  is an indicator denoting whether the Mega Millions game was available to play at time  $t$ . To account for discontinuous jumps in sales when there were large jackpots,  $\pi_t$  is an indicator denoting whether the maximum jackpot for day  $t$  was greater than \$200M.

### ***Game Choice***

We then examine the choice of which lottery draw game to play. Using a market share model, conditional on ticket purchase (i.e., no outside good option, as we modeled the decision to play the lottery previously), we assess choices between lottery products as a function of time, game, and demographic variables. We assume that the utility that a consumer  $i$  in location  $z$  derives from game  $j$  on day  $t$  is as follows:

$$u_{izjt} = \delta_{zjt} + \varepsilon_{izjt}, \text{ with} \quad (2)$$

$$\delta_{zjt} = \delta_1 \log(J_{jt}) + \delta_2 [\log(K_z) \times \log(J_{jt})] + \delta_3 X_{zjt}. \quad (3)$$

In Equation 2, the term  $\delta_{zjt}$  represents the deterministic part of the consumer utility, and  $\varepsilon_{izjt}$  is a stochastic term, assumed to have a Type I extreme value distribution. In equation 3, the deterministic part of the utility is a function of jackpot size  $J_{jt}$  and the interaction between jackpot size  $J_{jt}$  and income  $K_z$ , to capture variation in preferences by different demographics for switching between games when the jackpot reaches different levels. The term  $X_{zjt}$  contains observed game characteristics: game dummy variables; an indicator for whether the jackpot is greater than \$200M; game-income intercepts, to account for heterogeneous preferences in game formats across income groups (Garrett and Coughlin 2009; Jackson 1994; Mikesell 1989); and an indicator variable representing whether game  $j$ 's draw occurs on day  $t$ . The price does not enter the utility function because all three games were are priced at \$1 per ticket for the entire data analysis period. A significant and positive value for  $\delta_2$  would signify that higher income areas have greater increases in sales with increasing jackpot sizes; that is, they are more responsive to fluctuations in

the size of the jackpot. This explanation expands on the single-product findings from Oster (2004) by considering choices across different lottery products.

Given the distribution of  $\varepsilon_{izjt}$ , we can write the market share of game  $j$  in location  $z$  at time  $t$  as

$$s_{zjt} = \frac{\exp(\delta_{zjt})}{\sum_{k=1}^J \exp(\delta_{zkt})}. \quad (4)$$

We set Fantasy 5 as the baseline alternative,  $j = 1$ , and then invert the shares to obtain the difference in mean utilities

$$\ln(s_{zjt}) - \ln(s_{z1t}) = \delta_{zjt} - \delta_{z1t}, \quad (5)$$

for all games  $j \neq 1$ . As such, the dependent variable for our analysis is the difference between the log of the daily shares of each game and the log of the daily shares of Fantasy 5.

### ***Choice of Using Random Numbers***

After purchasing a ticket for a specific draw game, consumers choose between picking lottery numbers themselves or having numbers randomly selected by a lottery terminal, a feature called Quick-Pick. Prior research shows that the choice of number combination has a significant impact on the expected earnings of the ticket (Cox, Daniell, and Nicole 1998). A manual combination choice tends to generate similar sets of numbers (Wang et al. 2016), which leads to more winners should a popular combination be drawn. With more winners, prize pools must be split among a greater number of players, leading to lower expected earnings. Using the random selection feature improves expected earnings by increasing the chances of selecting less popular number combinations.

Although the optimal strategy would be to ponder which combinations other players might prefer and manually choose different numbers, we ignore this possibility in the subsequent analysis and assume that random selection is a superior alternative for two reasons. First, other than using simple rules of thumb, such as avoiding the so-called lucky number seven, deducing which combinations to avoid is difficult for players. Second, by construction, only a small percentage

of consumers would behave this way, since if most consumers were attempting to avoid others' selections, they would prefer certain combinations (again leading to lower expected earnings).

We model the choice to use Quick-Pick as a function of game and demographic variables using the following logit specification:

$$QP_{izjt} = \lambda_0 + \lambda_1 \log(K_z) + \lambda_2 \log(J_{jt}) + \lambda_3 X_{izjt} + e_{izjt}, \quad (6)$$

where  $i$  indexes an individual transaction for game  $j$  on day  $t$  at location  $z$ . The dependent variable  $QP_{izjt}$  represents a binary indicator of whether an individual transaction involved the use of the random Quick-Pick feature (i.e., whether the number combination on the ticket was generated randomly). The coefficients  $\lambda_1$  and  $\lambda_2$  capture the effect of income per capita and the size of the jackpot on the propensity to use Quick-Pick respectively.<sup>9</sup> A positive value for  $\lambda_1$  would signify that consumers in higher-income areas use the Quick-Pick option more frequently. The control variables  $X_{izjt}$  include similar location-specific demographic information as in the previous stages (i.e. percentage of the population over 65 years of age and the percentage of African Americans, and Hispanics) along with the total number of tickets purchased in the transaction and an indicator representing whether the transaction occurred on game  $j$ 's draw day.

### ***NonRandomness in Number Combinations***

Finally, we explain the variation in manual number selection behavior across income levels by evaluating the level of nonrandomness in number choice. Previous research offers multiple methods for estimating the level of nonrandomness given the number of winners across draws. We use Baker and McHale (2011)'s method because it better represents correlations among the number of winners across prize tiers. This approach estimates three parameters:  $f$ ,  $\lambda$ , and  $\gamma$ . The frequency of random choice,  $f$ , represents the proportion of consumers who choose their numbers perfectly randomly, which we assume to be consumers who use the Quick Pick option. For other

---

<sup>9</sup>We also estimated the model with the interaction between jackpot and income in this stage. Only two of the nine interaction coefficients are significantly different from zero, and the overall results stay substantively equivalent.

consumers,  $\lambda$  represents the degree to which number combinations are chosen nonrandomly; as  $\lambda$  increases, certain groups of combinations are increasingly likely to be chosen. We use the clustering coefficient,  $\gamma$ , to group similarly preferred combinations; a larger value of  $\gamma$  signifies that preferences for combinations are more heterogeneous. Without accounting for this clustering, the correlation between prize tiers would be overestimated.<sup>10</sup>

Using data on sales and number of winners across draws, the approach employs the following likelihood function to estimate the probability of observing the realized number of winners in each prize bracket  $b$ , for game  $j$ , on draw day  $t$ , at location  $z$ :

$$L_{zjt} = \frac{1}{\bar{N}_j} \sum_{n=1}^{\bar{N}_j} \prod_{b=1}^{B_j} (p_{nbzjt} S_{zjt})^{N_{bzjt}} \frac{\exp(-p_{nbzjt} S_{zjt})}{N_{bzjt}!}, \quad (7)$$

with the log-likelihood to be maximized, for a given location  $z$  and game  $j$ , being the sum of the log of all draws,

$$L_{zj} = \sum_{t=1}^T \log(L_{zjt}). \quad (8)$$

In Equation 7,  $\bar{N}_j$  is the total number of available combinations to choose from for game  $j$ ,  $B_j$  is the number of prize brackets for game  $j$ ,  $N_{bzjt}$  is the number of winners for prize bracket  $b$  at location  $z$  for game  $j$  for draw  $t$  in the data, and  $S_{zjt}$  is the total sales for the draw. The term  $p_{nbzjt}$  gives the probability that a player chooses a combination that wins the  $b^{th}$  prize when  $n$  is the winning combination. The magnitude of  $p_{nbzjt}$ , therefore, depends on the prize bracket and thus may contain multiple probabilities. For example, for the top prize tier, in which all the numbers in the combination are matched,  $p_{nbzjt}$  is the probability that the player chooses the winning combination  $n$ . For the next prize tier, matching four out of the five numbers, the denominator of  $p_{nbzjt}$  includes all combinations that share four out of the five winning numbers. The same pattern applies to the other prize tiers.

The probability of a player at location  $z$  choosing a combination  $c$  for game  $j$  for draw  $t$ ,  $p_{czjt}$ ,

---

<sup>10</sup>For additional details, see Baker and McHale (2011).

is given by

$$p_{czjt} = \frac{f_{zjt} + (1 - f_{zjt}) [\lambda_z \exp\{-\lambda_z(c-1)/\bar{N}_j\}] / \{1 - \exp(-\lambda_z)\}}{\bar{N}_j}, \quad (9)$$

where  $c$  indexes the combination,<sup>11</sup>  $f_{zjt}$  is the proportion of players whose numbers are selected randomly, and  $\lambda_z$  is the level of nonrandomness exhibited in choosing numbers manually for location  $z$ . This formulation models the probability of choosing any given combination as partly random, representing consumers whose numbers are selected randomly (in this case using Quick-Pick), and partly nonrandom as an exponential distribution with parameter  $\lambda$  that captures the degree to which consumers prefer certain combinations.

For our application, we modify Baker and McHale (2011)'s method as follows: Given that our transaction-level data include the percentage of total tickets purchased that use the Quick-Pick tool, we supply the observed  $f_{zjt}$  to the likelihood function without estimating it. As such, we use the Quick-Pick frequencies to derive a unique  $p_{czjt}$  probability distribution for each draw in the data set. Doing so enables us to capture the manual nonrandomness in number selection appropriately, given the draw-to-draw fluctuation in Quick-Pick usage, by estimating  $\lambda_z$  for each location  $z$ . We then compare these estimates across locations and income per capita values.

### ***Parameter Identification***

For the first stage, the decision to play, we can determine the coefficient for income per capita through the variance in income across the 5,807 census tracts. The same applies for the other demographic variables. We identify the sales response to jackpot size through the changes in the maximum jackpot size over the time span of the data set. Identification for the second stage, the choice of which game to play, follows a similar pattern: Variation across census tracts allows us to identify the coefficients associated with income per capita, while changes in jackpot sizes (daily for Fantasy 5 and semi-weekly for Mega Millions and SuperLotto Plus) across the three games allows

---

<sup>11</sup>We order the combinations lexicographically (an increasing order based on the numbers within the combination). We model the correlations between winners across prize brackets by separating the indexes of the combinations into  $\gamma$  equal length groups and then interchanging even-numbered groups.

us to identify the coefficients tied to jackpots. The introduction of the Mega Millions game during our data period enables additional identification of the impact of jackpot size on game choice. The third stage, the Quick-Pick decision, follows the same identification strategy for income and jackpot-related variables, albeit across 65 locations, given the structure of the transaction-level data set.

In the fourth stage, we identify the degree of nonrandomness in the distribution of combinations through the variation in the number of winners across draws. For a winning combination, if there are a large (small) number of winners in a lottery location for a given sales volume, then there is a preference (dislike) for those numbers. Greater variation in the number of winners across draws indicates that players have stronger preferences for certain combinations and therefore are more nonrandom. Completely random selection would lead to the same number of winners across draws, given constant sales.

### ***Player and Retailer Earnings Simulation***

We use the estimated parameters from the previous sections to disaggregate the impact of each of these decisions on player and retailer earnings through a series of simulations and to measure how the expected earnings for all three games vary across differing income levels. The simulation consists of 5,000 iterations, with each iteration representing a week of sales for each of the three games.

The simulation is comprised of a series of lottery draws with sales, game shares, Quick-Pick decisions, and manual selection corresponding to the estimated models from the previous sections. For one iteration of the simulation, we generate a week of sales for each location in the market-level data set using parameters drawn from the sales model. Jackpot sizes, and lower tier prizes, are also sampled from the prize statistics in the data. We calculate the subsequent stages, game shares and Quick-Pick frequencies, in a similar manner using the same set of demographic and jackpot variables and sampling from the estimated parameters of the relevant model. Finally, we sample the estimated coefficients for the nonrandomness in number selection across three income

groups and use them to compute the combination selection distributions.

For each jackpot drawing of that week, a winning number combination is drawn and a set of combinations is generated for each income group, a fraction being Quick-Pick. After all the combinations for the tickets have been generated, we compare them with the winning combination generated for that draw. We tabulate the number of winning tickets per prize tier (five out of five matching numbers, four out of five, and so on) and divide the prize pools for that draw in a pari-mutuel manner.<sup>12</sup> Given that the odds of winning a prize are the same for each combination, we use the winning amounts to compute the expected earnings per ticket across the low-, middle-, and high-income groups and the total population. We compute retailer earnings in a similar manner. Retailers earn a base commission depending on the number of tickets sold, and cashing bonuses are awarded for winning tickets that fall within the thresholds outlined in Figure 1.

## Results

To check for robustness, we model the relationship between income and consumer choice using two specifications for each decision stage. The first treats income per capita as a continuous variable, with variation across census tracts helping identify its impact. The second specification uses an indicator variable for income per capita representing the income group of each location. Due to the limited draws and the few observable winners per jackpot draw in our data, we can only estimate parameters for a limited set of income groups during the final decision phase (the non-randomness in choosing combinations) - we assign locations to three groups as defined in the data section. The simulations leverage the results from all four stages across these three income groups to measure differences in expected earnings.

---

<sup>12</sup>For Fantasy 5, we compute the expected earnings of the lowest prize tier, a free play for matching two numbers, using the expected earnings of the current draw.

## *Decision to Play Lottery Games*

Using the aggregate-level data set, we test the regressivity of lottery sales by estimating Equation 1 using a log-log regression.<sup>13</sup> Table 5 reports the results from the continuous and discrete specifications. We cluster standard errors by location because demographic information is location-specific, while lottery sales are at the location-day level.

The results across these formulations are consistent. In column (1), the log-log specification indicates that the estimated coefficient on the income per capita variable represents the income elasticity of demand. We estimate this coefficient to be greater than 0 and less than 1. This suggests that sales increase with income, but not proportionately so, with consumers in lower-income locations spending a larger fraction of their income on lottery products. The discrete specification in column (2) shows similar results. The coefficient on the interaction between jackpot size and income per capita is positive and significant, indicating that consumers in higher-income areas buy more lottery tickets in response to increases in jackpot size than consumers in lower-income areas. This is consistent with the behavior Oster (2004) observes.

Given the log-log specification, the coefficient on the number of retailers per capita represents the retailer elasticity of demand. The coefficient is significant and positive but less than 1 (inelastic), indicating that sales per capita increase with a greater number of lottery retailers but not proportionally so. Thus, there are diminishing returns to an increasing number of retailers selling lottery tickets in a given location. The coefficients for the control variables in Table 5 indicate that large jackpots increase total sales and that, in our twelve week sample, sales per capita were higher before the introduction of Mega Millions. Moreover, locations with older and more educated populations tend to have higher sales per capita. In all, our results are consistent with prior research in that though sales increase with income, low-income players spend disproportionately more on lottery products (Borg and Mason 1988; Clotfelter and Cook 1987; Farrell and Walker 1999; Suits 1977) and high-income players are more responsive to higher jackpot sizes (Oster 2004). However, we

---

<sup>13</sup> Alternative formulations such as quasi-Poisson and negative binomial regressions yielded similar parameter and elasticity estimates.

Table 5. Log-Log Regression of Lottery Sales on Income Per Capita

Dependent Variable: log(Sales Per Capita)		
	(1)	(2)
log(Income per capita)	.266*** (.043)	
Middle-income group		.301*** (.032)
High-income Group		.415*** (.045)
log(Maximum jackpot)	.369*** (.001)	.334*** (.003)
log(Maximum jackpot) × log(Income per capita)	.041*** (.002)	
log(Maximum jackpot) × Middle-income group		.021*** (.003)
log(Maximum jackpot) × High-income group		.054*** (.003)
log(Number of retailers per capita)	.875*** (.012)	.922*** (.0339)
log(Number of retailers per capita) × log(Income per capita)	-.016 (.019)	
log(Number of retailers per capita) × Middle-income group		-.043 (.039)
log(Number of retailers per capita) × High-income group		-.070 (.038)
Large jackpot (>\$200M)	.354*** (.005)	.354*** (.005)
Mega Millions available	-.364*** (.003)	-.364*** (.003)
% age 65+	.006*** (.002)	.007*** (.002)
% college degree	.002* (.001)	.006*** (.001)
% African American	.0002 (.001)	.0002 (.001)
% Hispanic	-.002** (.001)	-.001 (.001)
Constant	-3.483*** (.046)	-3.940*** (.061)
Fixed effects	Year, Day of the week	Year, Day of the week
Clusters	Location	Location
Adj. $R^2$	.567	.570
Number of observations	475,713	475,713

We demean the variables log(Income per capita), log(Maximum jackpot) and log(Number of retailers per capita). Given the interaction term between the two, we interpret the coefficient for log(Income per capita) as the effect of income on the dependent variable when log(Maximum jackpot) is at its mean. We interpret the coefficient for log(Maximum jackpot) and log(Number of retailers per capita) in the same manner. Robust standard errors are in parentheses. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001.

add to this research by detailing the relationship between the number of lottery retailers and lottery ticket sales.

### ***Game Choice***

Table 6 reports the results for the logit model on game choice. We cluster the standard errors by location and draw because demographic and jackpot information are location and game-specific respectively, while game shares are at the location-game-day level.

First, the game-income fixed effects show that higher-income areas have a preference for SuperLotto Plus, a game with higher jackpots and lower odds than the baseline alternative, Fantasy 5. This is consistent with the findings of Garrett and Coughlin (2009); Jackson (1994); Mikesell (1989), in which higher-income consumers have a preference for high-jackpot games. However, Table 6 builds on these findings by showing that higher-income consumers are more likely to move across games with changes in jackpot size. Across both the continuous and discrete specifications, the interaction between jackpot size and income per capita is positive and significant, indicating that higher-income areas derive more utility from higher jackpots than lower-income areas, and, as such, are more willing to switch games to chase the jackpot. Finally, the indicator variables for the game draw days show that each game is more popular on its own draw day.

### ***Choice of Using the Random Number Generation Tool***

Table 7 shows the results from the logistic specification when the outcome variable is the decision to randomly select number combinations (by transaction). We cluster standard errors by location because demographic information is location-specific, while Quick-Pick decisions are at the transaction-location-game-day level. For both specifications, continuous and discrete income definitions, the results show a positive coefficient of income per capita on the decision to randomly select numbers, indicating that high-income areas are more likely to use the Quick-Pick option. In addition, the coefficient on the number of tickets purchased in each transaction is negative and significant, implying that people are more likely to choose their numbers manually when they buy more

Table 6. Game choice logit share model results

Dependent Variable: Difference of Log Shares		
	(1)	(2)
Mega Millions	-1.168** (.423)	-1.160** (.468)
SuperLotto Plus	1.133*** (.116)	.977*** (.131)
Mega Millions $\times$ log(Income per capita)	$-5.487 \times 10^{-4}$ (.064)	
SuperLotto Plus $\times$ log(Income per capita)	.145* (.057)	
Mega Millions $\times$ Middle-income group		.001 (.056)
Mega Millions $\times$ High-income group		-.016 (.097)
SuperLotto Plus $\times$ Middle-income group		.116* (.053)
SuperLotto Plus $\times$ High-income group		.211* (.086)
log(Jackpot size)	.260*** (.067)	.224** (.074)
log(Jackpot size) $\times$ log(Income per capita)	.044*** (.010)	
log(Jackpot size) $\times$ Middle-income group		.022* (.009)
log(Jackpot size) $\times$ High-income group		.056*** (.015)
Large jackpot (>\$200M)	1.098*** (.124)	1.098*** (.125)
Mega Millions draw day	1.263*** (.037)	1.263*** (.037)
SuperLotto Plus draw day	1.128*** (.040)	1.129*** (.040)
Clusters	Location, Draw	Location, Draw
Adj. $R^2$	.410	.405
Number of observations	684,279	684,279

We demean the variables log(Income per capita) and log(Jackpot size). Given the interaction term between the two, we interpret the coefficient for log(Jackpot size) as the effect of the size of the jackpot on the dependent variable when log(Income per capita) is at its mean. Robust standard errors are in parentheses. Significance given for the null hypothesis that the coefficient equals zero; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

tickets. This result is mostly driven by small purchases: Players who buy a single ticket tend to use Quick-Pick more frequently than those who buy two or three tickets. The coefficient on whether a transaction occurs on the game's draw day is significant and negative, indicating that players are more likely to choose numbers manually if the draw of the winning numbers occurs later that day. None of the other demographic variables are significant.<sup>14</sup> In summary, our results show that players in lower-income areas use the random selection option less frequently, which corresponds to lower expected earnings per ticket.

### ***Randomness of Manually Chosen Number Combinations***

For the estimation in this stage, the likelihood function in Equation 7 is computationally demanding because of the total number of winning combinations and combinatorial subsets of the different prize brackets needed to compute the values for  $p_{nbzjt}$ . As such, we approximate the likelihood function as the mean of a random sample of the total  $\bar{N}$  combinations and compute only random subsets of all the lower-tier prize brackets for each  $n$  (see Appendix B in Baker and McHale (2011) for more details). Using this approximation, we then maximize the log-likelihood function using simulated annealing and compute the standard errors using the covariance matrix calculated from the curvature of the likelihood function.

Given that the validations data encompass only 65 retailers, we make several assumptions to ensure adequate sample sizes when running the estimation procedure. First, SuperLotto Plus and Mega Millions have relatively few draws in our data set (just twice a week, for 456 draws in total) and few winners per retailer (an average of less than 0.1 winners per draw for the prize tiers in our data) as a result of the low odds created by the mega number. Our validations data also contain only a subset of the prize brackets for each game. As such, we jointly estimate the nonrandomness parameter  $\lambda$  across all three games, to have an adequate number of draws. We further aggregate the data by running the estimation on groups of locations, instead of each individual location, to increase the observed sales and winning tickets per draw. We collect these aggregate groups by

---

<sup>14</sup>In the transactional data set (65 locations), education is highly correlated with income per capita and is therefore omitted.

Table 7. Logit Model of Random Selection by Game

Dependent Variable: Choice of Quick-Pick						
	Fantasy 5		Mega Millions		SuperLotto Plus	
	(1)	(2)	(1)	(2)	(1)	(2)
log(Income per capita)	-.030 (.264)		.361* (.174)		.352* (.150)	
Middle-income group		.236 (.355)		.044 (.222)		.022 (.131)
High-income group		.214 (.421)		.129 (.307)		.212 (.190)
log(Number of tickets)	-.471*** (.104)	-.469*** (.103)	-.178*** (.050)	-.177*** (.050)	-.248*** (.042)	-.247*** (.041)
log(Jackpot size)	.101* (.050)	.104* (.047)	.108*** (.013)	.109*** (.013)	.096*** (.009)	.096*** (.009)
Purchase on draw day	-.423*** (.121)	-.428*** (.117)	-0.256*** (.038)	-.258*** (.039)	-.280*** (.020)	-.282*** (.021)
% age 65+	-.023 (.030)	-.026 (.030)	-.012 (.023)	-.009 (.026)	-.004 (.019)	.001 (.021)
% African American	.0002 (.007)	.001 (.007)	-.006 (.006)	-.009 (.006)	-.005 (.005)	-.006 (.005)
% Hispanic	-.007 (.006)	-.006 (.006)	-.001 (.003)	-.003 (.004)	.0003 (.003)	-.001 (.003)
Constant	.831 (2.712)	.281 (.798)	-3.51* (1.76)	.069 (.447)	-3.469* (1.509)	-.075 (.336)
Clusters	Location	Location	Location	Location	Location	Location
Number of observations	262,916	262,916	399,755	399,755	1,563,156	1,563,156

Robust standard errors are in parentheses. Significance given for the null hypothesis that the coefficient equals zero; \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001.

Table 8. Nonrandomness in Manual Number Selection

	Nonrandomness ( $\lambda$ )
Low-income group	59.570* (32.396)
Middle-income group	2.400 (4.427)
High-income group	14.352* (8.566)

Standard errors are in parentheses, generated through inversion of the hessian at the optimal solution. Significance levels are given for the null hypothesis that the nonrandomness parameter is equal to zero (random selection), \*  $p < 0.05$ .

separating the data into groups based on income per capita, corresponding to those in the previous models.

Table 8 shows the results for the estimation of the nonrandomness in number selection, from which we arrive at two conclusions. First, the low- and high-income groups are significantly different from zero; that is, players in these locations choose their combinations significantly nonrandomly. Second, nonrandomness in the combination selection is U-shaped across the three income groups. That is, the lowest- and highest-income groups have relatively greater preferences for similar combinations than the middle-income group, matching the model-free evidence provided in the data section.

### ***Simulation of Consumer and Retailer Earnings***

Using the coefficients of each stage, the observed data, and the assumption in error distributions, we ran simulations of earnings for consumers and retailers. We performed 5,000 iterations of the simulation, with each iteration consisting of simulated sales, game choice, Quick-Pick decisions, and combination selections for a week's worth of lottery draws across the three games. We tabulate prizes won by players across income groups for each game's draws and then aggregated them by game. We then averaged players and retailer earnings across simulation iterations.

Tables 9-11 report the results from the player and retailer simulations. Table 9 shows the expected earnings for consumers per ticket less the price of the ticket (\$1). The manual selection rows show earnings when players choose numbers manually, while the Quick-Pick rows show

Table 9. Simulated Player Expected Earnings Per Ticket

	Consumer Earnings Per Ticket		
	Fantasy 5	Mega Millions	SuperLotto Plus
Manual Selection			
Low-income group	-.461 (.250)	-.731 (.049)	-.457 (.069)
Middle-income group	-.334 (.054)	-.722 (.017)	-.415 (.018)
High-income group	-.386 (.115)	-.727 (.027)	-.439 (.058)
Quick-Pick	-.313 (.035)	-.720 (.004)	-.408 (.007)
All tickets	-.334 (.001)	-.722 (.001)	-.420 (.001)

Standard deviations across simulation runs in parentheses.

earnings for players using the random selection tool. The *All Tickets* row shows the overall earnings across all regions, for both manual and random selection. Across the games, SuperLotto Plus and Fantasy 5 have the highest expected earnings, and players who play Mega Millions lose considerably more per ticket. Across the three games, players who choose numbers manually tend to earn less per ticket than those who choose the Quick-Pick feature. Across the three income groups, the expected earnings are U-shaped with the low- and high-income groups losing more per ticket than the middle-income group. Moreover, the low-income group loses the most, per ticket, as a result of manual selection preferences. In terms of aggregate earnings, higher-income areas spend more per capita on lottery tickets and, as a result, incur greater total losses. All games are designed such that the player loses in expectation, so buying more tickets leads to higher losses. SuperLotto Plus purchases result in the highest aggregate loss for consumers, followed by Mega Millions and Fantasy 5.

With the values from Table 9, we can summarize the relative impact of each customer decision on the final earnings across the three income groups. Table 10 disaggregates the impact on consumer earnings across the different decision stages and shows how much the low- and middle-income groups could improve (or worsen) their earnings using the decisions from the high-income

Table 10. High-Income Group Earnings, Per Ticket, Relative to Other Groups

	Low-Income Group	Middle-Income Group
When to play	+0.61%	+0.24%
Which game to play	+7.01%	+3.86%
Choice of Quick-Pick	+1.30%	+0.11%
Choice of combination	+2.75%	-2.58%

Values by stage are the changes in earnings when each stage of the high-income group is independently substituted into the earnings calculation for the other income groups.

Table 11. Simulated Retailer Expected Earnings

	Earnings Per Ticket		Earnings Per Retailer	
	Mean	% Difference	Mean	% Difference
Low-income group	\$.0631 (.0013)	-.193% (.319)	\$39.46 (7.900)	-45.062% (.333)
Middle-income group	\$.0632 (.0013)	-.090% (.113)	\$57.28 (11.549)	-20.273% (.175)
High-income group	\$.0633 (.0014)		\$71.86 (14.550)	

Values in the table are for one week of draws. Standard deviations of the estimates across simulation runs are in parentheses.

group. The values in the table show the change in earnings when the corresponding decision from the high-income group is substituted into the earnings calculation of the low- and middle-income groups. For example, low-income consumers' earnings would improve by 1.30% if they were to use the Quick-Pick option as frequently as players in the high-income group. Positive values therefore denote the decisions when the high-income group generates comparatively higher earnings.

The table shows that lowest-income group makes inferior choices for all playing decisions. This group chooses worse days to play the lottery (adopting the high-income group's decisions on when to play would improve earnings by 0.61%) and its preference for lower payout draws leads to lower earnings than the other groups (7.01%). In addition, the low-income group performs worse because of a less frequent use of Quick-Pick (1.3%). Finally, for the manual choice of combinations, the low-income group is more attracted to commonly chosen numbers than the high-income group and, as a result, would improve its earnings by 2.75% if it adopted better number selection tactics.

Table 11 shows the effects of these differences in expected earnings for retailers. Retailer earnings mostly consist of the sales commission (6%) on the number of lottery tickets sold, which explains the values close to \$.06 (6% of the \$1 ticket) in earnings per ticket. The results for retailer earnings per ticket are consistent with those from the player earnings table: Retailers in lower-income areas earn 0.19% less per each sold ticket than those in higher-income areas. In aggregate, given all effects, retailers in lower-income areas generate 45% less in total revenue primarily because of lower sales volumes.

## **Managerial and Public Policy Implications**

### ***Consumers***

Lotteries provide an alluring opportunity for low-income people to overcome their low-income status (Haisley, Mostafa, and Loewenstein 2008). Lotteries are not only one of the few opportunities available to poor people to increase their wealth substantially and change their socioeconomic status but, importantly, are believed to be fair, presenting equal opportunity to win for everybody, given the consistent odds across all players (Brenner 1986; Haisley, Mostafa, and Loewenstein 2008; Lee and Chang 2008). However, our study demonstrates that low-income consumers make worse lottery play decisions than high-income consumers when it comes to choosing when to play (purchasing tickets for low-jackpot draws), using technology (Quick-Pick), and selecting numbers, resulting in lower expected earnings.

Given these findings, lottery operators could increase equality of outcomes by taking the following actions. First, they could, at least partially, move away from the pari-mutuel prize structure to one with fixed prizes. Doing so would eliminate differences in expected earnings across combinations because even the most popular combinations would receive the same prize regardless of the number of winners. The pari-mutuel format of Mega Millions, for example, is unique to California; all other participating states follow a fixed-prize format for the non-jackpot prize tiers. It is likely that this change would not significantly affect the “pleasure” of playing lottery and thus would not negatively affect the overall revenues to the state.

Second, potentially facing more resistance from customers, lottery operators could eliminate players' ability to choose their own numbers and assign all lottery tickets random combinations. Again, doing so would remove differences in expected earnings across combinations, even in a pari-mutuel setting. This practice is not uncommon outside of the U.S.: For example, in El Gordo, the annual Christmas lottery in Spain, people buy fixed numbers from 00000 to 99999 issued by the lottery administration and cannot pick their own numbers (Wikipedia contributors 2023). However, such change would also run the risk of alienating customers who enjoy picking their own numbers, which could affect the income groups differently and lead to less playing overall. This brings into focus the trade-off faced by state governments in having to weigh the well-being of their constituency against the revenue generated from inequitable game formats.

### ***Retailers***

Retailers also need to be cognizant of the consequences of the different consumer decisions, including choice of game and choice of number combinations (vs. randomly). As shown previously, retailers stand to gain if more consumers were to use the random selection option (because they would be less likely to choose popular combinations), or if they were to choose better games (in terms of expected earnings).

Retailer bonuses could also be aligned more closely with consumer earnings in several ways. First, including all prize tiers in the incentive structure would make retailers benefit from the larger pari-mutuel prize pools. This would then incentivize retailers to nudge consumers to choose number combinations that increase expected earnings. Second, sales bonuses for some jackpot prizes are capped (see Figure 1). As a result, for draws with large jackpots, retailers are indifferent to how much a potential jackpot-winning customer takes home, since the retailer would earn the same bonus. Removing this cap would further encourage retailers to maximize earnings by getting consumers to play optimally. Moreover, given the heterogeneity in nonrandom number selection, lottery agencies should be careful when altering retailer incentive structures. The decisions to remove bonuses for certain prize tiers, as some states have done (Texas Lottery Commission 2019),

could cause retailers to earn more per ticket should consumers choose less profitable combinations because of the tiered nature of the incentive structure.<sup>15</sup>

### *Advertising*

Prior research shows that it is in lottery agencies' best interest not to disclose information about expected value or secondary prizes, as it may drive people away from playing the games (Stearns and Borna 1995). These omissions are especially detrimental to poorer, less educated customers because of their greater susceptibility to traditional lottery advertising (Lee and Chang 2008). Our research suggests that lottery advertising, and general communication of game information to consumers, should be subjected to increased scrutiny. Although information on game odds and jackpot size is available, the expected payout by jackpot draw is not. Providing such information could help shift customers away from less profitable draws and games and mitigate the differences in game choice between low- and high-income customers. Our results suggest that equivalent distributions in game choice across income groups would improve earnings for low-income consumers by more than 7%. Furthermore, thorough explanations of the mechanics of the pari-mutuel format, specifically its implications on number selection strategies, could improve earnings for low-income consumers by approximately 3%. As shown by Stearns and Borna (1995), interventions that explain expected lottery earnings per ticket have significant impact on which lottery products consumers choose. Overall, requiring lottery advertising to disclose and prominently feature the expected earnings of tickets and to educate consumers about the consequences of relying on manual numbers selection would be a necessary first step toward ensuring equal outcomes (equal expected earnings) across lottery players of different income levels.

---

<sup>15</sup>For example, California retailers earn a cashing bonus when a consumer wins between \$99 and \$599 (see Figure 1). If a theoretical prize tier awards an average winning ticket \$650, retailers could benefit from consumers choosing similar combinations, since this could, on occasion, generate enough winners to drop the winning ticket prize below the \$599 threshold.

## Conclusion and Future Research

Our comprehensive examination of the impact of income on lottery play uncovers significant differences in lottery purchasing and playing behavior across income levels. First, consistent with prior research, we find that lottery sales are regressive. That is, people with lower income spend disproportionately more on lottery products. Second, we find that players in poorer areas are less responsive to increases in jackpot size and are less likely to switch to games with higher jackpots and higher expected payouts. Third, we find that players in poorer areas are less likely to use Quick-Pick when playing the lottery. Given that players typically choose their numbers with the aid of a popular heuristic (e.g., birthdays, “lucky” numbers), those using Quick-Pick will, on average, earn more than those who do not. As such, low-income players win less per ticket because of their tendency to forgo the random selection option. Finally, for those who choose numbers manually, players in low-income areas exhibit the highest nonrandomness in how they choose number combinations, which, owing to the pari-mutuel nature of the games, leads to smaller prizes for winning tickets. As a whole, our findings show that players in low-income areas (vs. players in high-income areas) make inferior choices at each stage of playing the lottery, amounting to 10% lower expected earnings per ticket. Our results suggest that the inequality of the lottery across income groups is greater than previously identified.

Although our results are consistent with previous research on cognitive resources, lottery products as entertainment, and lottery advertising, future work, whether using additional secondary data or experiments, could identify the causal mechanisms tying income differences to heterogeneity in lottery play decisions. Future research could also further investigate how the relationship between income and nonrandomness in number selection might vary depending on cultural and social factors. Moreover, we only examine large jackpot draw games in this study. Future studies could also examine similar types of relationships in other game formats.

## References

- Ariyabuddhiphongs, Vanchai (2006), "A test of the social cognitive model of lottery gambling in Thailand," *International Gambling Studies*, 6 (1), 77–93.
- Baker, Rose D. and Ian G. McHale (2009), "Modelling the probability distribution of prize winnings in the UK National Lottery: consequences of conscious selection," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172 (4), 813–834.
- Baker, Rose D. and Ian G. McHale (2011), "Investigating the behavioural characteristics of lottery players by using a combination preference model for conscious selection," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174 (4), 1071–1086.
- Balakrishnan, PV Sundar, Sudhir K. Chawla, Mary F. Smith, and Brian P. Michalski (1992), "Mail survey response rates using a lottery prize giveaway incentive," *Journal of Direct Marketing*, 6 (3), 54–59.
- Beckert, Jens and Mark Lutter (2013), "Why the poor play the lottery: Sociological approaches to explaining class-based lottery play," *Sociology*, 47 (6), 1152–1170.
- Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2013), "Who is 'behavioral'? Cognitive ability and anomalous preferences," *Journal of the European Economic Association*, 11 (6), 1231–1255.
- Blalock, Garrick, David R. Just, and Daniel H. Simon (2007), "Hitting the jackpot or hitting the skids: Entertainment, poverty, and the demand for state lotteries," *American Journal of Economics and Sociology*, 66 (3), 545–570.
- Borg, Mary O. and Paul M. Mason (1988), "The budgetary incidence of a lottery to support education," *National Tax Journal*, 41 (1), 75–85.
- Brenner, Gabrielle A. (1986), "Why do people gamble? Further Canadian evidence," *Journal of gambling behavior*, 2 (2), 121–129.
- Burger, Martijn J., Martijn Hendriks, Emma Pleeging, and Jan C. van Ours (2020), "The joy of lottery play: Evidence from a field experiment," *Experimental Economics*, 23 (4), 1235–1256.
- California Lottery "Where To Play," <https://www.calottery.com/where-to-play> (2023) [Online; accessed 21-September-2023].
- Carpenter, Robert E., Evan L. Perlman, and Donald F. Norris (2010), "Who Pays for the Maryland Lottery?: Evidence from Point of Sale Data," *The Journal of Gambling Business and Economics*, 4 (1), 31–52.
- Clotfelter, Charles T. (1979), "On the regressivity of state-operated" numbers" games," *National Tax Journal*, 32 (4), 543–548.
- Clotfelter, Charles T. and Philip J. Cook "Implicit taxation in lottery finance," Technical report, National Bureau of Economic Research (1987).

- Clotfelter, Charles T. and Philip J. Cook (1990), "On the economics of state lotteries," *Journal of Economic Perspectives*, 4 (4), 105–119.
- Clotfelter, Charles T. and Philip J. Cook (1991), *Selling hope: State lotteries in America* Harvard University Press.
- Clotfelter, Charles T. and Philip J. Cook (1993), "The "gambler's fallacy" in lottery play," *Management Science*, 39 (12), 1521–1525.
- Cook, Philip J. and Charles T. Clotfelter "The peculiar scale economies of lotto," Technical report, National Bureau of Economic Research (1991).
- Coughlin, Cletus C. and Thomas A. Garrett (2009), "Income and lottery sales: Transfers trump income from work and wealth," *Public Finance Review*, 37 (4), 447–469.
- Coups, Elliot, Geoffrey Haddock, and Paul Webley (1998), "Correlates and predictors of lottery play in the United Kingdom," *Journal of Gambling Studies*, 14 (3), 285–303.
- Cox, Simon J., Geoffrey J. Daniell, and Denis A. Nicole (1998), "Using maximum entropy to double one's expected winnings in the UK National Lottery," *Journal of the Royal Statistical Society: Series D (the Statistician)*, 47 (4), 629–641.
- Dhar, Sanjay K., Claudia Gonzalez-Vallejo, and Dilip Soman (1995), "Brand promotions as a lottery," *Marketing Letters*, 6 (3), 221–233.
- Elemia, Camille and Mike Ives "433 People Won a \$4 Million Lottery. Was It Pure Luck?," <https://www.nytimes.com/2022/10/05/world/asia/philippines-lottery-jackpot.html> (2022) [Online; accessed 21-August-2023].
- Farrell, Lisa, Roger Hartley, Gauthier Lanot, and Ian Walker (2000), "The demand for lotto: the role of conscious selection," *Journal of Business & Economic Statistics*, 18 (2), 228–241.
- Farrell, Lisa and Ian Walker (1999), "The welfare effects of lotto: Evidence from the UK," *Journal of public Economics*, 72 (1), 99–120.
- Feehan, Patrick and David Forrest (2007), "Distribution of UK National Lottery grants across local authority areas," *Applied Economics Letters*, 14 (5), 361–365.
- Forrest, David, Robert Simmons, and Neil Chesters (2002), "Buying a dream: Alternative models of demand for lotto," *Economic Inquiry*, 40 (3), 485–496.
- Frederick, Shane (2005), "Cognitive reflection and decision making," *Journal of Economic Perspectives*, 19 (4), 25–42.
- Garrett, Thomas A. and Cletus C. Coughlin (2009), "Inter-temporal differences in the income elasticity of demand for lottery tickets," *National Tax Journal*, 62 (1), 77–99.
- Garrett, Thomas A. and Russell S. Sobel (2004), "State lottery revenue: The importance of game characteristics," *Public Finance Review*, 32 (3), 313–330.

- Goldsmith, Kelly and On Amir (2010), "Can uncertainty improve promotions?," *Journal of Marketing Research*, 47 (6), 1070–1077.
- Griffiths, Mark and Richard Wood (2001), "The psychology of lottery gambling," *International gambling studies*, 1 (1), 27–45.
- Grote, Kent and Victor Matheson (2011), "The economics of lotteries: A survey of the literature," *College of the Holy Cross, Department of Economics Faculty Research Series*, (11-09).
- Guillen, Mauro F. and Adrian E. Tschoegl (2002), "Banking on gambling: Banks and lottery-linked deposit accounts," *Journal of Financial Services Research*, 21 (3), 219–231.
- Guryan, Jonathan and Melissa S. Kearney (2008), "Gambling at lucky stores: Empirical evidence from state lottery sales," *American Economic Review*, 98 (1), 458–73.
- Haisley, Emily, Romel Mostafa, and George Loewenstein (2008), "Subjective relative income and lottery ticket purchases," *Journal of Behavioral Decision Making*, 21 (3), 283–295.
- Hansen, Ann, Anthony D. Miyazaki, and David E. Sprott (2000), "The tax incidence of lotteries: Evidence from five states," *Journal of Consumer Affairs*, 34 (2), 182–203.
- Hardoon, Karen K., Hayley R. Baboushkin, Jeffrey L. Derevensky, and Rina Gupta (2001), "Underlying cognitions in the selection of lottery tickets," *Journal of Clinical Psychology*, 57 (6), 749–763.
- He, Chen and Tobias J. Klein (2022), "Advertising as a reminder: Evidence from the Dutch State Lottery," *Marketing Science*.
- Hersch, Philip L. and Gerald S. McDougall (1989), "Do people put their money where their votes are? The case of lottery tickets," *Southern Economic Journal*, 56 (1), 32–38.
- Hirshleifer, Jack (1966), "Investment decision under uncertainty: Applications of the state-preference approach," *The Quarterly Journal of Economics*, 80 (2), 252–277.
- Holtgraves, Thomas and James Skeel (1992), "Cognitive biases in playing the lottery: Estimating the odds and choosing the numbers," *Journal of Applied Social Psychology*, 22 (12), 934–952.
- Jackson, Raymond (1994), "Demand for lottery products in Massachusetts," *Journal of Consumer Affairs*, 28 (2), 313–325.
- Kitchen, Harry and Scott Powells (1991), "Lottery expenditures in Canada: A regional analysis of determinants and incidence," *Applied Economics*, 23 (12), 1845–1852.
- Kubovy, Michael (1977), "Response availability and the apparent spontaneity of numerical choices.," *Journal of Experimental Psychology: Human Perception and Performance*, 3 (2), 359.
- Laitner, John (1999), "Means-tested public assistance and the demand for state lottery tickets," *Review of Economic Dynamics*, 2 (1), 273–290.

- Landman, Janet and Ross Petty (2000), ““It could have been you”: How states exploit counterfactual thought to market lotteries,” *Psychology & Marketing*, 17 (4), 299–321.
- Layton, Allan and Andrew Worthington (1999), “The impact of socio-economic factors on gambling expenditure,” *International Journal of Social Economics*, 26 (1/2/3), 430–440.
- Lee, Jennifer “Who Needs Giacomo? Bet on the Fortune Cookie,” <https://www.nytimes.com/2005/05/11/nyregion/who-needs-giacomo-bet-on-the-fortune-cookie.html> (2005) [Online; accessed 21-August-2023].
- Lee, Yu-Kang and Chun-Tuan Chang (2008), “A social landslide: Social inequalities of lottery advertising in Taiwan,” *Social Behavior and Personality: an international journal*, 36 (10), 1423–1437.
- Livernois, John R. (1987), “The redistributive effects of lotteries: Evidence from Canada,” *Public Finance Quarterly*, 15 (3), 339–351.
- Lorenz, Valerie C. (1990), “State lotteries and compulsive gambling,” *Journal of Gambling Studies*, 6 (4), 383–396.
- Matheson, Victor and Kent Grote (2008), “US lotto markets,” pages 503–524.
- McCaffery, Edward J. (1994), “Why people play lotteries and why it matters,” *Wis. L. Rev.*, page 71.
- McMullan, John L. and Delthia Miller (2009), “Wins, winning and winners: The commercial advertising of lottery gambling,” *Journal of gambling studies*, 25 (3), 273–295.
- Mikesell, John L. (1989), “A note on the changing incidence of state lottery finance,” *Social Science Quarterly*, 70 (2), 513.
- Miller, Keith C. (2019), “State Lotteries and Their Customers,” *UNLV Gaming LJ*, 9, 177.
- Miyazaki, Anthony D., Ann Hansen, and David E. Sprott (1998), “A longitudinal analysis of income-based tax regressivity of state-sponsored lotteries,” *Journal of Public Policy & Marketing*, 17 (2), 161–172.
- Narayanan, Sridhar and Puneet Manchanda (2012), “An empirical analysis of individual level casino gambling behavior,” *Quantitative Marketing and Economics*, 10 (1), 27–62.
- Oster, Emily (2004), “Are all lotteries regressive? Evidence from the Powerball,” *National Tax Journal*, 57 (2), 179–187.
- Rogers, Paul (1998), “The cognitive psychology of lottery gambling: A theoretical review,” *Journal of gambling studies*, 14 (2), 111–134.
- Scott, Frank and John Garen (1994), “Probability of purchase, amount of purchase, and the demographic incidence of the lottery tax,” *Journal of Public Economics*, 54 (1), 121–143.

- Semrush “*lottery.net*,” <https://www.semrush.com/website/lottery.net/overview/> (2023) [Online; accessed 21-August-2023].
- Stearns, James M. and Shaheen Borna (1995), “The ethics of lottery advertising: Issues and evidence,” *Journal of Business Ethics*, 14 (1), 43–51.
- Stranahan, Harriet A. and Mary O. Borg (2004), “Some futures are brighter than others: The net benefits received by Florida bright futures scholarship recipients,” *Public Finance Review*, 32 (1), 105–126.
- Suits, Daniel B. (1977), “Gambling taxes: Regressivity and revenue potential,” *National Tax Journal*, 30 (1), 19–35.
- Texas Lottery Commission “*Retailer Benefits*,” [https://www.texaslottery.com/export/sites/lottery/Retailers/Retailer\\_Benefits\\_x\\_Bonuses/index.html](https://www.texaslottery.com/export/sites/lottery/Retailers/Retailer_Benefits_x_Bonuses/index.html) (2019) [Online; accessed 2019].
- Thaler, Richard H. and William T. Ziemba (1988), “Anomalies: Parimutuel betting markets: Racetracks and lotteries,” *Journal of Economic perspectives*, 2 (2), 161–174.
- Tversky, Amos and Daniel Kahneman (1974), “Judgment under uncertainty: Heuristics and biases,” *Science*, 185 (4157), 1124–1131.
- Wang, Tong V., Rogier Jan Dave Potter van Loon, Martijn J. Van den Assem, and Dennie Van Dolder (2016), “Number preferences in lotteries,” *Judgment and Decision Making*, 11 (3), 243–259.
- Wikipedia contributors “Spanish Christmas Lottery — Wikipedia, The Free Encyclopedia,” [https://en.wikipedia.org/w/index.php?title=Spanish\\_Christmas\\_Lottery&oldid=1171349682](https://en.wikipedia.org/w/index.php?title=Spanish_Christmas_Lottery&oldid=1171349682) (2023) [Online; accessed 21-August-2023].
- Wohl, Michael J.A., Nyla R. Branscombe, and Jamey J. Lister (2014), “When the going gets tough: Economic threat increases financial risk taking in games of chance,” *Social Psychological and Personality Science*, 5 (2), 211–217.
- Wu, Roland Y. (1979), “A cross-section study on the demand for state lottery tickets,” *The American Economist*, 23 (2), 6–12.
- Yan, Dengfeng and Anaimalai V. Muthukrishnan (2014), “Killing hope with good intentions: The effects of consolation prizes on preference for lottery promotions,” *Journal of Marketing Research*, 51 (2), 198–204.