Are All Lotteries Regressive? Evidence from the Powerball

Emily Oster*
Harvard University

January 5, 2004

Abstract

The regressivity of lotteries has become an increasingly important issue in the U.S. as the number of state-run lotteries has increased. Despite this, we still know relatively little about the nature of lottery regressivity. I use a new dataset on Powerball lotto sales to analyze how regressivity varies with jackpot size within a single lotto game. I find that these large-stakes games are significantly less regressive at higher jackpot sizes. An out-of-sample extrapolation of these results suggest that the lottery becomes progressive at a jackpot around $806 million. This suggests that concerns about regressivity might be allayed by concentrating lotto games to produce higher average jackpots.

*I am very grateful to the Connecticut State Lottery for providing the data. David Laibson, Markus Mobius, Al Roth, Jesse Shapiro and two anonymous referees provided helpful discussion and comments. Contact Information: eoster@post.harvard.edu

1 Introduction

Currently in the United States 38 states either run their own lottery or participate in multi-state lottery consortiums. Lottery revenues amounted to $12 billion in 1998, and have continued to climb since that time.[1]

Despite their pervasiveness and obvious appeal (and the fact that the proceeds are generally used to finance public goods), lotteries are increasingly under attack by groups who feel that the games take advantage of the poor and poorly educated. All lottery products have an expected value lower than their cost so the lottery is implicitly taxed; if lotteries are disproportionately consumed by the poor, then this tax is regressive. There are a number of paper that analyze lottery regressivity as a whole, and nearly all find that, on average, lotteries are regressive (see, for example: Clotfelter and Cook (1987,1989), Kitchen and Powells (1991), Brown, Kaldenberg and Browne (1992), Scott and Garen (1994), Pirog-Good and Mikesell (1995), Stranahan and Borg (1998),Worthington (2001), Rubenstein and Scafidi (2002)). However, despite the volume of work on this topic we still know very little about the nature of this regressivity.

This paper differs from earlier work on lottery regressivity by analyzing not only whether the lottery is regressive, but whether the regressivity of a given lottery game varies with variations in prize. This issue has not been addressed previously in the literature, although Price and Novak (1999,2000) have noted differences in regressivity across lottery games. From a policy perspective, understanding regressivity more fully, and understanding when it is more or less prevalent, may be helpful for addressing popular concerns about the lottery tax.

Using a new panel dataset on Powerball lottery sales, I present evidence that the lottery is less regressive at higher jackpot levels. The dataset used has information on sales by

[1] http://www.publicgaming.org/lotben.html

zip code for 199 lottery drawings in Connecticut in which the prize varied from \$10 million to \$150 million. The primary advantage of this dataset over those used in the past is the ability to compare behavior across zip codes as the jackpot changes. Previous work has generally been constrained to look either at the change in sales for all consumers as the prize increases or at the behavior across income groups for a single prize. The richness of the dataset used here allows us to look at both factors simultaneously. The results show significant decreases in regressivity as the jackpot increases. Out-of-sample extrapolation suggests that this lottery would become progressive at a jackpot around \$806 million.

This result suggests that concerns about regressivity might be partially alleviated concentrating lotto games so jackpots are larger. In addition, to the extent that these results can be generalized out of sample, they suggests that instant lottery games are likely to be substantially more regressive than lotto.

2 Data

The Powerball is a multi-state lotto game run by a central consortium. Twenty-two states sell tickets and the revenues from each state's own ticket sales return to them. Powerball is a pari-mutual game, so as more tickets are purchased the jackpot increases. On a given ticket, a player chooses five numbers from a pool of 49 (without replacement) and then one from a pool of 42 (this last is the "powerball"). The numbers are drawn randomly and the chance of matching all 6 is one in 80,089,128 (note that this does not change based on the number of tickets purchased). There are smaller prizes for matching fewer than six numbers, down to \$3 for matching only the powerball. If more than one person matches all six numbers the jackpot is split.[2]

[2] In this sense, the value of a ticket may fall as more tickets are purchased because the chance of sharing the prize increases. However, the chance of matching all six numbers remains unaffected by others' purchases.

Powerball drawings are held twice a week. It is possible (even likely) that no one will win in a given drawing. If that happens the current jackpot "rolls over" and adds to the jackpot for the next drawing. In this way the jackpots can get very large. The minimum Powerball jackpot is $10 million and the jackpots have reached as high as $300 million.

The data used in this paper come from the Connecticut state lottery. The Connecticut lottery office provided sales information for each retailer in their state for every day between August 1999 and July 2001, encompassing 199 drawings. The sales data are aggregated by zip code. Demographic statistics about each zip code (including information on income, education, race, unemployment and urban population) are collected from the 2000 census. Although I have data on sales and prize in a panel format (across zip codes over time) the demographic data for each zip code is constant in all observations. Given that I use sales data spanning 1999-2001 and the demographic data is from the 2000 census it is likely to be a very good match.

Table 1 shows descriptive statistics for the lottery and demographic data. There are 265 zip codes in the census, although only 246 of them are home to any lottery retailers, and the 19 zip codes without any retailers are not used in the analysis. In addition, it should be noted that several zip codes are missing sales for specific drawings, due to administrative errors in data collection. However, the results presented below are robust to excluding all zip codes that are missing any observations. The primary demographic fact to note from Table 1 is that Connecticut is a state with a wide range of income levels, which makes it ideal for the analysis in this paper.[3]

[3]The only potential concern about using Connecticut for this analysis is that at very high jackpots there is occasionally an influx of ticket purchasers from New York into one of the richest areas in Connecticut. This is discussed in more detail in the results section.

3 Analysis

The primary result in this paper is shown in Figure 1, in which the percent of total state sales contributed by the richest 20% of zip codes and the poorest 20% are graphed against jackpot size.[4] This figure demonstrates that while poor zip codes buy more tickets at lower jackpot sizes, the series converge and, at the highest jackpot sizes sales are substantially greater in the richer areas. At a jackpot of only $10 million we expect the poorest 20% of zip codes to be contributing about 25% of the sales, and the richest 20% contributing only about 20%. However, at the highest jackpot levels the poorest 20% contribute only about 19% and the richest contribute close to 32%.

The figure is somewhat stylized as it considers only the richest and the poorest areas. To test whether the relationship is generally true, I run the following regression:

$$LogSales_{i,t} = \alpha + \beta_1(prize_t) + \beta_2(LogHHincome_i) + \beta_3(LogHHincome_i \times prize_t)$$

where i indexes zip codes and t indexes time. *Log HH income* is the log of the median household income in the zip code. The dependent variable is log sales per adult, implicitly controlling for population differences across zip codes. The coefficient of interest is β_3, which represents the income elasticity of sales. A positive value for β_3 implies that richer zip codes have a higher elasticity of sales with respect to prize – their play increases more than that of poor zip codes as the jackpot goes up. In other words, a positive value would imply that the lottery is less regressive at higher jackpot levels.

The results from the estimation of this equation can be seen in Column 1 of Table 2. The results indicate that β_3 is, indeed, positive and significantly different from zero. The income elasticity of sales is 0.0021 – implying that a 10% increase in income increases the responsiveness to prize by about 0.02%. Given that the gap between the richest and poorest

[4] For ease of interpretation, jackpots are grouped into intervals of $5 million, so this shows the average percent contributed for jackpots of $10 to $15 million, $15 to $20 million and so on.

zip code in Connecticut is close to \$100,000, the coefficient implies sizable elasticity differences.

Controls for other zip code demographics are not included in this regression for the simple reason that I am concerned solely with whether the tax is regressive and to what extent, rather than why. That is, finding that part of the effect of income is actually an effect of race does not change the calculation of the regressivity of the tax, which is a function simply of the relationship between sales and income. However, it is worth exploring whether the size of the interaction would change if additional controls were added. Column 2 of Table 1 adds to the primary regression some simple demographic controls and the interaction between these controls and prize. The coefficient on the interaction between income and prize is largely unaffected (it actually increases slightly), suggesting that the effect is due largely to income.

The decreasing regressivity of the lottery with jackpot size is further illustrated in Figure 2, in which the income elasticity of sales is graphed against jackpot size. This figure was produced by regressing sales on median household income for all drawings in the sample and graphing the results against jackpot size. It is easy to see that the elasticity is increasing as jackpot size increases. At the lowest jackpot sizes, the elasticity is about -0.70, while at \$150 million it is only -0.22. A linear regression of the elasticity on jackpot size yields the same result as in Table 2: a one million dollar increase in prize increases the income elasticity of demand by 0.0021. To give some sense of magnitude, the estimates here imply that a two-adult household at the 20th income percentile would purchase \$2.17 in lottery tickets at a jackpot of \$10 million and \$8.11 at a jackpot of \$100 million. In contrast, for a two-adult household at the 80th income percentile these values are \$1.53 and \$6.31. Their relative purchase amount therefore increases from 70% to 78% of the purchases of the poorer household.

There are several important issues with this analysis that deserve to be addressed here. First, the single \$150 million prize in this data has a very different sales composition than any other observation. This leads to concerns about the robustness of the results to exclusion of this particular jackpot. In fact, as can be seen in Column 1 of Table 3, exclusion of this jackpot decreases the magnitude of the interaction only very slightly, to 0.0020.

A more general concern is with the use of zip-code level data for this analysis. The analysis assumes that the zip code demographics are the same as the purchaser demographics. This will not be true if either there is substantial within-state, across-zip code migration for ticket purchases or if there is across-state migration. Although it is not possible to address the within-state issue using this data, there exists individual-level data from the 1999 National Gambling Impact Study that suggest ticket purchases are generally made close to home. When lottery-playing individuals were asked if they bought their lotto tickets in their neighborhood, 84% of people who played more frequently than once a year said they did. Since a neighborhood is probably smaller than a zip code in most cases, this number would likely be even higher were they asked about within zip code purchases. This provides some confidence in a high correlation between population and purchaser demographics.

Travel across state lines for ticket purchases is also an issue, particularly in Connecticut, which is bordered by two states (New York and Massachusetts) that do not offer the Powerball. In addition, three of the richest zip codes in the state are on the border closest to New York City. It is well known that these areas see an influx of purchasers when the jackpot is very large. Therefore, as a robustness check I run two regressions that exclude certain zip codes. In Column 2 of Table 3 I exclude the three zip codes on the border close to New York City, and in Column 3 I exclude all other border zip codes. The coefficient is essentially unchanged in Column 3, and although it decreases slightly in Column 2, it remains of similar magnitude and significance.

4 Discussion and Conclusion

There is a large literature in economics on why people play lotteries, and although it is not the goal of this paper to fully address this issue, it is interesting to speculate on what type of theory would produce the empirical results seen here. Given that even big jackpot lotteries are only very, very rarely positive expected value, most theories of why people play lotteries rely either on a "fun" component of gambling which increases lottery utility (for example, Conlisk (1993) and Caplin and Leahy (1998)), or on players having a poor understanding of the odds of the game (for example, Kahneman and Tversky (1979) and Camerer (2000)). In fact, it is easy to see how either theory could be consistent with the empirical fact presented here.

Consistent with the first case, it could be that the "fun" of the lottery is dependent in part on how large the prize is relative to the player's current wealth. If this is the case, richer individuals may only be willing to play at higher jackpots – they have a higher threshold for entry – which would produce this result. In the second case, if income tracks education, the rich may have a better understanding of the odds. This, too, could cause richer players to wait until higher jackpots to enter. The data used in this paper does not allow me to differentiate between these theories, although future theoretical work on why people play lotteries may be informed by this empirical finding.

The decreasing regressivity of the lottery product as jackpot increases suggests that lotteries are not always regressive. Although it is a substantially out-of-sample prediction, it is interesting to consider at what point the lottery tax becomes progressive, which can be done by extrapolating the results in Figure 2. The tax is progressive when the income elasticity is equal to one, in this case at a jackpot of $806 million. This should be taken with extreme caution, since it is out-of-sample and it is difficult to know how people would behave at jackpots of this magnitude. It is also difficult to know whether this jackpot level is something that could ever be achieved. It is worth noting, however, that the Powerball has reached jackpots of over $300 million, and yet the Powerball states made up only 22.5% of total lotto sales in the country in 1997 (National Gambling Impact Study Comission 1999). This suggests that a jackpots as high as the progressive level would not be out of the question with, for example, a national lottery.

It is worth noting, with caution, that there are some policy prescriptions that may come out of the empirical result in this paper. First, concentrating lotteries so there are fewer games, with longer odds and higher jackpots, could allay some fears about regressivity. In fact, there has been a trend in this direction even without the knowledge of the jackpot-regressivity relationship as lottery organizers have noticed that people appear to respond to the jackpot size much more than the expected value of a ticket (Forest, Simmons and Chesters 2002). Second, consistent with the evidence in Price and Novak (1999,2000), out of sample predictions here suggest that instant lottery games are substantially more regressive than higher-stakes lotto games. This suggests that, if regressivity is perceived to be a problem, interventions should focus more on the instant games and less on large jackpot lottery products.

References

Brown, D.J., D.O. Kaldenberg, and B.A. Browne, "Socio-Economic Status and Playing the Lotteries," *Sociology and Social Research*, 1992, *76*, 161–167.

Camerer, Colin, *Prospect Theory in the Wild: Evidence from the Field*, Cambridge, England: Cambridge University Press,

Caplin, A and John Leahy, "Anticipation, Uncertainty and Time Inconsistency," June 1998.

Clotfelter, Charles and P Cook, *Selling Hope: State Lotteries in America*, Cambridge: Harvard University Press, 1989.

___ **and Philip Cook**, "Implicit Taxation in Lottery Finance," *National Tax Journal*, 1987, *40* (4), 533–546.

Conlisk, J, "The Utility of Gambling," *Journal of Risk and Uncertainty*, 1993, *5*, 255–275.

Forest, David, R. Simmons, and N. Chesters, "Buying a Dream: Alternative Models of Demand for Lotto," *Economic Inquiry*, 2002, *40* (3), 485–496.

Kahneman, D and A Tversky, "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, 1979, *47*, 263–291.

Kitchen, H and S Powells, "Lottery Expenditures in Canada: A Regional Analysis of Determinants and Incidence," *Applied Econometrics*, 1991, *23*, 1845–1852.

National Gambling Impact Study Comission, "National Gambling Impact Study Commission Final Report," Technical Report, National Gambling Impact Study Comission 1999.

Pirog-Good, M and J Mikesell, "Longitudinal Evidence of the Changing Socio-Economic Profile of a State Lottery Market," *Policy Studies Journal*, 1995, *23*, 451–465.

Price, D and E.S Novak, "The Tax Incidence of Three Texas Lottery Games: Regressivity, Race and Education," *National Tax Journal*, 1999, *52*, 741–751.

___ **and** ___ , "The income Redistribution Effects of Texas State Lottery Games," *Public Finance Review*, 2000, *28*, 82–92.

Rubenstein, R and B Scafidi, "Who Pays and Who Benefits? Examining the Distributional Consequences of the Georgia Lottery for Education," *National Tax Journal*, 2002, *55* (2), 223–238.

Scott, F and J Garen, "Probability of Purchase, Amount of Purchase and the Demongraphic Incidence of the Lottery Tax," *Journal of Public Economics*, 1994, *54*, 121–143.

Stranahan, H and M Borg, "Horizontal Equity Implications of the Lottery Tax," *National Tax Journal*, 1998, *51* (1), 71–82.

Worthington, A, "Implicit Finance in Gambling Expenditures: Australian Evidence on Socioeconomic and Demographic Tax Incidence," *Public Finance Review*, 2001, *29*, 326–342.

Table 1

Descriptive Statistics for Connecticut

	Mean	Std. Deviation	Min	Max
Prize (in Millions)	31.32	23.43	10	150
State Sales	\$514,288	\$528,536	\$92,159	\$6,811,443
Adult Population	9,214	8,411	34	41,932
Median Household Income	\$60,758	\$21,844	\$11,913	\$146,755
Percent African-American	5.8%	11.6%	0%	86.5%
Percent (at least) High School Graduates	46.9%	5.5%	0%	100%
Percent Unemployed	4.6%	4.1%	0%	36.5%
Percent Urban	68.1%	37.6%	0%	100%

Table 2[a]

Differences in Prize Elasticity Across Zip Codes

Dependent Variable: Log Sales per Capita

	(1)	(2)
Explanatory Variables:		
Prize (in Millions)	.011***	.012***
	(11.15)	(4.94)
Log Median HH Income	−.709***	−.946***
	(−4.93)	(−5.3)
Log Median HH Income × Prize	.00214***	.00261**
	(3.64)	(2.26)
Percent HS Grad		−.491
		(−.20)
Percent African-American		−1.269***
		(−2.85)
Percent Urban		.483***
		(3.19)
Percent Unemployed		−1.722
		(−1.00)
Pct HS Grad × Prize		−.005
		(−.70)
Pct African American × Prize		.00022
		(−.11)
Pct Urban × Prize		.001
		(1.49)
Pct Unemployed × Prize		.002
		(.28)
constant	−1.066***	−.599
	(−4.16)	(−.47)
Number of Observations	48385	48385
R^2	.21	.26

[a] t-statistics in parenthesis, adjusted for zip-code level clustering.

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3[a]

Robustness Checks on Results

Dependent Variable: Log Sales per Capita

	(1)[b]	(2)[c]	(3)[d]
Explanatory Variables:			
Prize (in Millions)	.011***	.012***	.011***
	(10.88)	(11.61)	(10.18)
Log Median HH Income	−.704***	−.673***	−.698***
	(−4.89)	(−4.8)	(−4.5)
Log Median HH Income × Prize	.00196***	.00185***	.00204***
	(3.43)	(3.14)	(3.16)
constant	−1.052***	−1.126***	−1.099***
	(−4.10)	(−4.48)	(−4.02)
Number of Observations	48141	47792	44026
R^2	.19	.21	.21

[b] Excludes $150 Million Prize

[c] Excludes Three New York City Border Zip Codes

[d] Excludes All Other Border Zip Codes

[a] t-statistics in parenthesis, adjusted for zip-code level clustering

* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1
Percent of Sales by Richest and Poorest Zip Codes

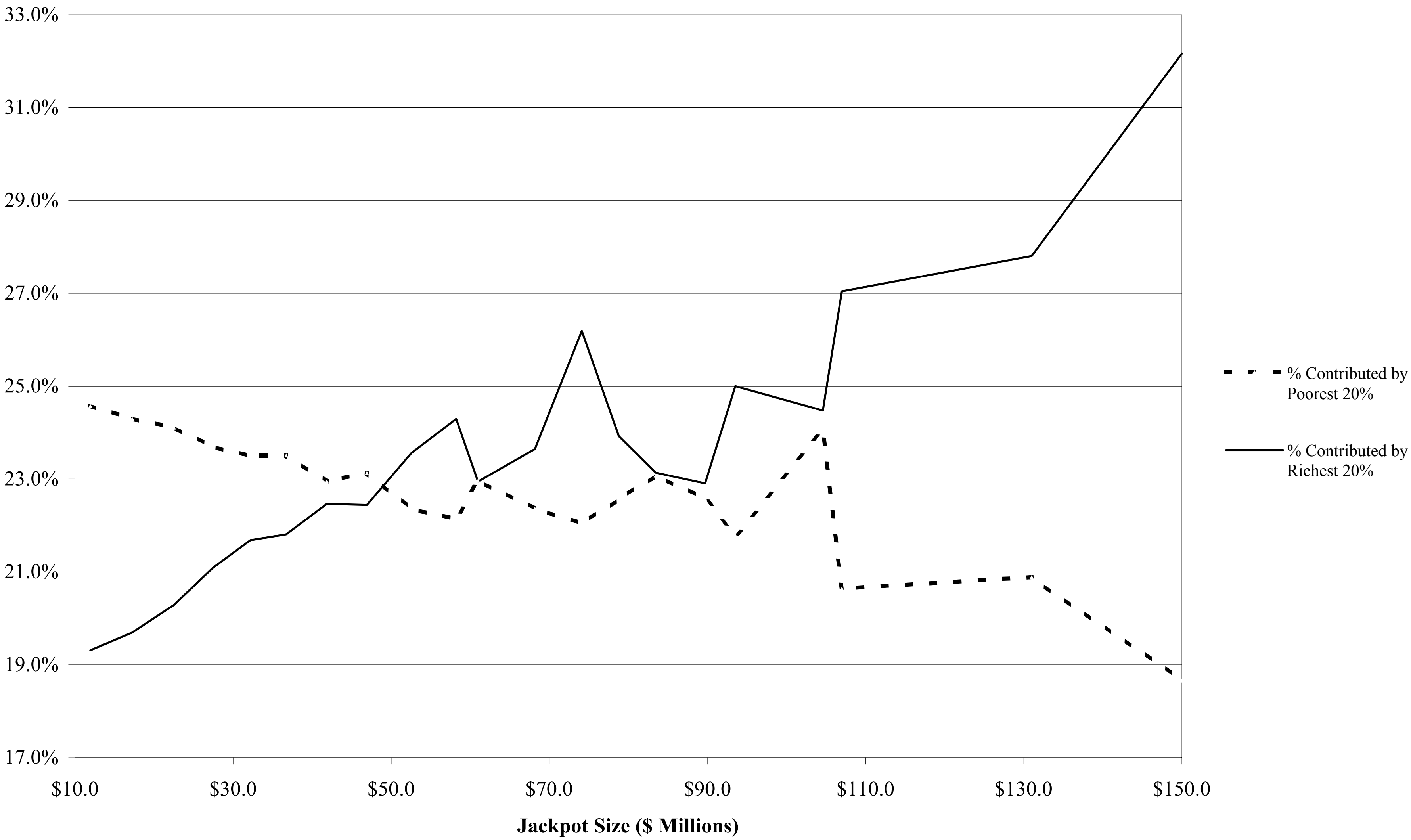

Figure 2
Income Elasticity of Sales Versus Prize