The Optimal Taxation of Lotteries
Who P(l)ays and Who Wins?

Hunt Allcott, Benjamin B. Lockwood, Dmitry Taubinsky*

February 6, 2020

PRELIMINARY; PLEASE DO NOT CITE OR CIRCULATE WITHOUT PERMISSION.

Abstract

The average American household spends $665 per year playing state-run lottery games. There is a long-standing debate as to whether these lotteries are a regressive "tax on people who are bad at math" or a "win-win" that generates both consumer surplus and government revenues. We study optimal lottery policy through the lens of optimal taxation, where lotteries are a taxed good whose consumers may be subject to behavioral biases. We derive new sufficient statistics formulas for optimal pricing and attributes of a government-provided good. We then estimate the key parameters using lottery prizes and sales data and a large new nationally representative survey. Individual-level lottery expenditures are highly correlated with survey measures of innumeracy and poor statistical reasoning, but our observable measures of behavioral bias statistically explain only about 15 percent of lottery purchases for the average household. We estimate that lottery demand is highly responsive to ticket prices and jackpot amounts, but not to smaller prizes. Using these empirical moments, we calibrate a structural model of lottery demand. In the model, lotteries are indeed a welfare-improving "win-win," and implicit tax is similar to the current norms in U.S. states.

JEL Codes: D12, D61, D91, H21, H42, H71.

Keywords: Behavioral public economics, behavioral welfare analysis, lotteries.

*Allcott: Microsoft Research, New York University, NBER, and JPAL. hunt.allcott@nyu.edu. Lockwood: Wharton and NBER. ben.lockwood@wharton.upenn.edu. Taubinsky: Berkeley and NBER. dmitry.taubinsky@berkeley.edu. We thank Charlie Clotfelter, Alex Imas, Melissa Kearney, and seminar participants at Berkeley and the 2019 National Tax Association conference for helpful comments. We are grateful to the Wharton Dean's Research Fund, the Wharton Behavioral Lab, Time-Sharing Experiments for the Social Science, and the Sloan Foundation for grant funding. We thank Raj Bhargava and Afras Sial for research assistance.

> "[A lottery] preys upon the hard earnings of the poor; it plunders the ignorant and simple."
>
> – U.S. Supreme Court, in Phalen v. Virginia (1850)

> "In our stressful world, the ability to dream is well worth the price of a lottery ticket ... The lottery is simply a form of entertainment that happens to benefit your state."
>
> – National Association of State and Provincial Lotteries (2020)

> "Since this regressive, addictive, partially hidden tax is here to stay, might a little improvement still be conceivable? ... Here's a modest suggestion: States should consider reducing their skim of the wagers."
>
> – Purdue University president and former Indiana governor Mitch Daniels (2019)

People have long debated whether states should run lotteries. Opponents argue that lotteries are a regressive tax on people who are bad at math. Proponents argue that lotteries are a win-win, generating both consumer surplus and government revenues. If states do run lotteries, there are further debates, such as the optimal "implicit tax"—the share of revenues that is allocated to the government instead of returned to prize winners. Economists are divided: a recent survey of the University of Chicago IGM experts panel found that 23 percent of leading economists believe that state-run lotteries increase social welfare, 28 percent disagree, and 45 percent were uncertain or had no opinion.[1]

These debates matter. Americans spent $85 billion, or a remarkable $665 per household, on lottery tickets in 2018 (NASPL 2020). This contributed $23 billion to state and local governments that year (NASPL 2020). This is more than the revenue raised by federal estate or tobacco taxes, and just less than the revenue raised by the federal gas tax. Americans spend more on lottery tickets than they do on cigarettes, and more than they do on music, sports tickets, movie tickets, books, and video games combined (Isidore 2015).

Embedded within these debates is a series of deeper empirical and theoretical questions. How much of lottery consumption is driven by entertainment and other normatively respectable preferences versus ignorance, innumeracy, and other behavioral factors that a social planner might want to mitigate? Do lower-income people really spend more on lottery tickets, and is this a good thing (because it reflects consumer surplus for people with higher marginal utility) or a bad thing (because it reflects exploitation of behavioral bias)? Broadly, how does one consider normative questions around lotteries in a principled way?

This paper considers state-run lotteries in a public finance framework, extended to incorporate possible behavioral bias. We present new empirical data and analyses that identify a set of statistics that determine optimal policy. We then use the empirical parameters to simulate the welfare effects

[1] See www.igmchicago.org/surveys/state-run-lotteries.

of lotteries and different design choices. In doing so, we provide the first benefit-cost analysis of state-run lotteries that theoretically and quantitatively considers the alleged behavioral biases at the heart of the policy discussion.

In our model, consumers with heterogeneous earning ability choose labor supply, lottery purchases, and numeraire good consumption to maximize their perceived utility. We leave utility general, so that a rational consumer might play the lottery because of anticipatory utility, entertainment value, the consumption from the possible lottery winnings, and/or any other reason. However, lottery consumption might also be affected by perceptual distortions, such as misunderstanding of small probabilities, overconfidence, self-control problems, and other behavioral biases that the social planner does not consider to be normatively relevant. The social planner has a lottery with exogenous prize structure, and sets the lottery purchase price, net lottery payout, and a non-linear income tax to maximize normative utility subject to a revenue raising constraint. The planner is inequality averse, placing higher welfare weights on people with lower earning ability. Setting a 100 percent implicit tax is equivalent to eliminating the lottery altogether.

The closest parallel to our model is the "optimal sin tax" framework of Allcott, Lockwood, and Taubinsky (2019). In the language of that framework, lottery tickets are a potential "sin good" (like alcohol, cigarettes, or sugary drinks) whose consumption may be affected by behavioral bias. A key difference is that in the framework in this paper, the state has direct control over a key product attribute: the lottery payout. Thus, policy directly targeting consumption of this potential "sin good" is multidimensional.

Put another way, the planner has two ways to change the implicit tax on lotteries: increase the price or decrease the payout. This introduces several important nuances, including the fact that biases such as perceptual distortions of probabilities would imply that mis-estimation of the normative utility of a lottery ticket is endogenous to the lottery payout. The optimal choice depends on which is a more progressive way of collecting revenue and reducing bias.

Generally, the social welfare effect of a state-run lottery depends on several sufficient statistics. More behavioral bias generally implies a higher implicit tax on lottery revenues. Relatively more behavioral bias among lower-income people implies a higher implicit tax because the corrective benefits of the implicit tax accrue to the poor. However, relatively more consumption among lower-income people imply a lower implicit tax because the mechanical utility losses from taxation accrue to the poor. The relative importance of the corrective benefits versus the mechanical losses depend on the demand slope: more responsive demand means that the corrective benefits of demand changes outweigh the mechanical losses, whereas less responsive demand means that the corrective benefits will be relatively small.

We gather a novel and extensive collection of data to estimate the empirical parameters required by the theory. Remarkably, to our knowledge, there are no nationally representative surveys of lottery expenditures since Clotfelter et al. (1999).[2] Using an online survey weighted for national representativeness, we provide updated estimates of Americans' lottery expenditures by income, finding that lottery spending takes an inverted-U shape in income: middle-income people spend about $25 per person per month, high-income people spend 10 percent less, and low-income people spend one-third less. Lotteries have changed substantially over the past 21 years, with the growth of instant and multi-state lotto games, and this pattern of expenditures by income is markedly different from the results in Clotfelter et al. (1999).

We use the same survey to provide new evidence on the relationship between lottery expenditures and proxies for innumeracy, overconfidence, and other behavioral biases. We gather survey measures of financial literacy, numeracy, statistical errors such as the Gambler's Fallacy and non-belief in the law of large numbers, overconfidence, incorrect beliefs about expected returns, and self-reported self-control problems. These survey measures of behavioral bias are strongly correlated with lottery expenditures: a one standard deviation increase in an overall bias index is associated with 54 log points higher expenditures, even after controlling for a rich set of preference measures and demographics. Furthermore, lower-income people have substantially higher bias indexes. These suggestive correlations are consistent with the Supreme Court's 1850 argument that a lottery "plunders the ignorant" and "preys upon the hard earnings of the poor."

We quantify bias using what Allcott, Lockwood, and Taubinsky (2019) call the "counterfactual normative consumer" strategy, predicting the "normative" consumption that people would choose if their survey answers (counterfactually) reflected zero behavioral bias. We find that about 15 percent of lottery consumption is statistically explained by our bias proxies. Of course, this estimate should be interpreted carefully, as one must make a strong unconfoundedness assumption to interpret the conditional correlation between bias proxies and consumption as a causal relationship.

The theory also requires us to identify the elasticity of lottery demand with respect to both prize amounts and prices. We collect a new nationwide dataset of lottery purchases, prize structures, and jackpot and other prize amount from state sources and La Fleur's, a large data provider. To identify the effect of ticket prices on demand, we study two recent cases where multi-state lotto games increased ticket prices from $1 to $2. To identify the effect of prize amounts on demand, we exploit random variation in lotto prize amounts over time as prizes are won (returning the prize to some reset value) or not (rolling over the prize into the next drawing). We find that demand is much more responsive to the expected value of prize winnings from the jackpot than from the second largest prize. This is the opposite of what would be expected for risk-averse consumers, but it could be explained by probability weighting or other factors that cause variation in the jackpot to be more salient than variation in smaller prizes.

Finally, we calibrate a structural model of lottery demand to match the reduced form empirical moments, and we simulate welfare at different prices and implicit taxes under different assumptions

[2]The Consumer Expenditure Survey significantly under-counts lottery purchases (Kearney 2005).

about bias. Even without behavioral bias—that is, if the social planner fully respects the revealed preferences of lottery consumers—an implicit tax on the order of 30 percent is optimal. This is because higher-income people consume more lottery tickets, so higher lottery taxes are a useful form of redistribution.

If we assume that more consumption is explained by behavioral bias, the optimal implicit tax increases in order to offset the overconsumption caused by bias. However, even if we calibrate bias to explain a share of consumption that is four times larger than our surveys suggest, our simulations still suggest lotteries raise welfare on the whole. Bias would have to be substantially larger than this for our model to predict that lotteries reduce welfare. Indeed, our preferred calibrations suggest that the current implicit tax rate for Powerball and Mega Millions, the two large multi-state lotteries, is close to the level that maximizes social welfare.

Our work connects to three existing literatures. The first is the literature in behavioral economics, judgement and decision-making, psychology, and other fields on decision-making under risk in general and with lottery purchases in particular (Kahneman and Tversky 1979; Haisley et al. 2008; Guryan and Kearney 2008; Post et al. 2008; Snowberg and Wolfers 2010; Aruoba and Kearney 2011; etc.). Our work adapts the insights from this literature into a quantitative welfare analysis.

The second area of related literature is on the public economics of lotteries (Clotfelter and Cook 1987, 1989, 1990; Price and Novak 1999, 2000; Grote and Matheson 2011; etc.). To our knowledge, our paper is the first to embed lottery policy questions into an optimal taxation model that accounts for behavioral bias.

The third area is the recent line of work evaluating behaviorally motivated public policy decisions.[3] Our paper extends this work to consider lotteries, and lays out a framework that we hope can provide the basis for a line of future work on economically optimal lottery design.

Sections I, II, III, and IV present the model, data, empirical results, and optimal policy simulations, respectively.

I Theoretical Framework

I.A A static model of lottery demand

We begin by setting lottery consumption in a standard optimal tax setting, where we treat lottery tickets as a standard consumption good. In the body of the paper we consider a simplified model in which exogenous windfalls of money do not significantly affect lottery demand or labor supply,

[3] See, e.g., Allcott and Rafkin 2020; Allcott, Kim, Taubinsky, and Zinman 2019; Handel 2013; Taubinsky and Rees-Jones (2018); Rees-Jones and Taubinsky (forthcoming); Lockwood 2019; Allcott and Taubinsky, 2015; Allcott, Lockwood, and Taubinsky, 2019; Gruber and Kőszegi, 2001; Handel and Kolstad, 2015; Handel, Kolstand, and Spinnewijn, 2016.

and in which demand for a particular type of lottery is binary. In the appendix, we present results for a continuous choice model that allows for income effects on lottery demand and labor supply.

We consider the binary demand assumption to correspond to our modeling of lottery purchases in a particular week for a particular game. Individuals rarely buy more than several lottery tickets at a time, perhaps because the "fun" of having a chance to win does not increase with the number of tickets. Empirically, we also do not find that shocks to individuals' incomes have a meaningful impact on lottery demand, consistent with our modeling. Finally, the negligible labor supply income effects assumption is supported by Gruber and Saez (2002), who find small and insignificant income effects on labor supply, and by Saez, Slemrod, and Giertz (2012), who review the empirical literature on labor supply elasticities and argue that "in the absence of compelling evidence about significant income effects in the case of overall reported income, it seems reasonable to consider the case with no income effects."

Formally, individuals choose earnings z, numeraire consumption c, and lottery consumption $s \in \{0, 1\}$. Earnings are subject to a nonlinear income tax $T(z)$, and the price of the lottery ticket is denoted p. The price of numeraire consumption is normalized to one.

Individuals have types $\theta \in \mathbb{R}$ distributed according to F_θ that determine their preferences for lotteries and their earnings ability. We assume types θ are unidimensional to simplify exposition, but we prove results under more general assumptions for the binary choice model in the Appendix. To model preferences that exhibit no (causal) income effects on lottery purchases, we assume that individuals' decisions can be approximated as maximizing a utility function $U = G(c + su(a; \theta, \varepsilon) - \psi(z; \theta))$, where m is the perceived utility from purchasing a lottery ticket and $\psi(z; \theta)$ is the labor cost of generating earnings z. The parameter a corresponds to lottery attributes manipulable by the policymaker, such as the the prize amounts and the various probabilities. The parameter $\varepsilon \in \mathbb{R}$, distributed according to F_ε, is a random preference shock that we think of as varying within individual over time.

The function u can reflect a number of possible motives for playing the lottery. First, u of course includes the material utility of winnings from the lottery. Second, it can reflect entertainment value from playing the lottery (Conlisk, 1993; Kearney, 2005), or belief utility from thinking about a chance of winning (e.g., Loewenstein, 1987; Caplin and Leahy, 2001; Brunnermeier and Parker, 2005). Third, we allow for perceptual distortions, such as over- or under-estimating the likelihood of winning or imperfect processing of small probabilities (Steiner and Stewart, 2016; Woodford, 2012). We allow for this third possibility by positing that individuals' normative utility is $v(a; \theta, \varepsilon)$. Probability weighting (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992) could be reflected both in non-standard motives such as anticipatory utility (Brunnermeier and Parker, 2005; Gottlieb, 2014) or in perceptual distortions (Steiner and Stewart, 2016; Woodford, 2012).

We let $\bar{s}(p, a, \theta)$ denote aggregate demand by type θ consumers, and we let $\bar{s}_V(p, a, \theta)$ denote the aggregate demand if consumers instead maximize $V = G(c + sv(a; \theta, \varepsilon) - \psi(z; \theta))$. Our results apply to cases in which $\bar{s}$ and $\bar{s}_V$ apply to only a single purchase occasion, as well to cases in which $\bar{s}$ and $\bar{s}_V$ apply to $N > 1$ purchase occasions. With $N \geq 1$ purchase occasions, the demand functions are given by $\bar{s} = NPr\left(u(a;\theta,\varepsilon) \geq p\right)$ and $\bar{s}_V = NPr\left(v(a;\theta,\varepsilon) \geq p\right)$, respectively. We assume that preference shocks are smoothly distributed conditional on θ, so that $\bar{s}$ and $\bar{s}_V$ are smooth function of attributes a and price p for each θ. When no ambiguity arises, we sometimes suppress some arguments and write, for example, $\bar{s}(\theta)$ for concision.

For simplicity, in the core theoretical analysis and simulations we require a to be unidimensional; in particular, a corresponds to the overall tax-inclusive expected monetary value of the lottery. Concretely, we consider the state offering a lottery with probabilities π_k and prizes aw_k, where we normalize so that $\sum_k \pi_k w_k = 1$. Given a price p per ticket, the total expenditures on lotteries are $p\bar{s}$, where $\bar{s}$ is aggregate population demand. The amount of money paid out by the state is $a\bar{s}$, which implies that share of expenditures retained by the state—"the implicit lottery tax"—is $1 - a/p$.

The policymaker selects T, p and a to maximize normative utility, aggregated across consumers using type-specific pareto weights $\mu(\theta,\varepsilon)$

$$\max_{T,p,a}\left[\int_\theta \mu(\theta,\varepsilon)U(c,s,a,z;\theta,\varepsilon)dF_\theta(\theta)dF_\varepsilon(\varepsilon)\right], \tag{1}$$

subject to a government budget constraint,

$$\int_\theta ((p-a)\bar{s}(\theta) + T(z(\theta)))dF(\theta) \geq R \tag{2}$$

and to consumer optimization

$$\{c(\theta), s(\theta), z(\theta)\} = \arg\max_{\{c,s,z\}} U(c,s,z,a;\theta,\varepsilon) \quad \text{s.t.} \quad c(\theta) + ps(\theta) \leq z - T(z) \text{ for all } \theta. \tag{3}$$

We let λ denote the shadow-value of public funds (the multiplier on the government budget constraint at the optimum), and we let $g(\theta,\varepsilon) = \mu(\theta,\varepsilon)U_c'$ denote the weighted marginal utility from consumption of type θ,ε individual. Following Saez (2002) and others we assume that $g(\theta,\varepsilon)$ are equal for all z-earners at the optimal tax system. We thus use $g(z)$ to denote the social marginal welfare weight of z-earners. Under the assumption of negligible income effects on labor supply, $E[g(z)] = 1$.

I.A.1 Elasticity concepts and sufficient statistics

The optimal sin tax depends on three types of sufficient statistics: elasticities, money-metric measure of bias, and the "progressivity of bias correction." These statistics are understood to be endogenous to the tax regime (t,T), though we suppress those arguments for notational simplicity. Because we assume that social marginal welfare weights depend only on income, the relevant statistics are functions of income z, rather than of types θ. We write $\bar{s}(z)$ to denote lottery demand of z-earners, and we use $\bar{s}$ to denote population-wide demand.

We define the price and attribute semi-elasticities as $\zeta^p(z) = \frac{d \ln \bar{s}(z)}{dp}$ and $\zeta^a(a) = \frac{d \ln \bar{s}(z)}{da}$, and we use $\bar{\zeta}^p$ and $\bar{\zeta}^a$ to denote population-wide semi-elasticities. We denote individuals' willingness to pay for a marginal increase in a by $\kappa(z) = -\zeta^a(z)/\zeta^p(z)$. We let $\bar{\kappa} = E[s(z)\kappa(z)]/\bar{s}$ denote the average willingness to pay by all buyers of the lottery ticket.

We define the bias of type θ consumers, $\gamma(a,\theta)$, to be the wedge $u(a,\theta) - v(m,\theta)$. This definition has a simple price-metric interpretation, consistent with the bias definition in ALT: γ is equal to the price reduction that produces the same change in demand as the bias does. Formally, $\bar{s}^V(p-\gamma, a, \theta) = \bar{s}(p, a, \theta)$. Figure 1 provides a graphical illustration of bias. The price-metric measure of bias, γ, is the vertical difference between $\bar{s}$ and $\bar{s}^V$. The horizontal difference between $\bar{s}$ and $\bar{s}^V$, $\Delta\bar{s}$ is the difference in demand induced by the bias. Notice that once $\Delta\bar{s}$ is known, the vertical distance can be obtained by measuring the price elasticity of demand. Our empirical strategy for quantifying bias will proceed along these lines.

Following ALT, we define $\bar{\gamma}(z) := E\left[\gamma(a,\theta)|z(\theta) = z\right]$ to be the bias of z earners, and we define $\bar{\gamma} := \frac{E\left[\bar{\gamma}(z)\frac{d\bar{s}(z)}{dp}\right]}{E\left[\frac{d\bar{s}(z)}{dp}\right]}$ to be the average bias of individuals marginal at price p (and attribute a). Because our framework includes redistributive motives, and because our bias definition is in money-metric terms, the welfare effects of behavior change will depend not just on the average bias but on whose behavior is being changed. All else equal, the welfare change is more positive when the benefits of counteracting individuals' biases are concentrated on low-income individuals. To take these redistributive concerns into account, we define $\sigma_p = Cov\left[g(z), \frac{\bar{\gamma}(z)}{\bar{\gamma}}\frac{d\bar{s}(z)}{dp}\right]/\frac{d\bar{s}}{dp}$ and $\sigma_a = Cov\left[g(z), \frac{\bar{\gamma}(z)}{\bar{\gamma}}\frac{d\bar{s}(z)}{da}\right]/\frac{d\bar{s}}{da}$ as the benefits of bias correction with respect to changes in p and a, respectively. These statistics quantify the extent to which the benefits of bias correction accrue to low-income individuals, per unit change in $\bar{s}$.

Finally, because biases may be endogenous to the attribute a, the welfare effects of varying lottery attributes will depend on how biases change with a. We define $\rho(z) := \frac{d\gamma(a,z)}{a}$, to be the change in z-earners' bias with respect to a. Analogous to $\bar{\kappa}$, we define $\bar{\rho} = E[\bar{s}(z)\rho(z)]/\bar{s}$ to be the average change among those who buy the ticket. Consequently, $\bar{\kappa} - \bar{\rho}$ is the average impact of increasing a on individuals' *normative* utility.

I.A.2 Simple example

To take a simple and illustrative example, to which we will return when we study our optimal tax formulas, suppose that utility is linear in the numeraire consumption, and that the lottery offers a single large prize aw with a small probability π. Taking into account anticipatory utility, consumers' perceived utility from a lottery ticket is given by $u(a,\theta,\varepsilon) = (1+\omega_\theta(\pi))\pi m(aw) + \varepsilon$, where the decision weights $\omega_\theta(\pi) > 0$ for small probabilities π and where m is concave. This formulation approximates a utility function that is relatively linear in numeraire consumption relative to the diminishing marginal returns of increase a very large prize of hundreds of millions of dollars.

Individuals may imperfectly understand small probabilities, however, so that their normative utility is in fact given by $v(a, \theta, \varepsilon) = (1 + \omega_\theta^v(\pi))\pi m(aw) + \varepsilon$. In this simple example, the bias is $\gamma(\theta) = (\omega_\theta(\pi) - \omega_\theta^v(\pi))\pi m(aw)$.

Our framework also allows for misperception of lottery value. For example, if individuals overestimate their happiness from winning by a factor $\tilde{\gamma}$, so that $u = (1 + \omega_\theta(\pi))\pi m(aw) + \varepsilon$ while $v = (1 + \tilde{\gamma})(1 + \omega_\theta(\pi))\pi m(aw) + \varepsilon$. In this case, the bias is $\gamma(\theta) = \tilde{\gamma}(1 + \omega_\theta(\pi))\pi m(aw)$

I.B A formula for the optimal lottery "tax"

To obtain intuition for our main result, consider how a change dp in the price p affects social welfare:

- First, it mechanically decreases individuals' incomes by $s(z)dp$, and mechanically increases government revenues by $dp\bar{s}$. The net mechanical effect is thus $E[\bar{s} - s(z)g(z)]dp = -Cov[\bar{s}(z), g(z)]dp$
- Second, it leads to substitution away from lottery purchases, which changes government revenue by $(p - a)\frac{d\bar{s}}{dp}dp$
- Third, the substitution has a net internality reduction benefit of $-\bar{\gamma}(1 + \sigma_p)\frac{d\bar{s}}{dp}dp$

Taking these terms together, the net effect on social welfare is

$$dW = -Cov[\bar{s}(z), g(z)]dp + (p - a)\frac{d\bar{s}}{dp}dp - \bar{\gamma}(1 + \sigma)\frac{d\bar{s}}{dp}dp$$

which implies that the optimal lottery price must satisfy the first order condition

$$p - a = \bar{\gamma}(1 + \sigma_p) + \frac{Cov[\bar{s}(z), g(z)]}{\frac{d\bar{s}}{dp}} \qquad (4)$$

In the absence of biases ($\gamma(z) = 0$) and redistributive motives ($Cov[s(z), g(z)] = 0$), the optimal lottery ticket price p must equal a, the expected winnings of the lottery. That is, the state breaks even at the optimum. It is optimal for the state to offer a net subsidy—i.e., offer net positive expected value lotteries—if the lottery is preferred by lower-income consumers, or if it is undervalued by individuals because, e.g., they believe that expected revenues are lower than they are. However, a positive $\bar{\gamma}$ pushes p to exceed a. That is, if consumers overvalue lotteries due to various behavioral biases, then it is optimal to discourage lottery demand by setting $p > a$.

We next consider the impact of changing a. To maintain the same signs as in the analysis of a price increase, we consider the impact of a small increase da in $1 - a$. The welfare effects of this are as follows:

- First, there is a mechanical effect on individuals' utility and on state revenues. The impact on perceived utility is $-E[s(z)\kappa(z)]da$. However the perceived utility of a lottery ticket differs from normative utility by $\gamma(a,z)$ at each value of a. Thus, the impact on normative utility is $-E[s(z)(\kappa(z)-\rho(z)]da$. The mechanical effect on state revenue is $\bar{s}da$. The net mechanical effect is therefore $\bar{s}da - E[\bar{s}(z)(\kappa(z)-\rho(z))g(z)]da$. We can rewrite this as $\bar{s}(1-\bar{\kappa}+\bar{\rho}) - Cov[\bar{s}(z)(\kappa(z)-\rho(z)), g(z)]da$
- Second, this leads to substitution away from lottery purchases, which impacts revenue by $(p-a)\frac{d\bar{s}}{d(1-a)}da$
- Third, the welfare impact of this substitution through bias reduction is $-\bar{\gamma}(1+\sigma_a)\frac{d\bar{s}}{d(1-a)}da$

Putting these effects together leads to the first-order condition

$$p-a=\bar{\gamma}(1+\sigma_a)-\frac{\bar{s}(1-\bar{\kappa}+\bar{\rho})}{\frac{d\bar{s}}{d(1-a)}}+\frac{Cov[\bar{s}(z)(\kappa(z)-\rho(z)),g(z)]}{\frac{d\bar{s}}{d(1-a)}} \tag{5}$$

Combining the first-order conditions in (4) and (5), we have the following result, which we formally derive and generalize in the Appendix.

Proposition 1. *If pand a are optimal and $a>0$,*

$$\begin{aligned} p-a &= \bar{\gamma}(1+\sigma_p)+\frac{Cov[\bar{s}(z),g(z)]}{\frac{d\bar{s}}{dp}} \\ \bar{\kappa} &= 1+\bar{\rho}-\bar{\zeta}^a\left[\bar{\gamma}(\sigma_a-\sigma_p)+\frac{Cov[\bar{s}(z)(\kappa(z)-\rho(z)),g(z)]}{\frac{d\bar{s}}{d(1-a)}}-\frac{Cov[\bar{s}(z),g(z)]}{\frac{d\bar{s}}{dp}}\right] \end{aligned} \tag{6}$$

If $a=0$ is optimal then

$$\bar{\gamma}(1+\sigma_a)-\frac{\bar{s}(1-\bar{\kappa}+\bar{\rho})}{\frac{d\bar{s}}{d(1-a)}}+\frac{Cov[\bar{s}(z)(\kappa(z)-s(z)),g(z)]}{\frac{d\bar{s}}{d(1-a)}}<0$$

I.B.1 Interpreting the formulas and their implications

To provide intuition for the implications of our formulas, we consider a number of special cases.

No bias, homogeneous preferences. When $\gamma(\theta)\equiv 0$ and $\bar{s}(z)$, $\kappa(z)$ and $\rho(z)$ are constant in z, the result reduces to $p=a$ and $\bar{\kappa}(a)=1$. In other words, the lottery is revenue neutral, and its scale a is set such that individuals' average willingness to pay for a is equal to the cost of increasing a.

In the simple example where $u(a,\theta)=(1+\omega(\pi))\pi m(aw)+\varepsilon$, this implies the that p and a are determined by the conditions $p=a$ and $(1+\omega(\pi))\pi w m'(aw)=1$. The second condition uniquely determines a, which then determines the price p. For example, if $m(x) = \ln(1+x)$, then optimal lottery structure is $p = a = \max\left(\frac{1}{w}\left((1+\omega(\pi))\pi w - 1\right), 0\right)$. Since we normalize $\pi w = 1$, this implies that lotteries are offered if and only if $\omega(\pi) > 0$; i.e., additional entertainment utility leads individuals to value lotteries above their monetary value.

Revenue-maximizing lottery structure. The revenue maximizing lottery structure can be obtained from our general result by setting $g(z) \equiv 0$ and ignoring the bias terms. In this case, the formula again reduces to $p = a$ and $\bar{\kappa}(a) = 1$. Thus, when there is no bias and no variation in preferences by income, the welfare-maximizing lottery is also the revenue-maximizing lottery.

Homogeneous bias and preferences. When $\bar{s}(z)$, $\gamma(z)$, $\kappa(z)$ and $\rho(z)$ are homogeneous across the income distribution, the first-order conditions reduce to $p - a = \bar{\gamma}$ and $\bar{\kappa}(a) = 1 + \bar{\rho}(a)$. In this case, the price is set above a lottery ticket's expected value when $\bar{\gamma} > 0$, so as to discourage lottery consumption. Moreover, the amount of prize winnings retained by individuals, a, is set such that $\bar{\kappa}(a) > 1$ when $\bar{\rho} > 1$ (individuals not only overvalue the lottery ticket, but also changes in a). This implies that the optimal choice of a is lower than what it would be when individuals overvalue the marginal benefits of increasing a (under the reasonable assumption that $\bar{\kappa}(a)$ is decreasing in a due to concavity of m). Thus, individuals are effectively taxed in two ways relative to the no-bias benchmark. First, the price is set to be higher than the expected monetary value of a lottery ticket. Second, the expected monetary value of the lottery ticket is lower than what it would be in the absence of bias.

In the simple example where $u = (1+\omega(\pi))\pi m(aw) + \varepsilon$ and $v = (1+\omega(\pi) - \tilde{\gamma}(\pi))\pi m(aw) + \varepsilon$, we have $\bar{\kappa}(a) = w(1+\omega(\pi))\pi m'(aw)$, $\gamma = \tilde{\gamma}(\pi)\pi m(aw)$ and $\rho = \tilde{\gamma}(\pi)\pi w m'(aw)$. We thus have the first-order conditions $p - a = \tilde{\gamma}\pi m(aw)$ and $\omega(\pi) w m'(aw) = 1 + \tilde{\gamma}\pi w m'(aw)$. Under the assumption that m is concave, there exists a high enough $\tilde{\gamma}$ such that the optimal choice of the lottery payout is $a = 0$. Thus, large-enough biases may imply that it is optimal to eliminate state-run lotteries, even if there is positive demand for them.

For example, if we specify $v(x) = \ln(1+x)$, then the optimal value of a is $a = \max\left(\frac{1}{w}\left((1+\omega(\pi) - \tilde{\gamma}(\pi))\pi w - 1\right)\right.$ Thus, a is decreasing in $\tilde{\gamma}$. Moreover, a high enough level of bias makes the choice of $a = 0$ optimal. If, e.g., $\tilde{\gamma}(\pi) = \omega(\pi)$, so that all probability weighting is due to incorrect beliefs or perceptual distortion, then it is optimal to eliminate state-run lotteries.

Homogeneous bias, variation in preferences by income. Next suppose that $\bar{s}(z)$ varies with z, but that $\gamma(z)$ and $\rho(z)$ do not vary by income. For example, this holds in a simple model in which $u = (1+\omega_\theta(\pi))\pi m(aw) + \varepsilon$ and $v = (1+\omega_\theta(\pi) - \tilde{\gamma})\pi m(aw) + \varepsilon$, so that the bias at each income z is $\gamma(z) = \tilde{\gamma}\pi m(aw)$. The first-order conditions are now

$$p - a = \bar{\gamma} + \frac{Cov[\bar{s}(z), g(z)]}{\frac{d\bar{s}}{dp}} \tag{7}$$

$$\bar{\kappa} = 1 + \bar{\rho} - \bar{\zeta}^a \left[\frac{Cov[\bar{s}(z)(\kappa(z) - \rho(z)), g(z)]}{\frac{d\bar{s}}{d(1-a)}} - \frac{Cov[\bar{s}(z), g(z)]}{\frac{d\bar{s}}{dp}} \right] \tag{8}$$

Compared to the case of homogeneous preferences case, lottery design is now affected by the extent to which increasing price is progressive or regressive, $Cov[\bar{s}(z), g(z)]$, and the extent to which decreasing a is progressive or regressive, $Cov[\bar{s}(z)(\kappa(z) - \rho(z)), g(z)]$. Consider first condition (7). This condition states that if low-income individuals are less likely to purchase lottery tickets, so that $Cov[\bar{s}(z), g(z)]$ is negative, then $p - a > \bar{\gamma}$. The extent to which p falls below a depends on the extent of the progressivity, and on the elasticity of demand. The less responsive individuals are to changes in the price, the higher is the optimal price. Intuitively, this is because of standard "Laffer curve" reasoning—if individuals substitute quickly away from high lottery "taxes," then optimal progressive tax will not be effective in generating revenue if it is too high.

As before, the optimal value of a is implicitly determined by the first-order condition for $\bar{\kappa}(a)$. Compared to the case of homogeneous preferences, the new bracketed term in equation (8) corresponds to the difference in the redistributive consequences, per unit change in demand, between increasing the implicit tax $1 - a$ versus decreasing p. To see this, first note that $s(z)(\kappa(z) - \rho(z))$ is the impact on z-earners' normative utility of increase a, and thus $Cov[\bar{s}(z)(\kappa(z) - \rho(z)), g(z)]$ measures the extent to which increasing $1 - a$ is progressive or regressive. Second, note that a change in $1 - a$ and a change in p will generally produce different changes in total demand $\bar{s}$. Thus an apples to apples comparison of the redistributive consequences of increasing p versus increasing $1 - a$ involves normalizing the redistributive effects by how demand responds to changes in those parameters.

When increasing the implicit lottery tax $1 - a$ is a more progressive way of collecting revenue than increasing p, the bracketed term will be positive. Since $\bar{\zeta}^a$ is negative, this leads $\bar{\kappa}(a)$ to be higher and thus the optimal choice of a to be lower. Conversely, when increasing the price is a more progressive way of collecting revenue than increasing the implicit tax $1 - a$, the bracketed term will be negative, which leads to a lower optimal value of $\bar{\kappa}(a)$. In words, the implicit lottery tax is increasing in the degree to which it is a more progressive way of collecting revenue than simply increasing the price.

Heterogeneous bias, variation in preferences by income. When bias is further allowed to be heterogeneous, as in the general result in Proposition 1, the intuition above is modified in two additional ways. First, the optimal price p will be increasing in σ_p, the extent to which bias correction is progressive. Second, the optimal implicit tax $1 - a$ will also depend $\sigma_a - \sigma_p$, the difference between how progressively increases in $1 - a$ versus increases in p counteract individuals' biases.

II Data

We consider three categories of lottery games. The first is lotto games, in which players pick a set of numbers, and they win if their numbers match those drawn in prize drawings, which are typically held daily, bi-weekly, or weekly. If no player wins the jackpot, the jackpot becomes larger in the next drawing. If one or more players win the jackpot, they split the jackpot and jackpot returns to some reset value in the next drawing. The largest lotto games in the U.S. are the two major multi-state lotto games, Mega Millions and Powerball, but many states also run their own games. Tickets typically cost $1 or $2.

The second category is instant or "scratch-off" games. Players buy a paper card and scratch off parts of it to reveal hidden numbers, letters, or signs. Players win if the hidden values match some pattern, for example if all symbols in a given row are the same. Tickets typically cost $1 to $20. The third category is other games, including Keno and video lottery terminals. As shown in Appendix Figure tk, instant game consumption has grown substantially since the mid-1990s, while lotto consumption has dropped slightly, and other games grew during the late 2000s but have returned to their 1990s levels.

II.A Survey Data

We designed a survey to measure lottery expenditures as well as biases and preferences affecting purchase of lottery tickets. We fielded the survey in November 2018 using Qualtrics Panels. We obtained complete responses from 2767 individuals who passed an attention check.[4]

Table 1, panel a, summarizes the demographics of our sample. Panel b of the table summarizes the respondent-level data. Appendix TKTK gives the exact text of the survey questions.

Lottery expenditures. We asked three questions to measure lottery expenditures. First, we asked survey respondents how much they spend on lottery games with prize drawings (e.g., MegaMillions, Poweball, Keno) in an average month. Second, we respondents how much they spend on instant or scratch-off games in an average month. Third, we asked them for total lottery expenditures, with a consistency check to make sure that their answer to this third question totaled the sum of their answers to the first two questions.

[4]Specifically, the attention check had the following instructions: *"In order to facilitate our research, we are interested in knowing certain factors about you. Specifically, we are interested in whether you actually take the time to read the directions; if not, then the data we collect based on your responses will be invalid. So, in order to demonstrate that you have read the instructions, please ignore the next question, simply select "Any comments?" and write "I read the instructions" in the box provided. Thank you very much."* The options given listed marital status.

Preferences for playing. We asked two questions to assess normatively respectable preferences for playing the lottery. First, we asked participants to what extent the statement "For me, playing the lottery seems fun" applies to them. This question gauges the fun and excitement of playing. Second, we asked participants to what extent they "enjoy thinking about how life would be if I won the lottery." This question gauges anticipatory utility from playing.

To assess biases in evaluation of lotteries, we created a battery of possible

Expectations about lottery payouts. To gauge whether participants understand how much of the lottery expenditures are returned back to participants, we asked them what percent of total expenditures was given out the players in prizes. To gauge whether participants are overconfident, in the sense that they think they are luckier or more skilled than than the average lottery player, we asked two additional questions. The first question asks how much the survey respondent would win back, on average, for every $1000 they spend on lottery tickets of their choice. The second question asks how much they think the average person would win back. The difference in answers to these questions is our measure of overconfidence.

Statistical reasoning. A possible bias in the evaluation of lotteries is that individuals have trouble assessing prize structures with small probabilities, which leads to perceptual distortions over-weighting the small probabilities. Our hypothesis was that this is most likely to arise for individuals who have generally poor statistical reasoning skills. To assess statistical reasoning, we tested whether respondents (i) are able to calculate expected values of gambles; (ii) whether respondents exhibit the gambler's fallacy (TK cites), as measured by questions about the likelihood that an unbiased coin lands hands after a streaks of heads; and (iii) whether respondents exhibit non-belief in the law of large numbers (TK cites), as measured by a question about the likelihood that out of 1000 coin flips, the number of heads will be between 491 and 419. For (i) and (ii) our constructed proxies are the fraction of incorrect answers. For (iii) our constructed proxy is the difference between the correct answer and the respondent's answer.

Forecasted utility benefits of winning. As argued by Kahneman et al. (2006) and others, people may exaggerate the causal impact of wealth on well-being. Such a bias would cause people to be overestimate the utility gains from winning a lottery. To assess whether respondents correctly forecast the impact of winning large prizes on well-being, we asked them to predict the results from the Lindqvist, Ostling, and Cesarini (forthcoming) study of well-being among Swedish lottery players winning different amounts. Our proxy for bias is the difference between a respondent's prediction and the estimate in Lindqvist, Ostling, and Cesarini (forthcoming). This proxy may be positive or negative.

Financial literacy and numeracy. In addition to statistical reasoning, we hypothesized that individuals who are generally more financially literate and numerate, and thus more equipped and experienced in evaluating risky prospects, would be more likely to correctly evaluate lotteries. This greater sophistication could be associated both with fewer perceptual distortions of small probabilities and better understanding of how monetary windfalls affect well-being. We measure financial literacy using five questions from Lusardi and Mitchell (2014). We measure numeracy using three questions from Banks and Oldfield (2007). Our proxies are the fraction of questions answered correctly.

Self-control. Finally, to assess potential self-control problems we asked participants if they feel that they should play the lottery more, less or the same as they currently do. We allowed participants to state "less" to mitigate potential experimenter demand effects, and to allow for the possibility that for some subjects buying a lottery ticket is an activity with delayed gains (e.g., the anticipatory utility of thinking about winning) and delayed costs (e.g., the hassle cost of obtaining a lottery ticket).

II.B Lottery Sales and Prizes

We have four separate data sources on lottery sales and prizes. Panel (c) of Table 1 presents descriptive statistics on the sales and odds data from the four sources. All sales data are in dollars, not adjusted for inflation.

First, we purchased lottery sales data at the state-game-week level from 1994 to 2017 from La Fleur's. The data include week-level sales data for every lotto-style game for which we have prize and probability data in addition to aggregated instant game sales and other separately-reported game sales, which differ across states. We only use the two latter types of game sales in our national cross-game substitution analysis, while the lotto-style game data is used throughout the national analysis.

The data include 275 unique games in addition to aggregated instant game sales and span 41 states and D.C. We exclude some games from our analysis if we do not observe prize and probability data from the datasets described below. Mega Millions and Powerball are the only multi-state games included. As of 2017, 44 states and D.C. were operating lotteries, so the data include most, but not all states with lotteries during the period of interest. Puerto Rico and the U.S. Virgin Islands also operate lotteries during the period and are not included in the data. Furthermore, the data is not entirely complete for the state-years included, in that there do exist lotteries in some states in some years which are not included. This only affects our national cross-game substitution estimates, which depend on aggregated sales within a state.

Second, we have imputed lottery sales at the draw-game level in California from 2003 to 2017. These draw-level data are important in addition to the La Fleur's data because many lotto games have multiple draws (typically with different jackpots) in each week, so weekly data are not sufficiently precise. We use these data in the sub-jackpot demand elasticity analysis.

Specifically, the California data include draw-level sales data for California's SuperLotto Plus from 2003 to 2017, Mega Millions from 2005 to 2017, and Powerball from 2013 to 2017. California joined Mega Millions in 2005 and Powerball in 2013, while SuperLotto Plus, a state-specific game, began before 2003. The data span three different structures of Mega Millions, two of Powerball, and one of SuperLotto Plus. Within our complete dataset, these are the only three games that are both pari-mutuel—resulting in variation in their sub-jackpot prize values—and that regularly experience rollovers and hence significant build-ups in the second prize pool, allowing us to identify a sub-jackpot elasticity off of significant variation in the second prize.

We impute sales using prize pool amounts, prize pool allocation, and LaFleur's weekly sales data. Each game has nine prize pools, one for the jackpot and each smaller prize. The amount in each prize's pool is either reported when the prize rolls over or equals the product of the number of winners and prize amount per winner when won by at least one player. The increase in the prize pool from one draw to the next when a rollover occurs or from zero to the next draw's prize amount when the prize is won (since the pool is depleted by the winners' prizes) is a function of sales. Administrators allocate a portion of the sales in each draw, typically around 50% (~45-67%), to prizes. They then allocate a portion of the total allocation to prizes to each of the nine prize pools in proportions fixed by regulation, for which we have data. In each game-format, we use prize data to obtain prize-specific allocations and then use the regulatory data to estimate the total allocation to prizes. We then compare the week-averaged LaFleur's data and week averages of the total allocation to estimate the portion of sales within each draw allocated to prizes. We assume this portion remains fixed within each game-format. Finally, we use our estimates of the total allocation to prizes in each draw and share of sales allocated to prizes in each game-format to estimate the sales in each draw. We complete this exercise for each of the three games.

Third, we have lottery sales at the draw-game level from 2005 to 2018 at the national level for Mega Millions and Powerball from the Lotto Report (lottoreport.com). The data span three different formats of each game. We use these data to study one structural change in Mega Millions and two structural changes in Powerball. In addition, we employ it in our sub-jackpot analysis to accurately compute the jackpot expected value of Mega Millions and Powerball in each draw, as the risk of splitting the jackpot is increasing in national sales.

These data also include draw-level data for Louisiana and the U.S. Virgin Islands for Mega Millions only from October 2010 to August 2012. We use these data to maintain a constant composition of states within Mega Millions' national sales in the Powerball price change event analysis.

Fourth, we have lottery prize and probability data at the state-game-draw level from 1994 to 2017. We collected the data directly from state lottery commissions and the Multi-State Lottery Association as well as scraped online from "are your numbers lucky?" tools and results pages. The data include 16 unique games and span 41 states and D.C. In addition to Mega Millions and Powerball, we have Fantasy 5 and The Pick from Arizona; SuperLotto Plus from California; Lotto from Connecticut; Lotto, Lucky Money, and Mega Money from Florida; Lotto from Illinois; Lotto and Easy 5 from Louisiana; Classic Lotto and Rolling Cash 5 from Ohio; Lotto Texas from Texas;

and Lotto from Washington.

Only jackpot prizes and probabilities are used in the national demand elasticity analysis. Sub-jackpot prizes and probabilities for Mega Millions, Powerball, and SuperLotto Plus are used in the California draw sales imputation exercise, the Mega Millions and Powerball price elasticity and structural change analyses, and the sub-jackpot demand elasticity analysis. Jackpot prize amounts are reported as advertised when there is a difference between an estimate advertised and the actual prize awarded.

III Estimating Key Parameters for Optimal Lottery Design

III.A Introduction

In this section, we gather the empirical parameters needed to calibrate the optimal nationwide tax on sugar-sweetened beverages. First, we show how lottery expenditures vary by income, measuring s and $s(z)$. Second, we estimate bias γ, and how this varies by income. Third, we estimate the population prize semi-elasticity $\bar{\zeta}^a$. Fourth, we estimate the population price semi-elasticity $\bar{\zeta}^p$.

III.B Lottery Expenditures by Income

Appendix Figure A2 shows the histogram of expenditures. Overall approximately 40% of individuals do not spend any money on lotteries. The mean expenditure conditional on purchase is \$39.40 . The top 10 percent of spenders account for 60 percent of spending.

Figure 2 presents lottery ticket expenditures by income. Overall, the pattern of expenditures is mostly flat by income, although it features a modest hump at approximately median income. Households with incomes between \$40,000 and \$60,000 spend the most on lotteries. The lowest income individuals, those with incomes below \$20,000, spend the least on lotteries. Overall, these results indicate that lotteries do not appear to be regressive in the sense that they are purchased mostly by low-income individuals.

III.B.1 Causal income effects

In the survey we also obtained two measures of the causal effect of higher income on lottery expenditures. The cross-sectional variation of expenditures by income does not necessarily answer this question, since it is based both on the causal effect as well as on the relationship between preferences (or biases) and earnings ability.

Our first measure is obtained from a question that asked participants by what percent their lottery expenditures would increase if their income doubled. On average, participants stated that their lottery expenditures would increase by 2.08 percent (SE of 0.33), implying an income elasticity of 0.02. In other words, respondents anticipate that changes in their income would have essentially no effect on their lottery expenditures.

Our second measure is based on two questions. The first question asked respondents by what percent their household income increased relative to the prior year. The second question asked respondents by what percent their lottery expenditures increased relative to the prior year. A regression of change in expenditures on change in income produces a coefficient of 16.89 (SE 3.11), which implies an income elasticity of 0.17. Although this second measure may be confounded by changes in life circumstances that are related to changes in preferences for lottery expenditures, it is consistent with our first measure in implying a small causal income elasticity.

III.C Correlates of Lottery Consumption

In this subsection, we summarize reduced-form results about correlations between our various proxies of bias and tastes and lottery expenditures. In the next subsection we formalize by how these correlations can be used to produce estimates of bias.

In the analysis that follows, we let $\boldsymbol{b}_i$ denote a vector of indices measuring household i's bias, and let let $\boldsymbol{b}^V$ denote the value of $\boldsymbol{b}$ for a "normative" consumer who is fully knowledge and has full self-control. We assume that the normative benchmark $\boldsymbol{b}^V$ corresponds to correctly answering all financial literacy, numeracy, and statistical reasoning questions, to correctly predicting average lottery payout and the results of Lindqvist, Ostling, and Cesarini (forthcoming), and to having no self-reported self-control problems (i.e., reporting to playing the lottery as much as one would like).

III.C.1 Graphical summary

To summarize the relationship between our bias and proxies and lottery expenditures, we create an aggregate bias index of standardized bias proxy values. Specifically, for each bias proxy k, we define $\tilde{b}_{i,k} := \frac{b_{ik} - b_k^V}{\sqrt{Var[b_{ik}]}}$. That is, the standardized value $\tilde{b}_{i,k}$ is equal the difference between $b_{i,k}$ and the normative benchmark, normalized to have standard deviation equal to 1. We define the index $\bar{b}_i$ to equal the average of the standardized bias proxies: $\bar{b}_i := \frac{1}{K} \sum_{k=1}^{K} \tilde{b}_{i,k}$. The aggregate index $\bar{b}_i$ is equal to zero for an individual not subject to any biases. The index is positive for individuals who are on average biased toward over-purchasing lotteries. The index is negative for individuals who are on averaged biased toward under-purchasing lotteries—e.g., because they underestimate the expected monetary value of a lottery or underestimate the hedonic gains of winning. The average correlation between the bias proxies is 0.15, implying that each individual proxy contributes meaningfully to the index.

Figure 3 shows that on average, our bias proxies are strongly related to lottery expenditures. As can be seen in the figure, a one standard deviation increase in the bias index is associated with an increase of approximately 1 unit in ln(1 + lottery expenditure), meaning that it is associated with a more than doubling of lottery expenditures. Consistent with assumption 1, the lottery expenditures are approximately linear in the bias index.

Figure 4 shows that our bias proxies are also significantly related to income. Even though higher-income households have slightly higher lottery expenditures than lower-income households, they are significantly less biased. As seen in Figure 4, the bias proxy index differs by approximately 0.3 standard deviations between the highest- and lowest-income households.

III.C.2 Regression results

To formalize our results, we estimate

$$\ln(s_i + 1) = \boldsymbol{\tau} \boldsymbol{b}_i + \beta_a \boldsymbol{a}_i + \beta_x \boldsymbol{x}_i + \epsilon_i. \tag{9}$$

where $\boldsymbol{a}_i$ denotes the vector of preferences (fun of playing and anticipatory utility), $\boldsymbol{x}_i$ is the vector of household characteristics introduced in Table 1, and ϵ_i is the residual. In some of our specifications we replace the bias vector $\boldsymbol{b}_i$ by the standardized index $\bar{b}_i$ described in Section III.C.1, and we vary the inclusion of controls.

Table 2 presents estimates of Equation (9). Column 1 separately quantifies the relationship between each proxy and lottery expenditures, while controlling for income, education and normative preferences, as well as age, gender, and race. Collectively, the variables in this specification explain approximately 50% of the variation in self-reported lottery expenditures. Column (2) examines the stability of our bias proxy coefficients by removing all of the controls. In this specification the R^2 statistics drops to approximately 10%, and on net the coefficients on the bias proxies become somewhat more positive.

To provide transparent and formal statistical evidence for the explanatory power of our bias proxies, we aggregate them into the standardized index $\bar{b}_i$. As shown in column (3), the standardized bias index is highly significant: a one standard deviation in the index is associated with an increase of 0.54 in the logarithm of lottery expenditures (se 0.06). Removing the demographic controls has little effect on this coefficient, although it does have a moderate effect on the R^2 statistic, as seen in column (4). Column (5) shows that removing controls for the entertainment of utility of lotteries, does, however have a significant effect on the estimated effect of the bias index. The estimated coefficient on the bias index in column (5) is similar to the estimated coefficient on the bias index in column (6), where we include no controls. This implies that our bias proxies are largely explaining variation that is distinct from the variation explained by demographic controls.

Finally, column (7) formalizes the graphical evidence in Figure 2 that lottery expenditures are moderately increasing in income. A comparison of columns (1) and (6) shows that this relationship cannot be explained by other covariates such as education, financial literacy, or other demographics. Controlling for demographics and survey measures of bias only strengthens the positive relationship between income and lottery expenditures.

III.D Estimates of Bias in Lottery Expenditure

In section I we defined bias γ the price reduction that produces the same change in demand for lotteries as the bias does. Our *counterfactual normative consumer* empirical strategy directly implements this definition, using an approach that builds on Bronnenberg et al. (2015), Handel and Kolstad (2015), Allcott, Lockwood, and Taubinsky (2019) and other work.[5] Our formulation and implementation of this strategy follows Allcott, Lockwood, and Taubinsky (2019). The process is to use the estimated relationship between bias proxies and quantity consumed and use that relationship to predict the counterfactual quantity that would be consumed if consumers instead maximized normative utility. In section IV we combine these quantity estimates of bias with our estimates of prize and price elasticities to produce estimates of γ.

To formalize the approach, recall that money-metric bias γ is defined to satisfy $s(p,\theta) = s^V(p-\gamma,\theta)$. Log-linearizing this equation, and using an i subscript for each survey respondent in the data gives

$$\ln s_i = \ln s_i^V + \zeta_i^p \gamma_i, \tag{10}$$

where $\ln s_i^V$ denotes the log of the quantity that household i would consume in the absence of bias, s_i is reported consumption, and ζ_i is the price semi-elasticity of demand. As an example, imagine that bias increases quantity demanded by 10%, while a 1 dollar decrease in the price increases demand by 50%. Then the impact of bias on consumption is the same as a \$0.20 price reduction: $\gamma_i = 0.20$ dollars.

A limit to utilizing (10) is that it produces an estimate of bias for a particular set of lottery attributes. Thus, while the estimated bias from equation (10) can be used to determine the optimal price holding attributes a constant, it cannot be used to simultaneously determine the optimal attributes and the optimal price. In Section IV we therefore make additional parametric assumptions about the structure of bias to extrapolate how it changes when lottery attributes change.

Assumption 1. *Linearity:* $\zeta_i \gamma_i = \boldsymbol{\tau} \cdot (\boldsymbol{b}^V - \boldsymbol{b}_i)$*, where* $\boldsymbol{\tau}$ *comprises two parameters scaling the effects of nutrition knowledge and self-control.*

Assumption 2. *Unconfoundedness:* $\boldsymbol{b}_i \perp (\ln s_i^V | \boldsymbol{a}_i, \boldsymbol{x}_i)$.

In words, bias is conditionally independent of normative consumption. While such unconfoundedness assumptions are often unrealistic, this is more plausible in our setting because of our

[5] Bartels (1996), Cutler et al. (2015), Handel and Kolstad (2015), Johnson and Rehavi (2016), and Levitt and Syverson (2008) similarly compare informed to uninformed agents to identify the effects of imperfect information. All of these papers require the same identifying assumption: that preferences are conditionally uncorrelated with measures of informedness.

tailor-made survey measures of lottery entertainment utility. Equation 10 and Assumptions 1 and 2 imply the estimating equation (9).

Inserting our parameter estimates into Equation 10 and Assumption 1, we obtain estimates of counterfactual normative consumptions:

$$\ln \hat{s}_i^V = \ln s_i - \hat{\boldsymbol{\tau}}(\boldsymbol{b}^V - \hat{\boldsymbol{b}}_i) \tag{11}$$

$$\hat{\gamma}_i = \hat{\boldsymbol{\tau}}(\boldsymbol{b}^V - \hat{\boldsymbol{b}}_i)/\hat{\zeta}_i. \tag{12}$$

Under our assumptions, the statistics $\hat{\boldsymbol{\tau}}(\boldsymbol{b}^V - \hat{\boldsymbol{b}}_i)$ is an estimate of the effect of bias on lottery expenditures. This different from the money-metric measure of bias, which is given by $\hat{\boldsymbol{\tau}}(\boldsymbol{b}^V - \hat{\boldsymbol{b}}_i)/\hat{\zeta}_i$.

Using our regression results and equation (11) we find that on average, bias inflates lottery expenditures by 9 percent. Figure 5 plots how the share of expenditures attributable to bias varies by income. Predicted overconsumption is much larger for low-income households: it is 0.19 and 0.05 percent, respectively, for households with incomes below $20,000 and above $100,000.

Figure 6 quantifies the contribution of each bias proxy k to lottery consumption: $\hat{\tau}_i(b_k^V - \hat{b}_{i,k})$. These estimates are not simply proportional to the estimated coefficients $\hat{\tau}_i$ in Table 2 because the contribution of any proxy to overconsumption depends not only on the coefficient, but its deviation from the normative benchmark. For example, while our measure of overconfidence is significantly and positively related to lottery expenditures in Table 2, its overall contribution to overconsumption is zero because individuals are not overconfident on average. Instead, some are overconfident and others are underconfident.

III.E Jackpot Elasticity

III.E.1 Background

In this section, we estimate the semi-elasticity of lottery demand with respect to jackpots, $\bar{\zeta}_1^a$. Figure 7 illustrates the key identifying variation for the two large multi-state lotto games, Mega Millions and Powerball, in 2014. In most drawings, nobody wins the jackpot, so a predetermined share of revenues are added to the prize pool and it rolls over to the next drawing. When somebody does win the jackpot, the jackpot is split equally between all winners, and the jackpot returns to a "reset value" in the next drawing.

Since the winning number is randomly drawn, there is randomness in whether the jackpot is won, and thus randomness in the size of the next jackpot. Our strategy exploits this randomness, while addressing subtle endogeneity issues and considering substitution across games and time.

Figure 9 shows that sales are highly correlated with jackpot sizes. The figure is a binned scatter plot of the natural log of weekly ticket sales on the jackpot pool, controlling for quarter-of-sample fixed effects and game-state-format fixed effects. This figure suggests that the semi-elasticity is strongly positive. However, the slope is not our estimate of the semi-elasticity, as it is affected by simultaneity bias: the period t demand shock affects ticket sales (and thus the probability of a win) as well as the the jackpot pool size. Furthermore, these demand shocks might be serially autocorrelated. Our estimation strategy is designed to address these issues by exploiting the random part of the lotto drawing conditional on the best available prediction of demand.

III.E.2 Estimation Strategy

In this section, we index time (drawings or weeks) by t and lottery games by j. We let s_{jt} denote total sales for game j at time t, and we let w_{jt} denote the jackpot amount for observation jt, both in units of millions of dollars.

The jackpot size at time t depends on whether the prize rolls over after drawing $t-1$. We define a rollover indicator r_g as

$$r_{jt-1} = \begin{cases} 1 & \text{if } j\text{'s jackpot rolls over at } t-1 \\ 0 & \text{if } j\text{'s jackpot is won} \end{cases}. \tag{13}$$

The rollover probability $\mathbb{E}\left[r_{jt-1}\right]$ depends on ticket sales s_{jt-1}, but the realization is random *conditional* on s_{jt-1}.

The rollover realization is highly predictive of the actual jackpot w_{jt}. We define a jackpot forecast z_{jt} as

$$z_{jt} = \begin{cases} (1+\rho_j)w_{jt-1} & \text{if } r_{jt-1}=1, \text{ where } w_{jt} = (1+\rho_j)w_{jt-1} + \nu_{jt},\ E[\nu_{jt}]=0 \\ \bar{w}_j & \text{if } r_{jt-1}=0 \end{cases} \tag{14}$$

where $\bar{w}_j$ is the jackpot's reset value and ρ_j is the expected increase in the jackpot with each rollover. Here again, $\mathbb{E}\left[z_{jt}\right]$ depends on demand shocks and other information available at $t-1$, but z_{jt} is random conditional on that.

We then construct a control $f_j(H_{t-1})$ as the flexible best predictor of log sales given $H_{t-1} = (s_{jt-1}, s_{jt-2}, w_{jt-1}, ...)$, all information about sales and prize pools available in the previous period.

Define ξ_j as a vector of game fixed effects, and $\phi_{q(t)}$ as a vector of quarter-of-sample fixed effects. We estimate the semi-elasticity of demand with respect to the jackpot using the following regression:

$$\ln s_{jt} = \bar{\zeta}^a \pi_j w_{jt} + f(H_{jt-1}) + \xi_j + \phi_{q(t)} + \epsilon_{jt}, \tag{15}$$

instrumenting for w_{jt} with the jackpot forecast z_{jt}. Abstracting away from substitution across time and games, this delivers an unbiased estimate of $\bar{\zeta}^a$ under the assumption that $z_{jt} \perp \epsilon_{jt} | H_{jt-1}$.

III.E.3 Results

Appendix Figures A4 and A5 show the first stage and reduced form of our instrumental variables (IV) estiamtes of Equation (15). There is a strong relationship between the jackpot forecast and the actual jackpot, as well as between the jackpot forecast and sales.

Table 3 presents estimates of Equation (15). Column 1 presents our primary IV estimates with the richest specification of $f(H_{jt-1})$, including four lags and quadratic terms. Column 2 presents estimates with less rich controls: only two lags in H and no quadratic terms. The fact that the coefficient moves little between columns 1 and 2 suggests that any additional controls in H would have little impact. Column 3 presents OLS results. The coefficient is slightly (although not statistically significantly) larger, consistent with a moderate upward bias from positive demand shocks causing larger sales and larger jackpots.

Our model in Section I and our estimates so far consider one lottery in a static framework, so there is no substitution across time or across games. In reality, one might worry about dynamics, for example that buying tickets when jackpots are large then causes people to tire of lotteries and be less likely to buy tickets when the jackpot resets. Furthermore, one might wonder whether people substitute across games, for example buying more scratch-offs when lotto jackpots are low, and fewer scratch-offs when lotto jackpots are high.

We first consider dynamics. To do so, we repeat estimates of Equation (15), except including four additional lags of the jackpot expected value. We include one instrument for each lag $s \in \{1, 2, 3, 4\}$, constructed using the jackpot from the period before the earliest lag, w_{jt-5}, and rollovers r_{jt-s} between $t-5$ and $t-s$. The control $f(H_{jt-5})$ includes only sales and jackpot pools from before the earliest lag. As shown in Appendix Table A1, the jackpot forecast instrument for lag $t-s$ strongly predicts the actual jackpot for lag $t-s$, but does not strongly predict the jackpots for other lags.

Table 4 presents results. Columns 1-5, respectively, present estimates with four, three, two, one, and zero additional lags. The contemporaneous coefficient is much stronger than any of the lag coefficients, and the Akaike and Bayesian Information Criteria are minimized in column 5, the specification with zero lags. This suggests that any dynamic effects are quite limited.

We now consider substitution between lottos and other games. To do this, we regress sales levels for different game types in the same state and week on the jackpot expected value for a given lotto game, using the same IV strategy and controls as in Equation (15). Table 5 presents results. In column 1, the dependent variable is sales for the same lotto game. This parallels the results in Table 3, except that the dependent variable is in levels and not logs. In column 2, the dependent variable is sales for all other games in that state and week. The ratio of coefficients in column 2 to column 1 is a diversion ratio: the dollars of expenditures that are diverted from all other games due to an exogenous \$1 increase in expenditures on the average lotto game.

Different types of games might be more or less substitutable with lotto games. In particular, one might hypothesize that in states with multiple types of lotto games, the different lotto games are particularly close substitutes for each other. Columns 3 and 4 present estimates with all other lotto games and all instant (scratch-off) games as the dependent variables. In all of columns 2-4, we see no statistically detectable substitution toward or away from other games in the same state as lotto jackpots vary.

The implication of Tables 4 and 5 is that it may be reasonable to consider individual lotto games in estimation, with the prize elasticities as estimated above.

III.F Sub-Jackpot Prize Elasticity

III.F.1 Background

In this section, we estimate the semi-elasticity of lottery demand with respect to "sub-jackpot" prizes $\bar{\zeta}^a_{2+}$. The identification largely follows our approach from the previous section, but there is an additional challenge: in most lotto games, the jackpots vary over time, but other prizes are fixed. To address this challenge, we exploit California's unusual parimutuel rule. In 1996, a California Supreme Court ruling drew a distinction against "lotteries," where players play against other players, and "house-banked games," where players play against the house. The California Lottery Act allows only the former. As a result, all California lotto prize levels have parimutuel pools that roll over to the next draw if they are not won. Furthermore, Mega Millions and Powerball have California-specific parimutuel sub-jackpot prizes.

Figure 8 illustrates the key identifying variation for Powerball in 2014. The figure illustrates how the second prize amount evolves independently of the jackpot amount. Furthermore, while most of the expected value of a ticket is determined by the jackpot, the second prize pool can also contribute substantially to the overall expected value and even exceed the jackpot's contribution, as it did in June and July.

III.F.2 Estimation Strategy

Our estimating equation is very similar to Equation (15), except that we add the second prize expected value variable. We estimate

$$\ln s_{jt} = \bar{\zeta}^a \pi_j w_{jt} + \bar{\zeta}^a_2 \pi_j w_{jt} + f(H_{jt-1}) + \xi_j + \phi_{q(t)} + \epsilon_{jt}, \tag{16}$$

instrumenting for the jackpot and second prize expected value with the prize forecast instruments as described above.

III.F.3 Results

Table 6 presents estimates of Equation (16), in a format paralleling the jackpot semi-elasticity estimates from Table 3. If consumers were risk-neutral, they would respond equally to variation in expected value coming from the jackpot versus the second prize, and the two coefficients in each column would be equal. In reality, we estimate that consumers are much less responsive to variation in expected value from the second prize than from the jackpot. Appendix Figure A6 presents a binned scatterplot of the reduced form of this regression, confirming no visible relationship between sales and the second-prize expected value instrument.

This result could arise for several reasons. First, probability weighting could cause consumers to perceive that the jackpot probability is a larger proportion of its true probability than for the more likely lower prizes. Second, consumers could derive larger anticipatory utility from larger prizes, and anticipatory utility might be insensitive to probabilities. Third, the jackpots may be more heavily advertised and promoted, inducing consumers to pay more attention.

III.G Price Elasticity

III.G.1 Background

In this section, we estimate the semi-elasticity of lottery demand with respect to price, $\bar{\zeta}^p$. The challenge here is that unlike prizes, prices change little over time: tickets for most lotto games cost \$1. We exploit two price changes as event studies to estimate price elasticity. In the first event study, the Powerball ticket price increased from \$1 to \$2 in January 2012. In addition to this price change, the probabilities rose by a ratio of 39/35 for some prizes, and the jackpot reset value doubled to \$40 million. In the second event study, the Mega Millions price increased from \$1 to \$2 in October 2017. This event was also accompanied by a change in structure, wherein the odds of winning some of the prizes (those that require matching the "Mega Ball" value, five of the nine prize levels) increased by about 20%, whereas the odds of winning other prize levels decreased by a similar amount, shifting the overall prize allocation toward the jackpot.

Panels (a) and (b) of Figure 10 shows the natural log of weekly ticket sales for Mega Millions and Powerball for the two event studies. In each case, the raw data seem to suggest that ticket sales drop after the price change, relative to the "control" game whose price does not change at that time. There is no large apparent increase in sales of the other game, suggesting that substitution is negligible, so we use the game whose price does not change as a control.

Our empirical strategy exploits the raw data in this figure, except adding controls for the jackpot pool amount. These controls both eliminate noise from the time-varying jackpot and also control for changes in the jackpot amounts induced by the reset value change in the first event study.

III.G.2 Empirical Strategy

Define p_{jt} as the purchase price of game j tickets at time j, and define $\hat{\zeta}_2 EV_{jt}^{2+}$ as the predicted effect of the sub-jackpot prize on sales, as estimated in Table tk. Define $\eta_{d(t)}$ as a day-of-week fixed effect. We estimate the following regression:

$$\ln s_{jt} - \hat{\zeta}_2 EV_{jt}^{2+} = -\zeta_p p_{jt} + \zeta_1 w_{j1t} + f(H_{jt-1}) + \xi_j + \eta_{d(t)} + \phi_{q(t)} + \epsilon_{jt}, \tag{17}$$

instrumenting for the jackpot w_{j1t} using the jackpot forecast IV.

For our econometric purposes, these price changes would ideally have happened in isolation, without also changing the format. The jackpot pool value control w_{j1t} and the adjustment for sub-jackpot prizes $\hat{\zeta}_2 EV_{jt}^{2+}$ control for contemporaneous format changes. In some specifications, we replace the jackpot pool control with the simple jackpot expected value $\pi_{j1t} w_{j1t}$. Appendix A.A considers a third event study in which Powerball changed its format in 2015 without changing the purchase price. Figure A9 and A2 show that the jackpot pool control explains purchase changes after the format change better than the jackpot expected value control. This is why we use the jackpot pool control in our primary specifications, although it makes little difference for the estimate of ζ_p.

III.G.3 Results

Figure 11 presents the event studies after residualizing on the covariates (other than price) in Equation (17). In both cases, the price increase from \$1 to \$2 appears to cause sales to drop by about 50 log points relative to the other multi-state game.

Table 7 presents estimates of Equation (17). Columns 1 and 2 combine the two event studies, using the jackpot pool control or the jackpot expected value control, respectively. Columns 3 and 4 consider each event study individually. The estimate of ζ_p between -0.5 and -0.6 imply that a price increase of \$0.10 reduces demand by 5-6 percent.

IV Simulations

IV.A Setup

We now turn to the question of welfare and counterfactual policy questions. Progress here requires calibrating a structural model, for although the optimal policy conditions derived in Proposition 1 must hold at the optimum, the statistics on which they depend are endogenous to policy.

We begin by assuming a functional form for the decision utility that rationalizes consumers' observed behavior. As in Section I, a lottery is characterized by a ticket price p, a vector of probabilities π_k, and a vector of prizes aw_k, with $\sum_k \pi_k w_k = 1$ so that a represents the expected value of a purchased lottery ticket. For our baseline simulations, we focus on a setting in which a single representative lottery game is available, and we compute social welfare across a range of values of price and expected values.

As in Section (I), the decision about whether to purchase a lottery ticket (on a particular choice occasion) is binary. A consumer of type θ perceives a utility gain from playing the lottery of

$\sum_k (1+\omega_k(\theta))\, \pi_k u\,(aw_k;\theta)$, where $\omega_k(\theta)$ denotes the decision weight applied to prize k by consumer type θ, and u denotes the utility gain from an increase of aw_k in the consumer's continuation wealth, normalized by the marginal value of consumption.[6] Decision weights here are "reduced form" in the sense that we do not attempt to understand the mapping between the winning probability π_k and the weight $\omega_k(\theta)$. Rather, we assume that the total decision weight $(1+\omega_k(\theta))\pi_k$ is stable across variations in prizes and prices, and we consider only the latter sources of variation in our policy counterfactuals.

The consumer makes a binary choice, purchasing a ticket if doing so generates higher utility than the outside option of no purchase. Since our baseline assumption is that labor supply decisions are not affected by lottery policy, we can treat each agent's choice of earnings $z(\theta)$ as exogenous, and focus solely on their decision to purchase a lottery ticket. Formally, we can write the consumer's problem as maximizing utility by selecting $s \in \{0,1\}$ to maximize the decision utility function

$$U(s;\theta,\varepsilon) = z(\theta) - T(z(\theta)) - \psi(z(\theta)) + s \cdot \left(-p + \sum_k (1+\omega_k(\theta))\, \pi_k u(aw_k;\theta) + \alpha\varepsilon\right) \tag{18}$$

The term ε is an iid extreme value distributed preference shock, and α controls the elasticity of demand. We define $\tilde{U}(a,p;\theta) = -p + \sum_k (1+\omega_k(\theta))\, \pi_k u(aw_k;\theta)$ to denote the representative utility that type θ derives from a lottery ticket. Then the share of type θ consumers who purchase a lottery ticket (or, equivalently, the long-run average per-period consumption of lottery tickets among θ-types) can be written in closed form, thanks to the extreme value distribution of the ε shocks:

$$\bar{s}(p,a;\theta) = \frac{\exp\left(\tilde{U}(\theta)/\alpha\right)}{1+\exp\left(\tilde{U}(\theta)/\alpha\right)}. \tag{19}$$

Similarly, the average decision utility of θ-type agents, aggregating across the shocks ε, has an analytic solution:

$$\begin{aligned}\bar{U}(a,p;\theta) &= E\left[z(\theta) - T(z(\theta)) - \psi(z(\theta)) + s \cdot \left(-p + \sum_k (1+\omega_k(\theta))\, \pi_k u(aw_k;\theta) + \alpha\varepsilon\right)\right] \\ &= z(\theta) - T(z(\theta)) - \psi(z(\theta)) + \alpha \ln\left(1+\exp\left(\tilde{U}(a,p;\theta)/\alpha\right)\right) \end{aligned} \tag{20}$$

The expectation in Equation (20) need not represent normative welfare. Although the decision weights $\omega_k(\theta)$ may be nonzero for normatively valid reasons, such as anticipatory utility, they may alternatively (or additionally) be nonzero due to behavioral biases. To flexibly account for this

[6] Formally, if $\mathcal{U}(W;\theta)$ is the utility that type θ derives from continuation wealth W, and $y(\theta)$ denotes θ's continuation wealth if the prize is not won, then $u(x;\theta) = \frac{\mathcal{U}(y(\theta)+x)-\mathcal{U}(y(\theta))}{\mathcal{U}'(y(\theta))}$.

distinction, we let $\phi_k(\theta)$ denote the share of $\omega_k(\theta)$ which is a normative mistake or bias. Thus we can write normative utility as

$$V(s;\theta,\varepsilon) = z(\theta) - T(z(\theta)) - \psi(z(\theta)) + s \cdot \left(-p + \sum_k \left(1 + \left(1 - \phi_k(\theta)\omega_k(\theta)\right)\pi_k u(aw_k;\theta) + \alpha\varepsilon\right)\right.$$
$$= U(s;\theta,\varepsilon) - \sum_k \phi_k(\theta)\omega_k(\theta)\pi_k u(aw_k;\theta) \tag{21}$$

In the notation of Section (I), $\gamma(a,\theta)$ is equal to the second term in Equation (21). Combining Equations (20) and (21), average normative utility among θ-types can be written

$$\bar{V}(a,p;\theta) = \bar{U}(a,p;\theta) - \sum_k \phi_k(\theta)\omega_k(\theta)\pi_k u(aw_k;\theta)$$

We treat labor supply as separable from lottery consumption, so lottery design does not affect the governments income tax receipts and labor supply decisions can be ignored when considering optimal lottery design. However, earnings are relevant for the policymaker through their effect on the consumer's marginal utility of consumption. As a result, a policymaker may wish to redistribute from higher to lower earning consumers at the margin. We account for this policy motive by using reduced form Pareto weights $\mu(\theta)$, which can be interpreted as the social marginal welfare weights, taking as given the prevailing income tax. As a result, optimal lottery policies can be found by maximizing a narrowed representation of social welfare, which excludes the exogenous utility from non-lottery consumption and labor disutility. Formally, the optimal lottery maximizes

$$W(a,p) = \int_\theta \mu(\theta)\bar{V}(a,p;\theta)dF_\theta(\theta) + \lambda\bar{s}(p,a)(p-a), \tag{22}$$

where λ represents the marginal value of lottery revenues for public funds.

IV.B Calibration

We impose a number of parametric assumptions in order to calibrate this model. First, we assume that all consumers have homogeneous CRRA utility over continuation wealth, with coefficient of relative risk aversion σ, implying that θ's utility gain from winning x, normalized by marginal utility of continuation wealth, is $u(x;\theta) = \frac{y(\theta)}{1-\sigma} \cdot \left[\left(1 + \frac{x}{y(\theta)}\right)^{1-\sigma} - 1\right]$, where $y(\theta)$ represents the θ type's continuation wealth (absent any prize winnings).

We assume five distinct types of consumers, corresponding to the income bins displayed in Figure 2, and we assume consumer are homogeneous within these bins, aside from the preference shocks ε across which we have aggregated above. We multiply annual income by 20 to approximate expected lifetime continuation wealth. We calibrate marginal social welfare weights $\mu(\theta)$ to be proportional to $1/z(\theta)$, where $z(\theta)$ denotes average income within each bin; these weights are normalized so that they average to one across the population.

We assume that the jackpot decision weights $\omega_1(\theta)$ vary across types, and that the weights on all smaller prizes are the same within bins, denoted $\omega_{2+}(\theta) = \kappa\omega_1(\theta)$, with the proportion κconstant across types. In total, we have eight positive parameters to estimate in the demand model: the five type-specific jackpot decision weights $\omega_1(\theta)$, the smaller prize decision weight proportion κ, the continuation utility curvature σ, and the demand elasticity parameter α. These are exactly identified by calibrating to match eight empirical moments: average lottery consumption within each income bin, and the estimated elasticity of demand with respect to jackpot size, smaller prize size, and prices.

To calibrate the degree of behavioral bias, we assume a portion of each consumer's decision weights are due to bias. The bias share, denoted $\phi(\theta)$, is allowed to vary across consumer types, but is assumed to be constant across prize levels within each type. We compute policy counterfactuals for a range of assumptions about bias, including the case where $\phi(\theta)$ is adjusted to match our estimated quantity effect of bias across income in accordance with Figure 5.

We calibrate the model assuming a single representative lottery game, based on a current MegaMillions lottery ticket with a price of \$2 and a jackpot pool of \$300 million. We treat the policy choice of lottery expected value as a univariate decision, incorporating both the implicit tax rate and expected explicit income taxes paid on winnings. Therefore we include both factors in expected value parameter a, and we assume an income tax rate on lottery winnings of 40%. This leads to a status quo expected value of $a = 0.74$.

IV.C Results

The primary results of our calibrations are reported in Figure 12. Both panels display plots of social welfare (relative to a benchmark with no available lottery) across a range of possible lottery parameters. Panel (a) varies the expected value of the lottery by scaling up or down all prizes proportionally, holding fixed the status quo price of the lottery. Panel (b) holds fixed status quo prices, and instead varies the price of the lottery ticket. In both cases, we display results for three different assumptions about bias.

The first case assumes all observed demand is fully normative, that is, $\phi(\theta) = 0$ for all types. In this case it is unsurprising that lotteries generate surplus, as the usual revealed preference logic implies that consumers are made better off by the option to purchase a good from which they derive utility. Nevertheless, welfare is not maximized at the "laissez faire" benchmark of $a = p$ with zero lottery tax. Rather, the optimal implicit lottery tax is positive. This is due to the redistributive motive. Since lottery consumption is lower among the lowest earning consumers (Figure 2), imposing an implicit tax is a useful way to raise revenue, which could be refunded (for example) through a lump sum rebate, generating an overall improvement in progressivity. The second line in 12, panel (a), shows that this overall qualitative assessment remains true even after incorporating our calibrated estimates of behavioral bias. Although the total welfare generated by the lottery is somewhat lower, it remains positive. Indeed, welfare remains positive at the status quo expected value even assuming that the quantity effect of bias is four times as large as our estimates suggest, as evident in the third line.

Panel (b) of Figure 12 shows a similar plot of estimated welfare across a range of ticket prices, holding fixed the status quo expected value. Here too, welfare is naturally highest in the specification where all demand is assumed to be normative, but remains positive even when bias is substantial. In all specifications, welfare declines to zero as the price grows large, reflecting the fact that a sufficiently high price discourages nearly all consumption and thus resembles a market with no lottery. Very low prices generate substantial welfare when demand is normative, reflecting the fact that consumers the fiscal cost to the government is offset by the transfer received by consumers, who derive real utility greater than the marginal social cost of providing the lottery, e.g., due to anticipatory utility. In contrast, in the presence of bias, welfare falls as prices decline. This reflects that low prices induce many consumers to purchase lottery tickets even when their net normative utility from that purchase decision is negative.

Finally, it is possible to jointly solve for the optimal expected value and price. We perform this computation using a grid search, and we find the optimal price of a lottery ticket is \$0.74, and the optimal expected value (rescaling all prizes in fixed proportion) is 0.36. Together, these result in an implicit tax rate of 52%, remarkably close to the prevailing design of current lotto games like PowerBall and Mega Millions.

V Conclusion

There is a long-standing debate as to whether these lotteries are a regressive "tax on people who are bad at math" or a "win-win" that generates both consumer surplus and government revenues. In this paper, we derive new formulas that deliver optimal prices and attributes for a government-regulated good as a function of a set of sufficient statistics. We then provide new descriptive evidence on lottery consumption, behavioral biases, and demand elasticities. We find that individual-level lottery expenditures are highly correlated with survey measures of innumeracy and poor statistical reasoning, but our observable measures of behavioral bias statistically explain only about 9 percent of lottery purchases for the average household. Using these empirical moments, we calibrated a model that predicts that lotteries are indeed a welfare-improving "win-win," and the socially optimal ticket price and implicit tax are similar to the current norms in U.S. states. There are many caveats to our results, and we think of this paper as just a first step toward a new literature studying lotteries through the lens of optimal taxation.

References

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *Quarterly Journal of Economics* 134 (3):1557–1626.

Allcott, Hunt and Dmitry Taubinsky. 2015. "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market." *American Economic Review* 105 (8):2501–2538.

Banks, James and Zoe Oldfield. 2007. "Understanding Pensions: Cognitive Function, Numerical Ability and Retirement Saving." *Fiscal Studies* 28 (2):143 – 170.

Bartels, Larry M. 1996. "Uninformed Votes: Information Effects in Presidential Elections." *American Journal of Political Science* 40 (1):194–230.

Bronnenberg, Bart, Jean-Pierre Dubé, Matthew Gentzkow, and Jesse Shapiro. 2015. "Do Pharmacists Buy Bayer? Sophisticated Shoppers and the Brand Premium." *Quarterly Journal of Economics* 130 (4):1669–1726.

Brunnermeier, Markus K. and Jonathan A. Parker. 2005. "Optimal Expectations." *American Economic Review* 95 (4):1092–1118.

Caplin, Andrew and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *The Quarterly Journal of Economics* 116 (1):55 – 79.

Clotfelter, Charles T., Philip J. Cook, Julie A. Edell, and Marian Moore. 1999. "State Lotteries at the Turn of the Century: Report to the National Gambling Impact Study Commission." Tech. rep., National Gambling Impact Study Commission.

Conlisk, John. 1993. "The Utility of Gambling." *Journal of Risk and Uncertainty* 6:255–275.

Cutler, David, Amber Jessup, Donald Kenkel, and Martha Starr. 2015. "Valuing Regulations Affecting Addictive or Habitual Goods." *Journal of Benefit Cost Analysis* 6 (2):247–280.

Gottlieb, Daniel. 2014. "Imperfect Memory and Choice Under Risk." *Games and Economic Behavior* 85:127–158.

Gruber, Jon and Emmanuel Saez. 2002. "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics* 84 (1):1–32.

Gruber, Jonathan and Botond Kőszegi. 2001. "Is Addiction 'Rational'? Theory and Evidence." *Quarterly Journal of Economics* 116 (4):1261–1303.

Handel, Benjamin R and Jonathan T Kolstad. 2015. "Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare." *American Economic Review* 105 (8):2449–2500.

Handel, Benjamin R., Jonathan T. Kolstand, and Johannes Spinnewijn. 2016. "Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets." NBER Working Paper No. 21759.

Johnson, Erin M. and M. Marit Rehavi. 2016. "Physicians Treating Physicians: Information and Incentives in Childbirth." *American Economic Journal: Economic Policy* 8 (1):115–141.

Kahneman, Daniel, Alan B. Krueger, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2006. "Would You Be Happier If You Were Richer? A Focusing Illusion?" *Science* 312:1908–1910.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica* 47 (2):263 – 292.

Kearney, Melissa Schettini. 2005. "State lotteries and consumer behavior." *Journal of Public Economics* 89:2269 – 2299.

Levitt, Steven D. and Chad Syverson. 2008. "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions." *Review of Economics and Statistics* 90 (4):599–611.

Lindqvist, Erik, Robert Ostling, and David Cesarini. forthcoming. "Long-run Effects of Lottery Wealth on Psychological Well-being." *Review of Economic Studies* .

Loewenstein, George. 1987. "Anticipation and the Value of Delayed Consumption." *The Economic Journal* 97 (387):666–684.

Lusardi, Annamaria and Olivia S. Mitchell. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence." *American Economic Review* 52 (1):5 – 44.

Rees-Jones, Alex and Dmitry Taubinsky. forthcoming. "Measuring Schmeduling." *Review of Economic Studies* .

Saez, Emmanuel. 2002. "The Desirability of Commodity Taxation Under Non-Linear Income Taxation and Heterogeneous Tastes." *Journal of Public Economics* 83 (2):217–230.

Saez, Emmanuel, Joel Slemrod, and Seth H Giertz. 2012. "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review." *Journal of Economic Literature* 50 (1):3–50.

Steiner, Jakub and Colin Stewart. 2016. "Perceiving Prospects Properly." *American Economic Review* 106 (7):1601–1631.

Taubinsky, Dmitry and Alex Rees-Jones. 2018. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Review of Economic Studies* 85 (4):2462–2496.

Tversky, Amos and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5:297 – 323.

Woodford, Michael. 2012. "Inattentive Valuation and Reference-Dependent Choice." *Working Paper* .

Table 1: **Descriptive Statistics**

Panel A: Survey Demographics

	Qualtrics sample	US population
Income ($000s)	55.69	44.42
College	0.30	0.30
Male	0.39	0.48
White	0.78	0.78
Black	0.05	0.10
Age	50.97	49.29
Monthly lottery expenditure	21.96	24.19

Notes: This table provides a summary of the demographics of our Qualtrics Panel (N=2,767).

Panel B: Survey Responses to Bias and Preference Questions

	Obs.	Mean	Std. dev.	Min	Max
Lottery seems fun	2,767	0.16	0.65	-1.00	1.00
Enjoy thinking about winning	2,767	0.39	0.61	-1.00	1.00
Self-control problems	2,767	-0.04	0.32	-1.00	1.00
Financial illiteracy	2,767	0.25	0.25	0.00	1.00
Expected value miscalculation	2,767	0.61	0.37	0.00	1.00
Innumeracy	2,767	0.35	0.32	0.00	1.00
Gambler's fallacy / law of small numbers	2,767	0.21	0.34	0.00	1.00
Non-belief in law of large numbers	2,767	0.40	0.18	0.02	0.78
Overconfidence	2,767	-0.00	0.49	-4.95	4.95
Expected payout	2,767	24.82	18.60	5.00	95.00
Curvature	2,767	21.94	38.24	-100.00	1,000.00

Notes: This table summarizes participants' responses to our proxies for bias and entertainment utility of lotteries. Section II.A summarizes the coding of these variables.

Panel C: Lottery Sales and Prizes

	Years	Geo.	Obs.	Mean	Std. dev.	Min	Max
LaFleur's sales ($)	1994-2017	State	322,848	3.90	9.42	0.00	237.19
Jackpot pool ($)	1994-2017	State	125,971	64.03	80.29	0.05	1,500.00
Jackpot odds	1994-2017	State	125,971	142.62	87.08	0.44	302.58
Lotto Report sales ($)	2005-2018	National	2,209	31.85	45.91	10.55	1,270.21
Imputed sales ($)	2003-2017	CA	3,202	5.09	5.91	1.66	176.58
2nd prize pool ($)	2003-2017	CA	3,202	0.42	0.70	0.03	7.66
2nd prize odds	2003-2017	CA	3,202	5.52	5.71	1.59	18.49

Notes: All means, standard deviations, minimums, and maximums are reported in millions. Variables for which the geographic level is "State" have coverage over 41 states and D.C. We use LaFleur's sales and jackpot pool and odds data in our national demand elasticity analysis and Lotto Report and imputed sales and second prize pool and odds data in our price and format change analyses as well as our California-level sub-jackpot demand elasticity analysis.

Table 2: **Regressions of Monthly Lottery Expenditure on Bias Proxies**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Bias index			0.540***	0.458***	1.004***	1.082***	
			(0.063)	(0.061)	(0.080)	(0.076)	
Financial illiteracy	0.151***	0.254***					
	(0.030)	(0.038)					
Expected value miscalculation	0.027	0.032					
	(0.027)	(0.036)					
Innumeracy	0.044	0.069*					
	(0.030)	(0.039)					
Gambler's fallacy / law of small numbers	0.000	-0.003					
	(0.026)	(0.034)					
Non-belief in law of large numbers	0.032	0.095***					
	(0.024)	(0.032)					
Overconfidence	0.059**	0.106***					
	(0.024)	(0.031)					
Expected payout	0.082***	0.269***					
	(0.025)	(0.032)					
Curvature	0.029	0.258***					
	(0.025)	(0.032)					
Self-control problems	0.183***	-0.031					
	(0.025)	(0.032)					
Ln(income)	0.265***		0.266***		0.367***		0.165***
	(0.028)		(0.028)		(0.036)		(0.035)
Ln(years of education)	-0.815***		-0.816***		-1.897***		
	(0.209)		(0.205)		(0.264)		
Risk aversion	0.134		0.115		0.297*		
	(0.134)		(0.134)		(0.174)		
Lottery seems fun	0.968***		0.950***	0.987***			
	(0.033)		(0.033)	(0.034)			
Enjoy thinking about winning	0.203***		0.186***	0.154***			
	(0.033)		(0.033)	(0.034)			
Other controls	Yes	No	Yes	No	Yes	No	No
R^2	0.494	0.107	0.486	0.444	0.137	0.069	0.008
Observations	2,767	2,767	2,767	2,767	2,767	2,767	2,767

Notes: This table presents regressions of ln(1+monthly lottery expenditures) on the bias proxies and controls summarized in subsection II.A. "Other controls" includes age, ln (years of education), and dummies for male, white, and black. Standard errors are in parentheses. $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Table 3: **Demand Semi-Elasticity**

	(1) IV	(2) IV	(3) OLS
Jackpot expected value	0.7930***	0.7986***	0.9058***
	(0.0875)	(0.0832)	(0.0755)
Lags in H	4	2	0
Quadratic terms in H	Yes	No	No
R^2	0.71	0.67	0.60
Observations	59,789	59,960	60,128

Notes: The dependent variable in all specifications is the natural log of week-averaged game sales. The week-averaged simple jackpot expected value is instrumented by the simple expected value of the week-level jackpot forecast in the IV regressions. Includes linear and quadratic controls for the previous four week-average sales and jackpot pools in addition to quarter-of-sample and state-game-format fixed effects. Standard errors are clustered at the state and week level. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 4: **Intertemporal Substitution**

	(1)	(2)	(3)	(4)	(5)
Jackpot expected value (t)	0.8975***	0.8944***	0.8805***	0.9263***	0.7930***
	(0.0462)	(0.0445)	(0.0473)	(0.0436)	(0.0875)
Jackpot expected value (t-1)	0.1061***	0.0934***	0.1454***	-0.0504	
	(0.0167)	(0.0192)	(0.0330)	(0.0880)	
Jackpot expected value (t-2)	-0.0165	0.0397*	-0.1341		
	(0.0196)	(0.0228)	(0.0905)		
Jackpot expected value (t-3)	0.0528**	-0.1145			
	(0.0213)	(0.0866)			
Jackpot expected value (t-4)	-0.1211				
	(0.0822)				
Observations	59,421	59,513	59,605	59,697	59,789
Akaike Information Criterion	-8,044.68	-8,113.91	-8,553.20	-9,153.55	-13,925.26
Bayesian Information Criterion	-7,891.81	-7,961.01	-8,409.27	-9,045.59	-13,817.28

Notes: The dependent variable in all specifications is the natural log of week-averaged game sales at time t. The week-averaged simple jackpot expected values are instrumented by the simple expected values of their week-level jackpot forecasts. Includes linear and quadratic controls for the four week-average sales and jackpot pools prior to the earliest period in addition to quarter-of-sample and state-game-format fixed effects. Standard errors are clustered at the state and week level. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 5: **Cross-Game Substitution**

	(1) Own game sales	(2) All other games sales	(3) Other lotto games sales	(4) Instant games sales
Jackpot expected value	1.8833***	0.0887	0.0578	0.0452
	(0.3422)	(0.1655)	(0.1447)	(0.0598)
Observations	58,756	58,756	58,756	58,756

Notes: The dependent variable is the sum of the sales of all games within the same state of the type noted in each column. Observations with zero other within-state lotto-style or instant game sales are omitted. The simple expected value of the jackpot is instrumented by the simple expected value of the week-level jackpot forecast. Includes linear and quadratic controls for the previous four week-average sales and jackpot pools in addition to quarter-of-sample and state-game-format fixed effects. Standard errors are clustered at the state and week level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 6: **Sub-Jackpot Semi-Elasticity in California**

	(1) IV	(2) IV	(3) OLS
Jackpot expected value	0.7743***	0.8120***	0.9802***
	(0.0343)	(0.0367)	(0.0265)
2nd prize expected value	0.0712	-0.1245	-0.1610***
	(0.1226)	(0.0875)	(0.0519)
Lags included in H	4	2	0
H includes quadratic terms	Yes	No	No
R^2	0.74	0.70	0.62
Observations	3,101	3,110	3,201

Notes: The dependent variable in all specifications is the natural log of game sales at the draw level. The jackpot and second prize expected values are instrumented by their respective prize expected value forecasts. Includes linear and quadratic controls for up to the previous four draws' sales and jackpot and second prize pools in addition to day-of-week, quarter-of-sample, and game-format fixed effects. Standard errors are clustered at the week level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table 7: **Price Semi-Elasticity (Summary)**

	(1) Pooled	(2) Pooled	(3) Powerball	(4) Mega Millions
Price	-0.5583***	-0.5356***	-0.6031***	-0.5079***
	(0.0660)	(0.0624)	(0.1023)	(0.0652)
Jackpot pool	0.0040***		0.0059***	0.0032***
	(0.0003)		(0.0006)	(0.0003)
Jackpot expected value		0.9657***		
		(0.0696)		
Observations	416	416	208	208

Notes: The week-averaged simple jackpot expected value is instrumented by the simple expected value of the week-level jackpot forecast, as well as by the prior four draws' ticket sales, jackpot pools, and day-of-week, quarter-of-sample, and game fixed effects. Columns (3)-(4) are limited to the Powerball and Mega Millions price change event data, respectively, while columns (1)-(2) pool the data from both price change events. "Log ticket sales (adj.)" are the natural log of week-averaged sales adjusted by subtracting the estimated sales increase due to the increase in the sub-jackpot expected value using our preferred estimate of the sub-jackpot demand semi-elasticity. Includes game controls for each event. Standard errors are clustered at the week level. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure 1: **Graphical Illustration of Bias**

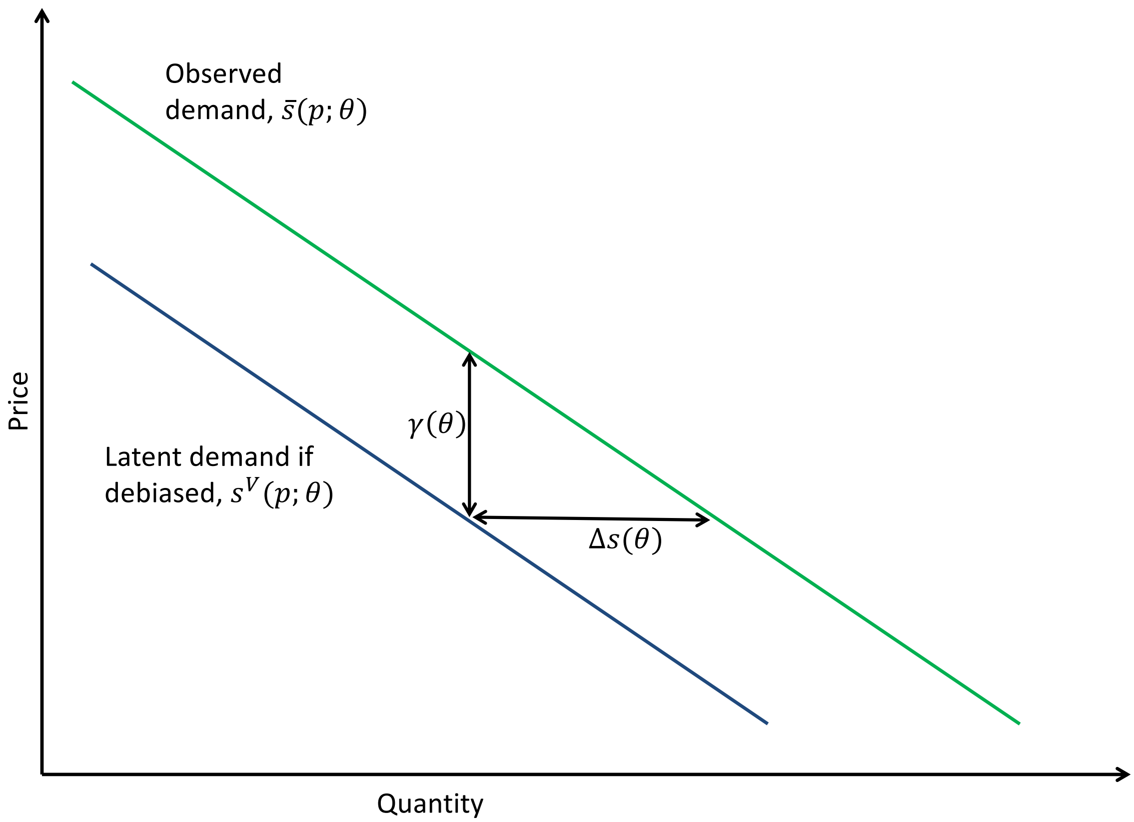
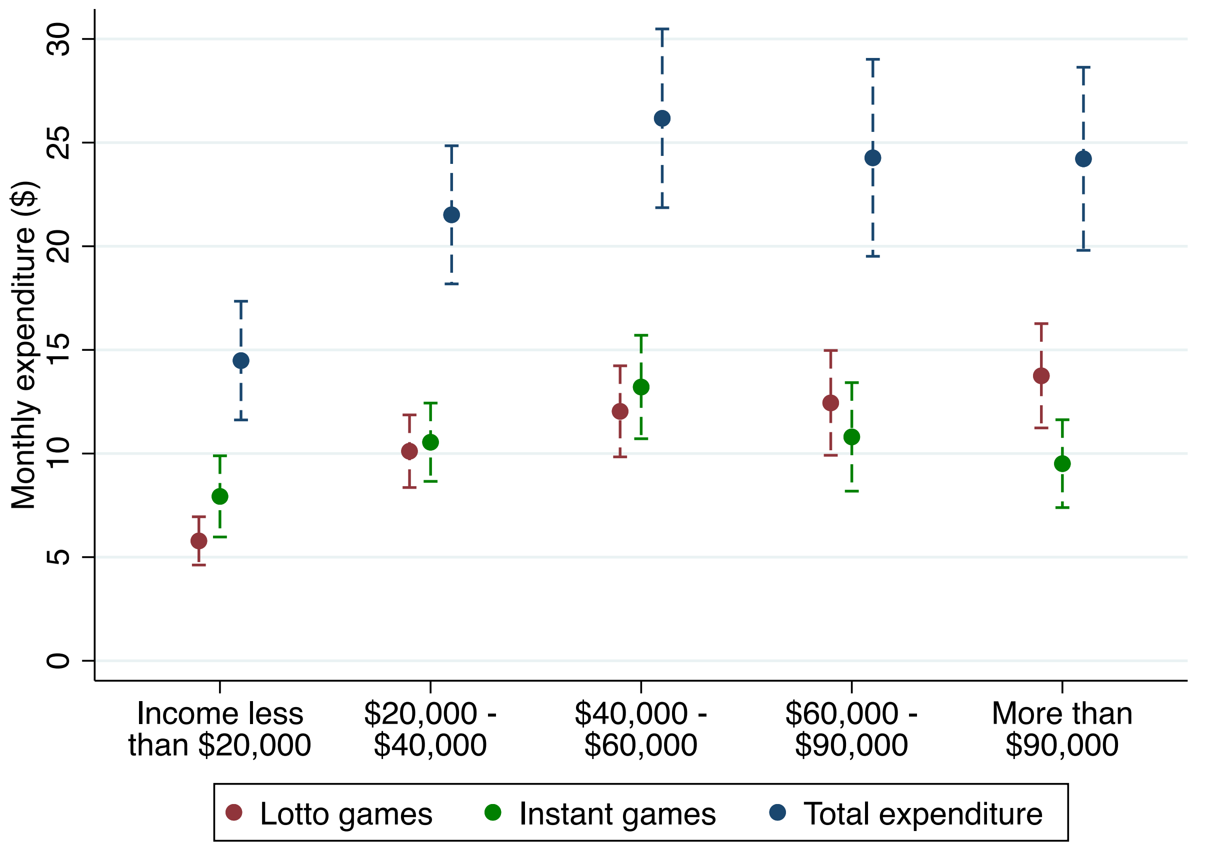
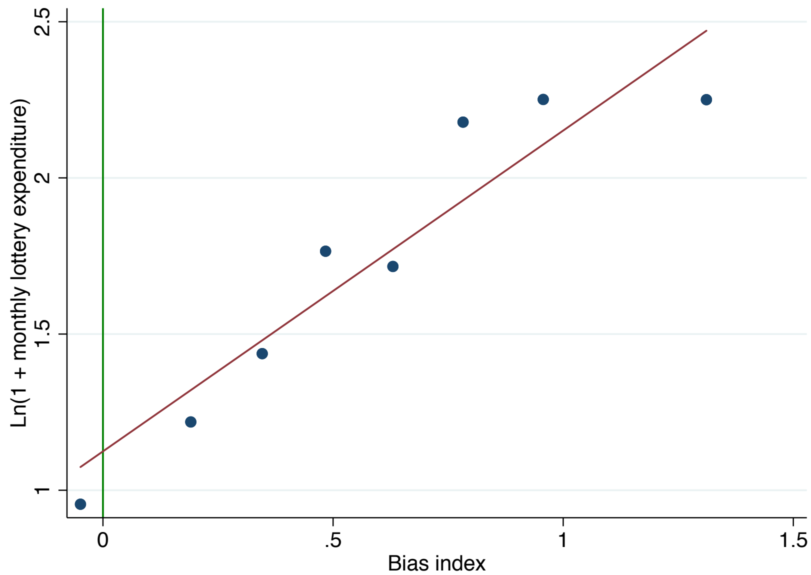


Figure 2: **Lottery Expenditures by Income Group**



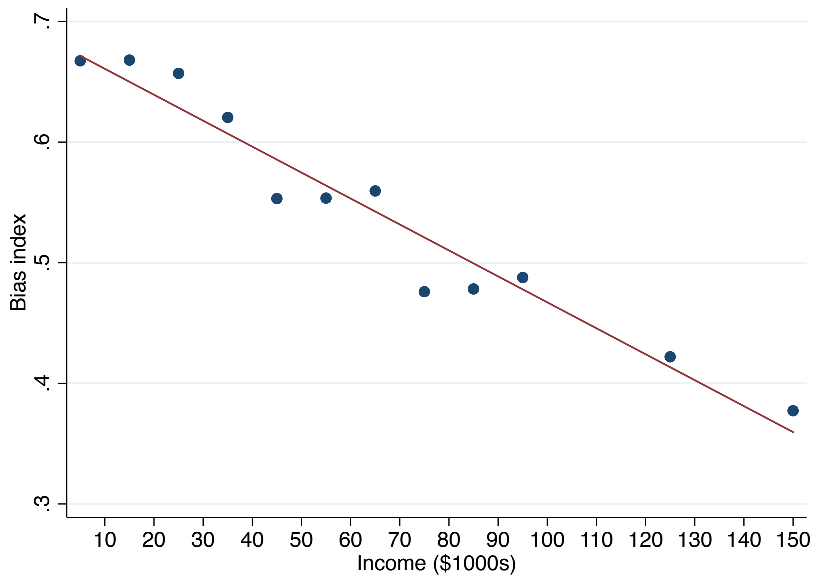
Notes: This figure presents estimates of lottery expenditures by income, along with 95% confidence intervals. "Lotto games" includes MegaMillions, Powerball, Keno, as well as other state games such as Daily 4, Pick 3, Megabucks, SuperLotto. "Instant games" includes scratch off tickets and other same-day games. "Total expenditure" refers to the sum of expenditures in these two categories.

Figure 3: **Relationship Between Lottery Expendiures and Index of Bias Proxies**



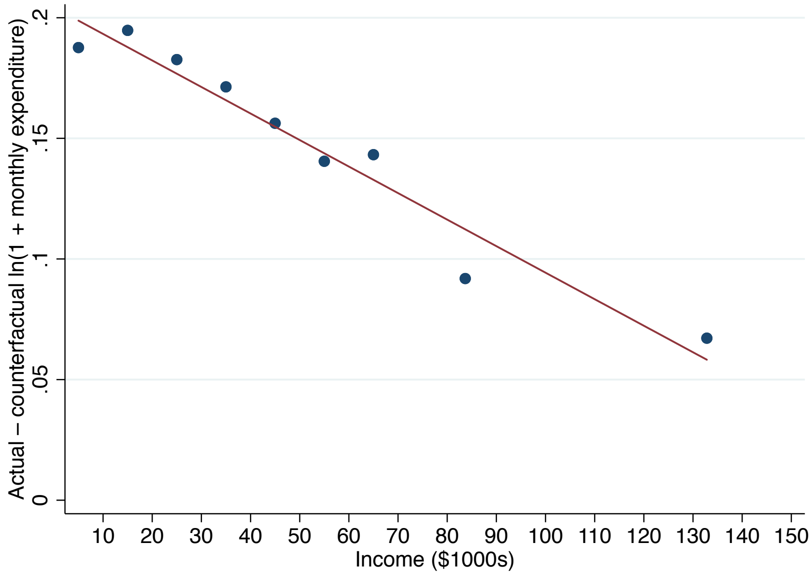
Notes: This figure presents a binned scatter plot of the natural log of lottery expenditures by our aggregated index of bias proxies. To create the index, we first center the bias proxies around the "normative values"—i.e., values that correspond to correct answers or to reported self-control problems. We then divide the proxies by their standard deviation and take the average. The vertical line corresponds to the the bias index value of "normative consumers" who are not subject to any biases or self-control problems. Negative values are possible if, e.g., individual underestimate the expected winnings of a lottery ticket.

Figure 4: **Index of Bias Proxies vs. Income**



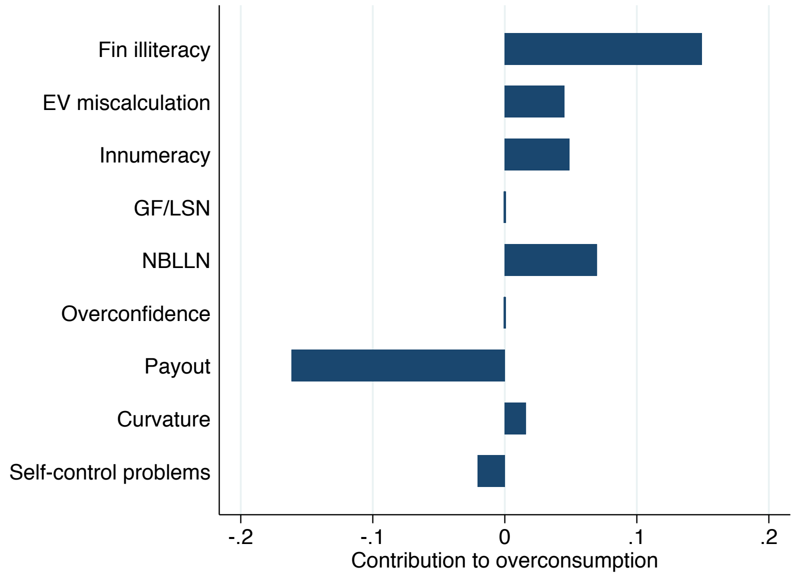
Notes: This figure presents the average value of our aggregated index of bias proxies by income. To create the index, we first center the bias proxies around the "normative values"—i.e., values that correspond to correct answers or to reported self-control problems. We then divide the proxies by their standard deviation and take the average.

Figure 5: **Share of Expenditures Attributable to Bias, by Income**



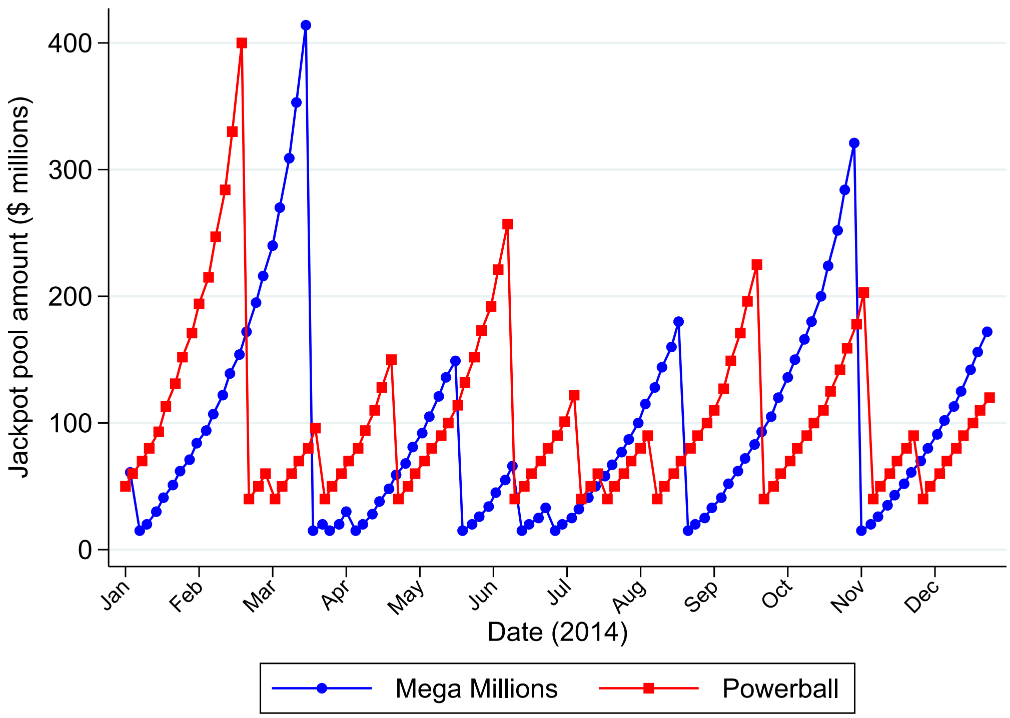
Notes: This figure presents the share of lottery expenditures attributable to bias. Specifically, the figure plots the average estimated value of $\ln(1 + \hat{s}_i^V) - \ln(1 + s_i)$ by income bin.

Figure 6: **Contribution to Overconsumption**



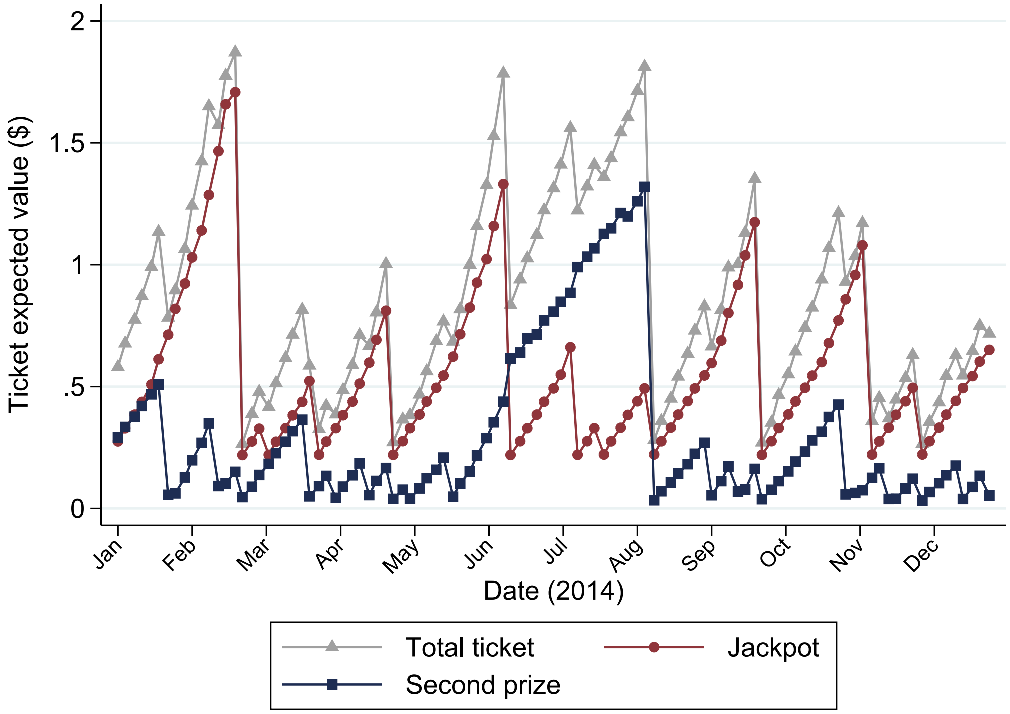
Notes: This figure plots the average contribution of each bias proxy b_k to lottery expenditures; i.e., $\hat{\tau}_i(b_k^V - \hat{b}_{i,k})$. Units are in fraction of consumption. Positive values indicate that the bias on average leads individuals to over-spend on lotteries, while negative values indicate that the bias leads individuals to underspend on the lottery, on average.

Figure 7: **Mega Millions and Powerball Jackpots in 2014**



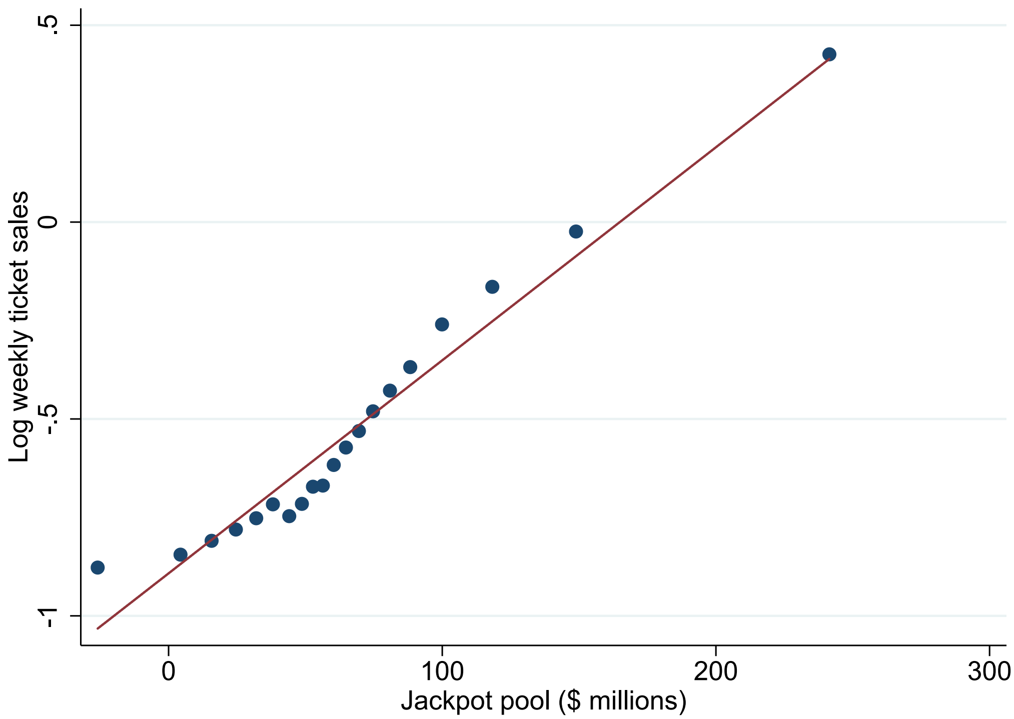
Notes: This figure shows jackpot amounts by drawing for Mega Millions and Powerball in an example year.

Figure 8: **California Powerball Jackpot and Sub-Jackpot Prizes in 2014**



Notes: This figure shows jackpot, second prize, and total ticket expected value for Powerball in California in an example year.

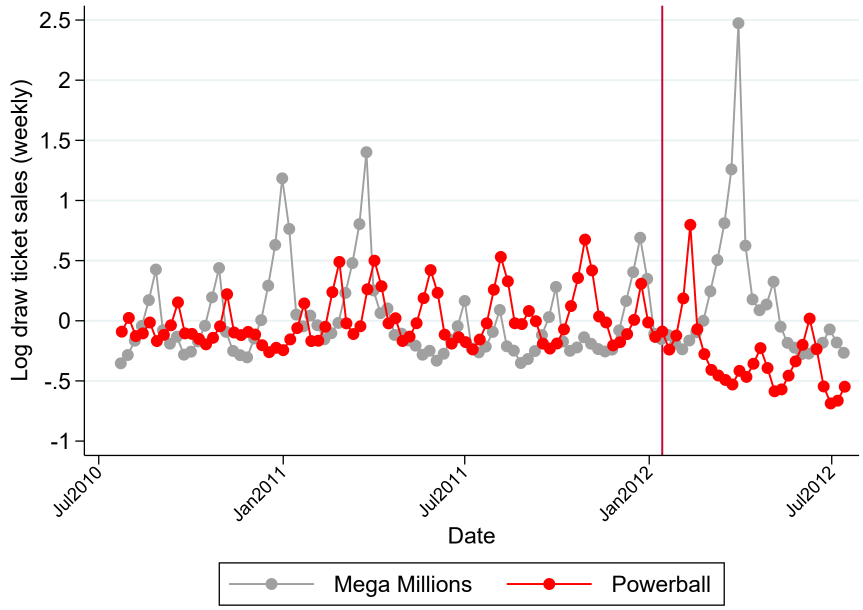
Figure 9: **Sales vs. Jackpot**



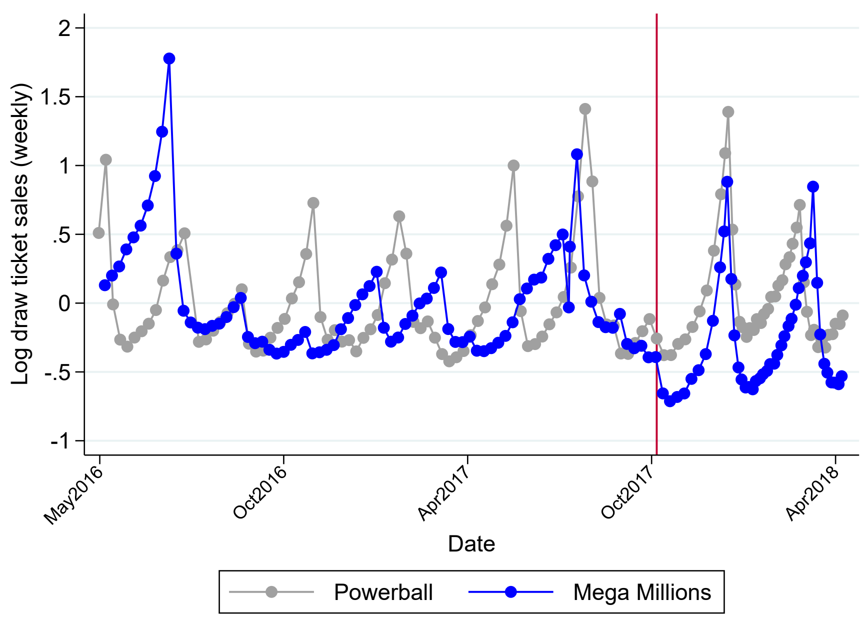
Notes: Jackpot pool amounts are averaged at the week level. Includes quarter-of-sample and state-game-format fixed effects.

Figure 10: **Format Change Event Studies**

(a) **Powerball Format Change**



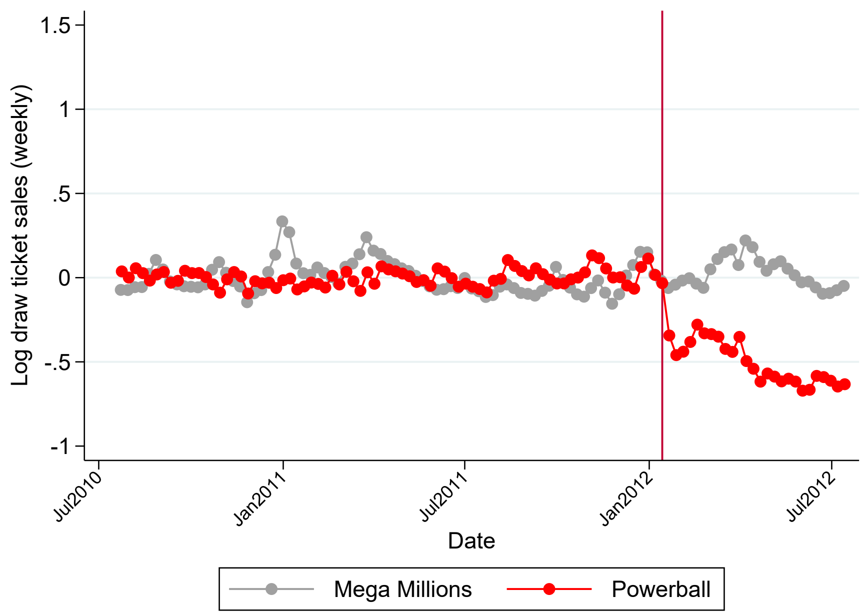
(b) **MegaMillions Format Change**



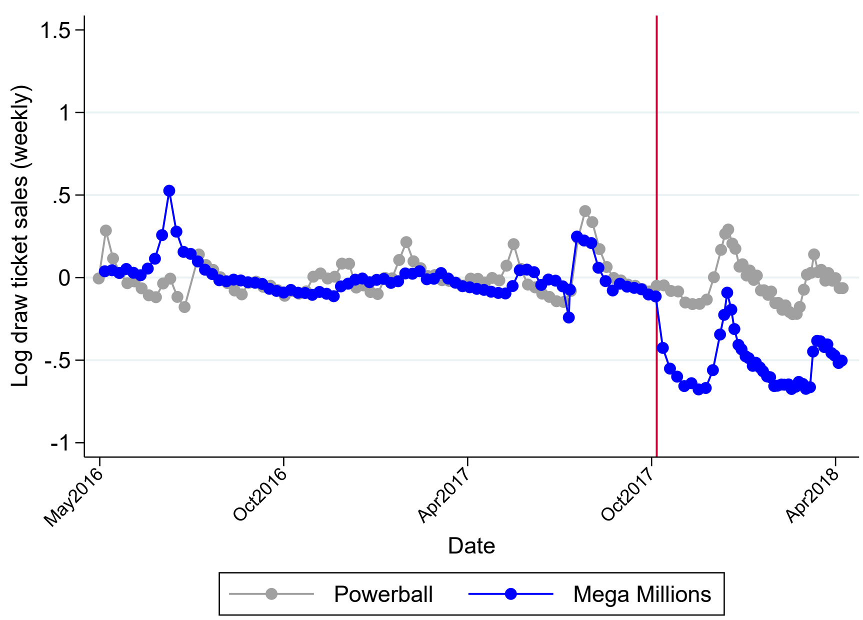
Notes: This figure presents the natural log of draw-level ticket sales before and after price increases, which are indicated by the vertical red lines. In January 2012, the Powerball ticket price increased from $1 to $2. In October 2017, the Mega Millions ticket price also increased from $1 to $2.

Figure 11: **Format Change Event Studies with Controls**

(a) **Powerball Format Change**



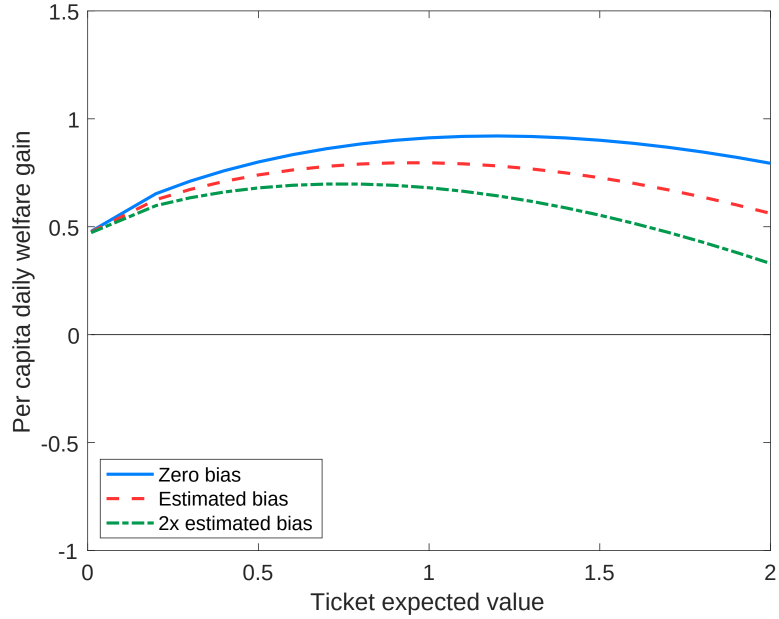
(b) **MegaMillions Format Change**



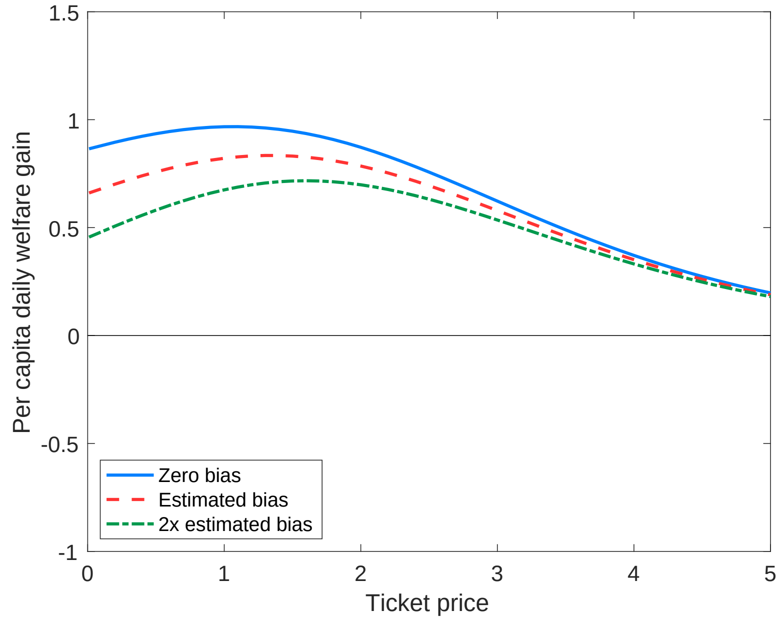
Notes: This figure presents the natural log of draw-level ticket sales after residualizing on controls before and after price increases, which are indicated by the vertical red lines. In January 2012, the Powerball ticket price increased from $1 to $2. In October 2017, the Mega Millions ticket price also increased from $1 to $2. Draw-level log sales are residualized on the jackpot pool amount, averaged over weeks, and expressed as differences from the pre-price change period mean.

Figure 12: **Effect of Lottery Attributes on Social Welfare**

(a) **Variation in Ticket Expected Value**



(b) **Variation in Ticket Price**



Notes: These figures plot the simulated social welfare gain, in dollars per person-day, from a representative lottery with varying expected value (Panel a) or price (Panel b). Values are computed relative to a hypothetical benchmark without a lottery product. The representative lottery is based on a standard $2 Mega Millions ticket with a jackpot pool of $300 million.

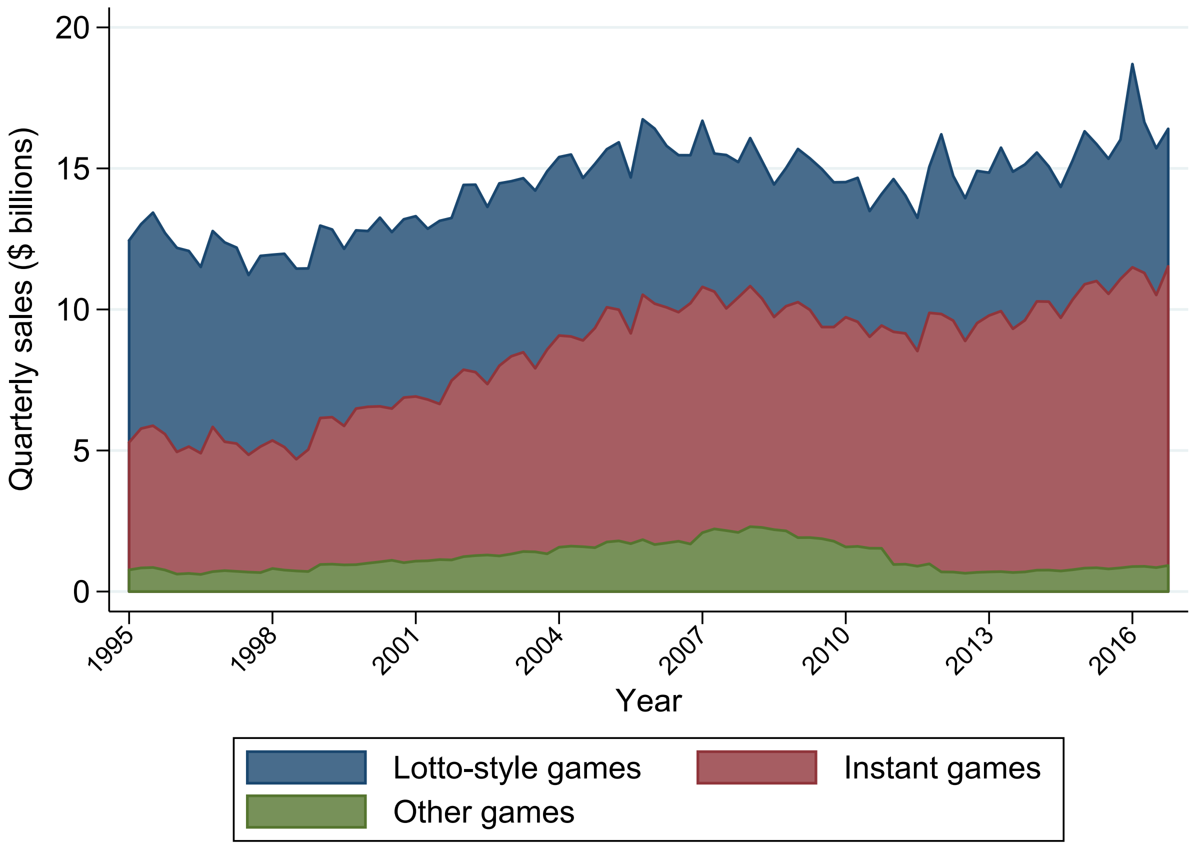
Online Appendix: Not for Publication

The Optimal Taxation of Lotteries

Hunt Allcott, Benjamin B. Lockwood, and Dmitry Taubinsky

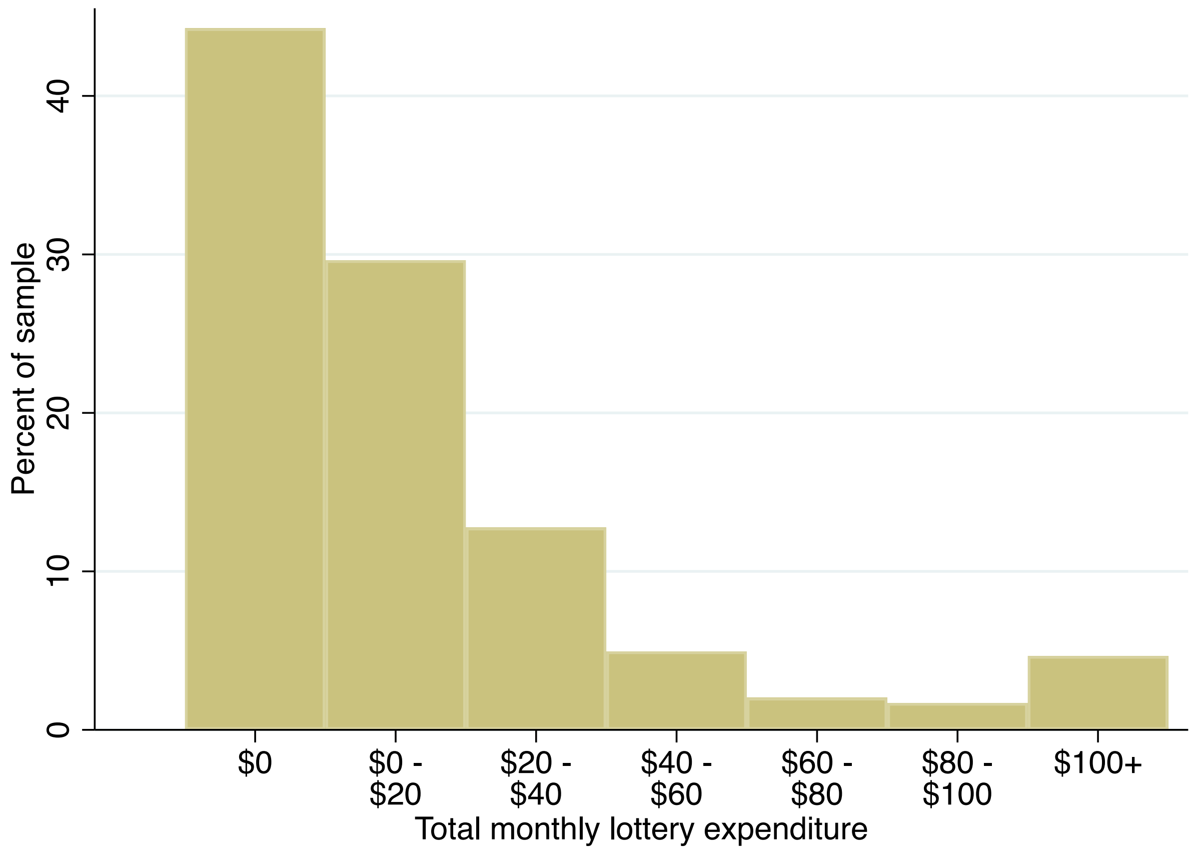
A Empirical Results Appendix

Figure A1: **Lottery Consumption by Category over Time**



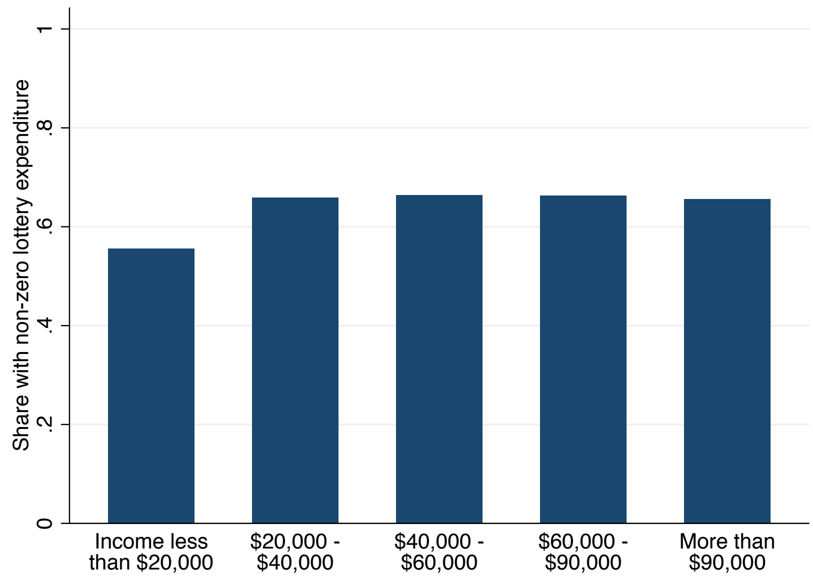
Notes: This figure shows total lottery sales by category, using data from La Fleur's.

Figure A2: **Distribution of Monthly Lottery Expenditures**



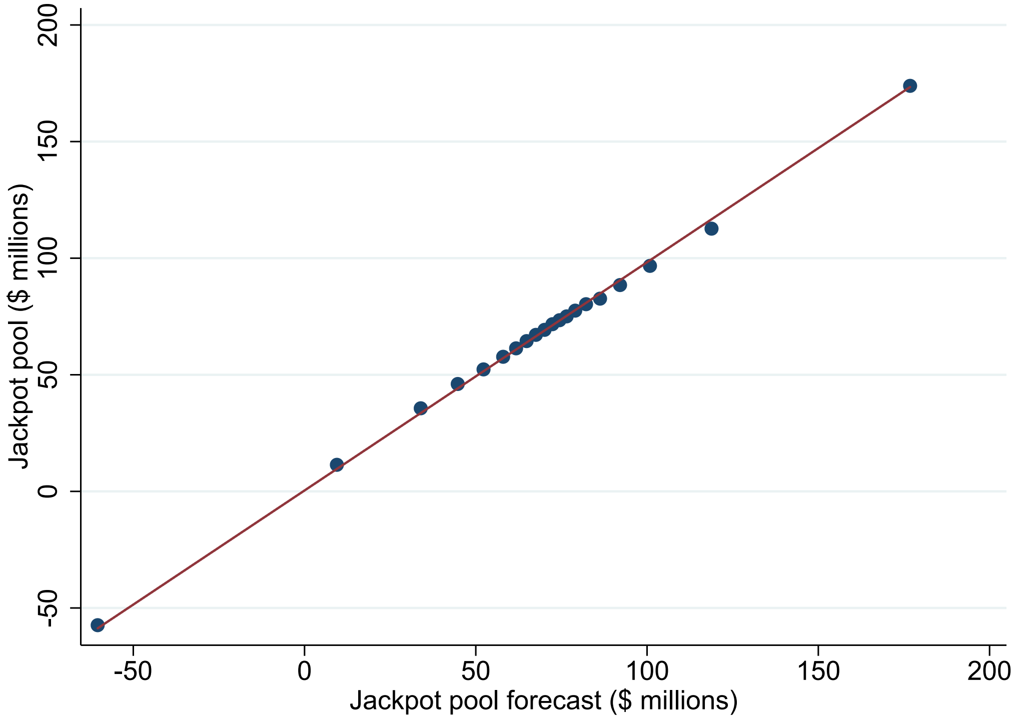
Notes: This figure presents a histogram of the monthly lottery expenditures of participants in our online sample.

Figure A3: **Non-Zero Lottery Expenditure by Income**



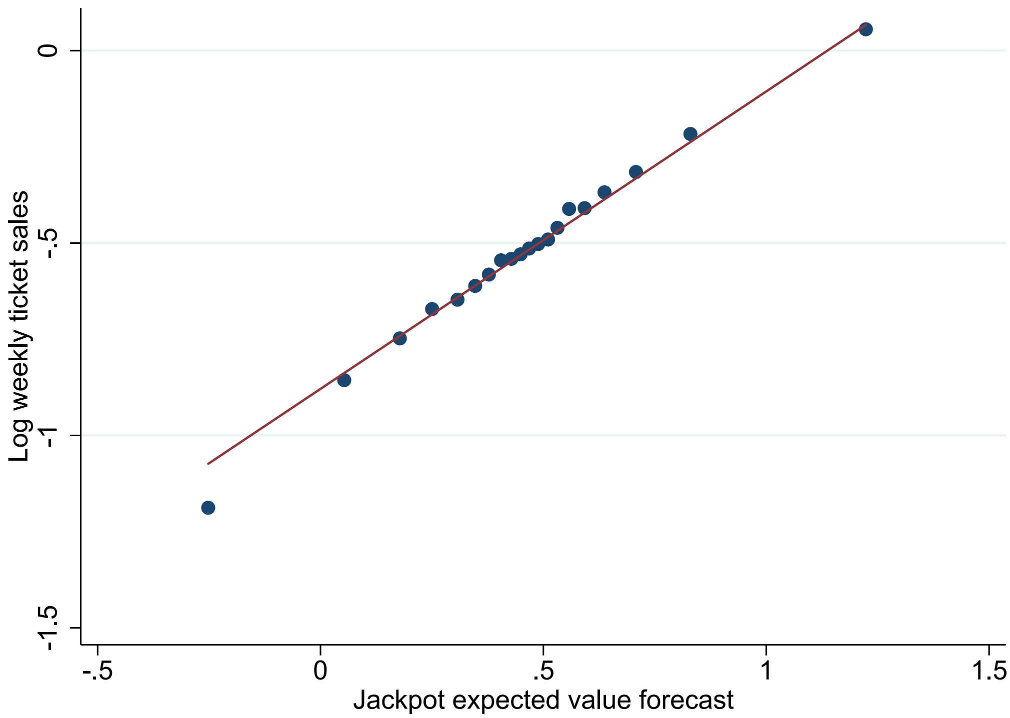
Notes: This figure reports the fraction of respondents who reported having non-zero total lottery expenditures in an average month, by income bin.

Figure A4: **Actual Jackpot vs. Jackpot Forecast Instrument**



Notes: This presents a binned scatterplot of the first stage of the instrumental variables estimate of Equation (15). Includes quarter-of-sample and state-game-format fixed effects.

Figure A5: **Sales vs. Jackpot Forecast Instrument**



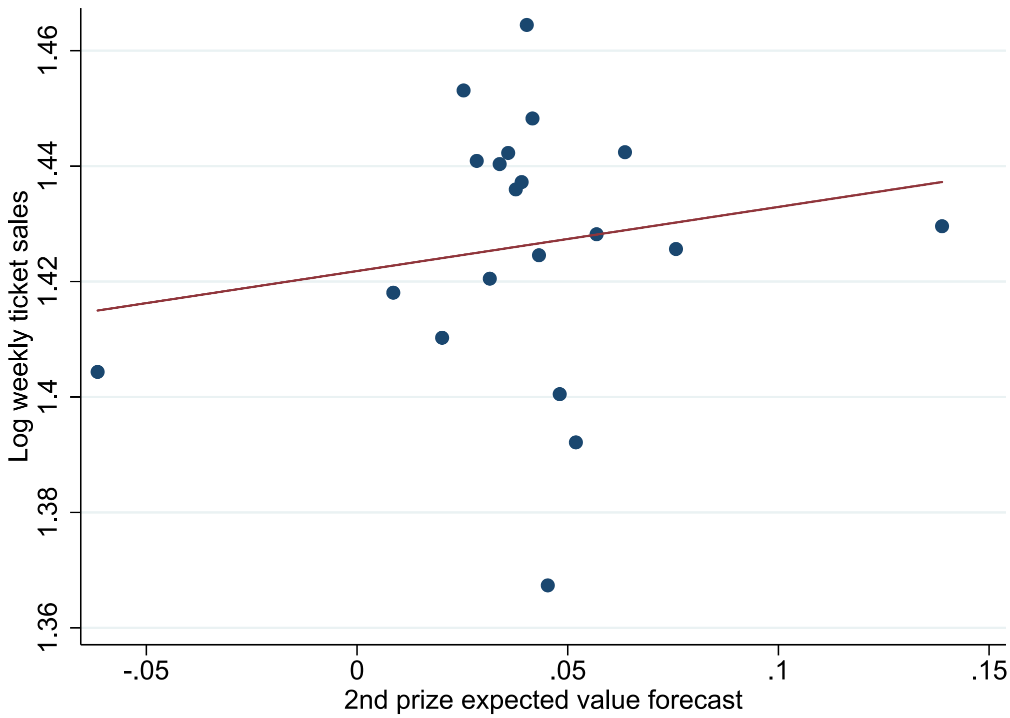
Notes: This presents a binned scatterplot of the reduced form of the instrumental variables estimate of Equation (15). Includes quarter-of-sample and state-game-format fixed effects.

Table A1: **Regressions of Jackpot Expected Value on Forecast Instruments**

	(1) EV (t)	(2) EV (t-1)	(3) EV (t-2)	(4) EV (t-3)	(5) EV (t-4)
Jackpot expected value forecast (t)	0.6040***	-0.0808***	-0.0023	0.0079	0.0015
	(0.0255)	(0.0196)	(0.0131)	(0.0096)	(0.0014)
Jackpot expected value forecast (t-1)	0.0097	0.7564***	-0.0788***	-0.0080	-0.0023
	(0.0110)	(0.0216)	(0.0147)	(0.0063)	(0.0018)
Jackpot expected value forecast (t-2)	-0.0229	-0.0029	0.8576***	-0.0653***	-0.0064
	(0.0137)	(0.0148)	(0.0210)	(0.0155)	(0.0040)
Jackpot expected value forecast (t-3)	-0.0311	-0.0584***	-0.0407**	0.9177***	-0.0527***
	(0.0194)	(0.0188)	(0.0191)	(0.0216)	(0.0096)
Jackpot expected value forecast (t-4)	0.0773	0.0818	0.0568	0.0588	1.0147***
	(0.0657)	(0.0652)	(0.0632)	(0.0593)	(0.0129)
R^2	0.73	0.81	0.86	0.91	0.97
Observations	59,421	59,421	59,421	59,421	59,421

Notes: The dependent variables are the week-averaged simple jackpot expected values at different points in time relative to the latest jackpot forecast instrument. The jackpot forecast instruments constitute the simple expected value of the week-level jackpot forecast at the corresponding point in time, constructed using prize information from prior to the earliest forecast and the actual prize rollover sequence. Includes linear and quadratic controls for the four week-average sales and jackpot pools prior to the earliest period in addition to quarter-of-sample and state-game-format fixed effects. Standard errors are clustered at the state and week level.

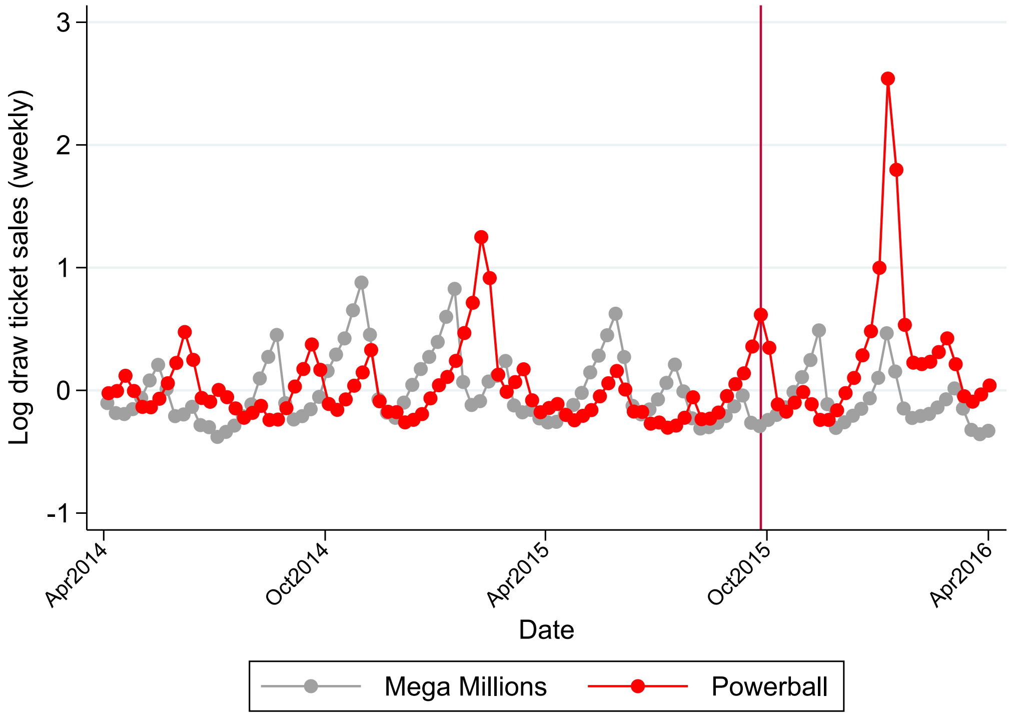
Figure A6: **Reduced Form: Second Prize in California**



Notes: Includes a control for the jackpot expected value forecast and linear and quadratic controls for the previous four draws' sales and jackpot and second prize pools in addition to day-of-week, quarter-of-sample, and game-format fixed effects.

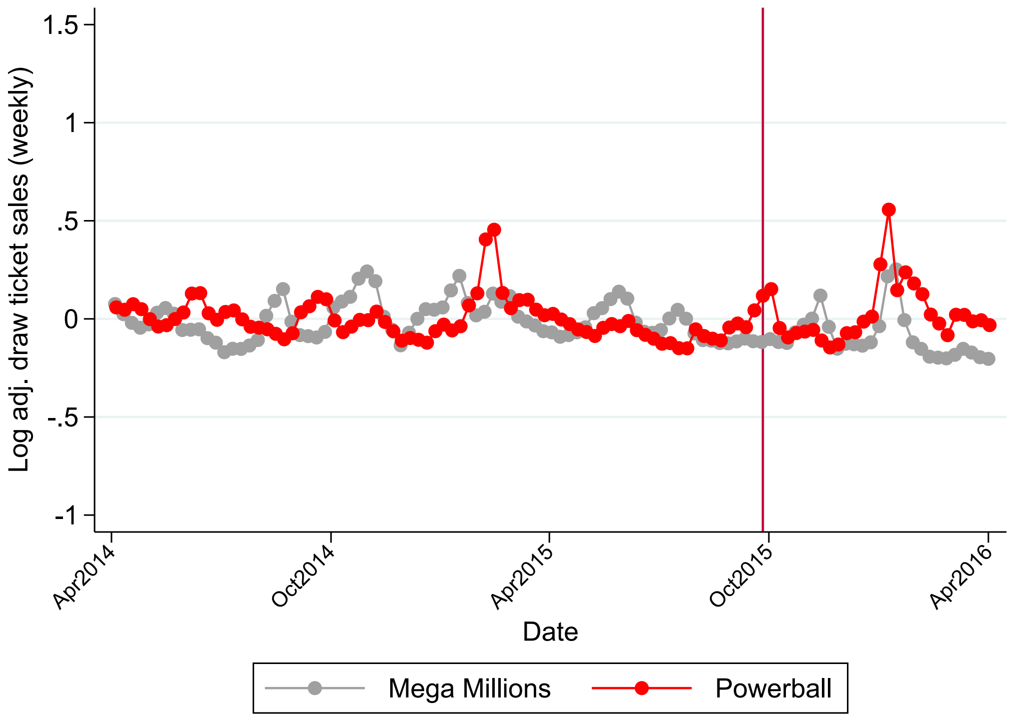
A.A Powerball 2015 Format Change

Figure A7: **Powerball Format Change Event Study**



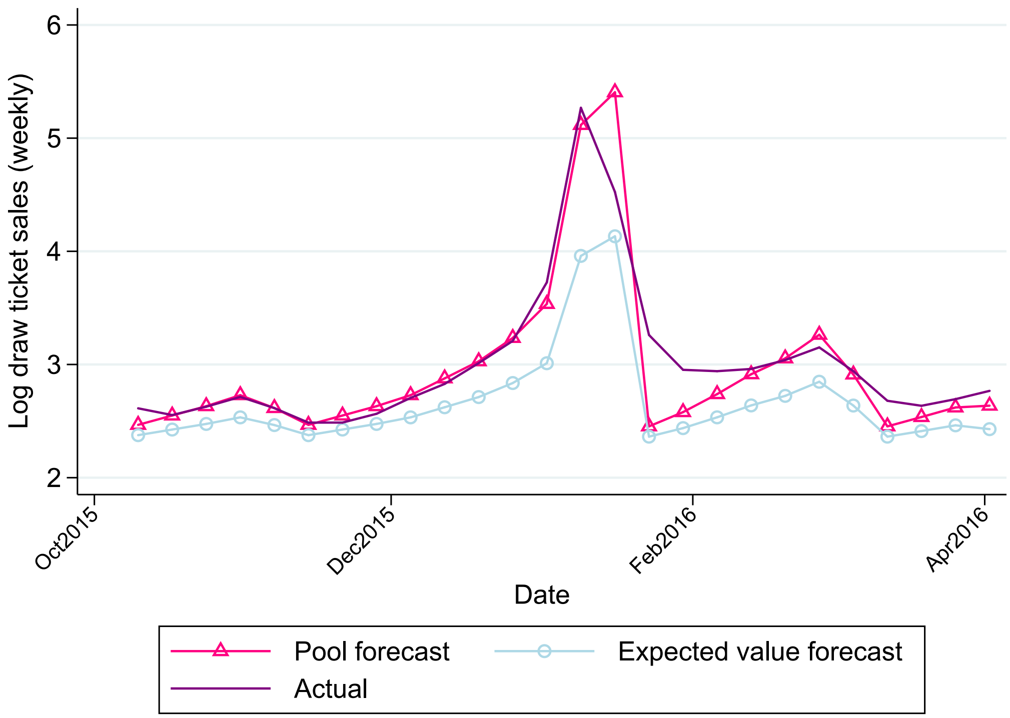
Notes: The red vertical line indicates the week prior to October 4, 2015, the date when Powerball changed its format.

Figure A8: **Powerball Format Change Event Study (Jackpot and Sub-jackpot Pool Controls)**



Notes: The red vertical line indicates the week prior to October 4, 2015, the date when Powerball changed its format. Draw-level log sales are residualized on the jackpot pool amount, averaged over weeks, are adjusted by subtracting the estimated sales increase due to the increase in the sub-jackpot expected value solely due to prize changes using our preferred estimate of the sub-jackpot demand semi-elasticity, and expressed as differences from the pre-regulation change period mean. We fix the probabilites in the sub-jackpot expected value across the regulation change to isolate the increase due to prize changes.

Figure A9: **Post-Format Change Forecasted Sales**



Notes: Limited to observations after October 4, 2015, the date when Powerball changed its format. Log sales are forecasted using quarter-of-sample and game fixed effects and either the jackpot pool or simple jackpot expected value with data from the previous year of Powerball sales and jackpots and the contemporaneous and previous year of Mega Millions sales and jackpots.

Table A2: **Difference-in-Differences: Format Change Event**

	(1) Log ticket sales (adj.)	(2) Log ticket sales (adj.)	(3) Log ticket sales
Powerball × Post	0.0706	0.2615***	0.3345***
	(0.0531)	(0.0525)	(0.1204)
Post	-0.1674***	-0.1764***	-0.1720**
	(0.0386)	(0.0374)	(0.0668)
Powerball	-0.3753***	-0.5291***	-0.2854***
	(0.0217)	(0.0296)	(0.0455)
Jackpot pool	0.0037***		
	(0.0004)		
Jackpot expected value		0.9070***	
		(0.0747)	
R^2	0.85	0.84	0.20
Observations	208	208	208

Notes: The week-averaged jackpot pool and week-averaged simple jackpot expected value are instrumented by the week-level jackpot forecast and simple jackpot expected value forecast in columns (1) and (2), respectively, as well as by the prior four draws' ticket sales, jackpot pools, and day-of-week, quarter-of-sample, and game fixed effects. Column (3) includes a linear time trend. "Powerball" is an indicator for the Powerball game observations. "Post" is an indicator for observations after the Powerball format change event. "Log ticket sales (adj.)" are the natural log of week-averaged sales adjusted by subtracting the estimated sales increase due to the increase in the sub-jackpot expected value using our preferred estimate of the sub-jackpot demand semi-elasticity. In column (1), we hold win probabilities constant across the regulation change when adjusting sales. Standard errors are clustered at the week level. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

B Survey measures of preferences and bias proxies

Variable	Question text
	Preferences
Lottery seems fun	To what extent do you agree or disagree with the following statement: *For me, playing the lottery seems fun.* [-5 Strongly disagree, -4, -3, -2, -1, 0 Neutral, 1, 2, 3, 4, 5 Strongly agree]
Enjoy thinking about winning	To what extent do you agree or disagree with the following statement: *I enjoy thinking about how life would be if I won the lottery.* [-5 Strongly disagree, -4, -3, -2, -1, 0 Neutral, 1, 2, 3, 4, 5 Strongly agree]
	Bias proxies
Financial illiteracy	Suppose you had $100 in a savings account and the interest rate was 2% per year. After 5 years, how much do you think you would have in the account if you left the money to grow? [More than $110, Exactly $110, Less than $110]

	Imagine that the interest rate on your savings account was 1% per year and inflation was 2% per year. After 1 year, how much would you be able to buy with the money in this account? [More than today, Exactly the same, Less than today]
	Please tell us whether this statement is true or false. "Buying a single company's stock usually provides a safer return than a stock mutual fund." [True, False]
	Normally, which asset displays the highest fluctuations over time? [Savings accounts, Bonds, Stocks]
	When an investor spreads her money among different assets, does the risk of losing money: [Increase, Decrease, Stay the same]
Expected value miscalculation	Imagine a nationwide lottery in which tickets cost $1, and 1 out of every 10 tickets wins $5. What percent of the lottery revenues are returned to the winners? [_%]
	Now imagine a different nationwide lottery in which 1 out of every 100,000,000 tickets wins $60,000,000, AND 1 out of every 1,000 tickets wins $50. What percent of the lottery revenues are returned to the winners? [_%]
Innumeracy	A second hand car dealer is selling a car for $6,000. This is two-thirds of what it cost new. How much did the car cost new? [$_]
	If 5 people all have the winning numbers in the lottery and the prize is $2 million, how much will each of them get? [$_]
	Let's say you have $200 in a savings account. The account earns 10% interest per year. How much will you have in the account at the end of two years? [$_]

GF/LSN (Gambler's Fallacy/Law of Small Numbers)	For the next few questions, imagine flipping a coin that has a 50% chance of landing heads and a 50% chance of landing tails. Imagine that after eight flips you observe the following result: tails-tails-tails-heads-tails-heads-heads-heads. What is the probability, in percent from 0-100, that the next flip is tails? [_%] Now imagine starting over and flipping a coin ten times. The first nine are all heads. What are the chances, in percent from 0-100, that the next flip is heads? [_%]
NBLLN (Non-Belief in the Law of Large Numbers)	Now imagine starting over and flipping a coin 1000 times. What are the chances, in percent from 0-100, that the total number of heads will lie within the following ranges? (Your answers should add up to 100.) [Between 491 and 519 heads _%, Between 520 and 1000 heads _%]
Overconfidence	Imagine **you** could keep buying whatever lottery tickets you want, over and over for a very long time. For every $1000 you spend, how much do you think **you** would win back in prizes, on average? [$0 to $100, $100 to $200, $200 to $300, $300 to $400, $400 to $500, $500 to $600, $600 to $700, $700 to $800, $800 to $900, $900 to $1000, $1000 to $1500, $1500 to $2000, $2000 to $5000, More than $5000] Imagine that the **average** lottery player in the country could keep buying whatever lottery tickets they want, over and over for a very long time. For every $1000 they spend, how much do you think they would win back in prizes, on average? [$0 to $100, $100 to $200, $200 to $300, $300 to $400, $400 to $500, $500 to $600, $600 to $700, $700 to $800, $800 to $900, $900 to $1000, $1000 to $1500, $1500 to $2000, $2000 to $5000, More than $5000]
Expected payout	Think about the total amount of money spent on lottery tickets nationwide. What percent do you think is given out in prizes? [0 - 10%, 10 - 20%, 20 - 30%, 30 - 40%, 40 - 50%, 50 - 60%, 60 - 70%, 70 - 80%, 80 - 90%, 90 - 100%]
Curvature	Now we'd like to ask you how much you think lottery winnings increase happiness. A recent study surveyed Swedish lottery winners about their psychological well-being.

	A typical person in the study had won between $100,000 and $800,000 in the lottery about 12 years before the survey. The study compared people who had won more vs. less money to determine the effect of additional lottery winnings. The four measures of well-being were (i) happiness, (ii) "how satisfied are you with life," (iii) how often you felt each of 12 positive and negative emotions, and (iv) "how satisfied are you with your finances?" People answered each question on a numerical scale. By how many percent do you think an additional $100,000 in lottery winnings increased or decreased well-being? (Enter 0 in either box for a 0% change) [Increased by _%, Decreased by _%]
Self-control problems	Do you feel that you should play the lottery more, less, or the same as you do right now? [-5 I should play less, -4, -3, -2, -1, 0 I should play the same amount, 1, 2, 3, 4, 5 I should play more]