

**Lucky Stores, Gambling, and Addiction:  
Empirical Evidence from State Lottery Sales**

April 2005

Jonathan Guryan  
University of Chicago, GSB  
and NBER

Melissa S. Kearney  
Brookings Institution  
and NBER

**Abstract:** There is a large body of literature in both psychology and economics documenting mistaken perceptions of randomness. Such mistakes typically take the form of a belief in the *gambler's fallacy* or the *hot hand* myth, both driven by a belief in the so-called *law of small numbers*. In this paper we demonstrate that people appear to believe that “lightning will strike twice” when it comes to lottery jackpots. First, we show that in the week following the sale of a winning ticket, retailers that sell a winning jackpot ticket experience relative increases in game-specific ticket sales of between 12 and 38 percent, with the sales response increasing in the size of the jackpot. In addition, the increase in sales experienced by the winning vendor increases with the proportion of the local population comprised of high school dropouts, elderly adults, and households receiving public assistance. We further show that this increase in retail-game sales initially reflects an increase in total sales at the retail and zip code level. Second, we show that the increase in sales is persistent at the winning retailer. However, the data do not provide clear evidence that the increase in sales at the zip code level is persistent. It thus appears that in the long run, consumers are persistent in their habit of buying lottery tickets at the “lucky” store; however, as the shock to total gambling dissipates, there is no evidence that lottery gambling itself is habit forming or addictive.

Comments welcome to [mkearney@brookings.edu](mailto:mkearney@brookings.edu) and [jguryan@gsb.uchicago.edu](mailto:jguryan@gsb.uchicago.edu). A previous version of this paper was titled “Perceptions of Randomness and Habit Formation among Lottery Gamblers.” The authors thank Josh Angrist, David Autor, Marianne Bertrand, Bill Dickens, Mark Duggan, Amy Finkelstein, Phil Levine, Toby Moskowitz, Kevin Murphy, Richard Thaler, Ping Zhang, and seminar participants at the NBER Summer Institute Labor Studies, University of Chicago GSB, University of Maryland Department of Economics, Federal Reserve Board of Governors, Heinz School of Public Policy at Carnegie Mellon, University of Illinois Urbana-Champaign, and Federal Reserve Bank of Chicago for helpful comments. We also thank Dwyer Gunn, Jillian Popadak, Andriy Protsyk, Eric Sun, and Daniel Theisen for valuable research assistance. Kearney gratefully acknowledges financial support from the Brookings Institution’s Andrew W. Mellon Foundation Early Career Fellowship.

## **I. Introduction**

In this paper we examine detailed sales data from the Texas Lottery Commission to determine whether consumers respond to the sale of a winning lottery ticket. The sale of a winning lottery ticket at a store does not affect the likelihood that the retailer will sell a winning ticket in the future. Therefore, if fully informed consumers treat lottery tickets as purely financial assets, then store-level sales should not respond to such an event. However, we show that retailers who sell a winning jackpot ticket experience relative increases in ticket sales of the winning game between 12 and 38 percent the following week, and that the sales response increases in the size of the jackpot. The random assignment of winning tickets to stores, conditional on contemporaneous sales, allows us to interpret this subsequent increase as being causally related to the sale of the winner. We discuss both rational and behavioral explanations for this response.

Having established that there is an initial demand response to the sale of a winning ticket, we test for substitution across games within the winning store, and from nearby stores. We show that net of any substitution of these types, the sale of a winning ticket leads to an aggregate increase in total zip-code level sales. We further show that store-level sales do not predict future sales of winning tickets, conditional on contemporaneous lottery sales. Therefore, the sale of a winning ticket constitutes an exogenous shock to lottery gambling consumption and can be used to test whether the immediate increase in lottery consumption is persistent. We discuss the conditions under which this persistence can be interpreted as evidence of habit formation or addiction.

Following the economics literature, we define a good as addictive if past consumption causally increases current consumption by increasing the marginal utility of

current consumption (e.g. Becker and Murphy, 1988). Thus, it follows that if the shock to lagged consumption—in this case the sale of a winning lottery ticket—is a valid instrument for lagged consumption in a structural equation relating current consumption and lagged consumption, then the Instrumental Variables (IV) estimate provides a direct test of addiction.

We find that stores that sell winning tickets experience increases in sales of the winning game for up to 40 weeks, conditional on contemporaneous sales. To interpret this persistence as addiction, we must show that the initial increase in aggregate lottery sales leads to subsequent aggregate increases. The data do *not* provide such evidence, as there is no evidence of a persistent increase in sales at the zip code level. The data are consistent, however, with the long-run increase in sales at the winning vendor being completely offset by a permanent substitution away from other retailers selling tickets on this game. We therefore do not find strong evidence of addiction of lottery gambling in the long run. Our estimates are, however, too imprecise to rule addiction out.

The long run persistence of heightened sales at the winning retailer could be interpreted as evidence that consumers continue to believe in the luck of that store or that consumers initially respond to a belief in a lucky store and then maintain the habit of buying their tickets at that vendor. Alternatively, the response could reflect a rational response to advertising. The sale of the winning ticket may act like advertising for the lottery. Consumers close to the winning seller may be more exposed to this information, and may in turn decide to purchase lottery tickets. However, the results of a phone survey suggest that all Texas lottery retailers use the same advertising materials provided by the Texas Lottery Commission and that the only substantive difference in formal advertising

between winning stores and non-winning stores is that the winning store advertises the fact that they sold a winner. There could also be greater levels of “chatter” or conversation about the lottery at the winning vendor driven by the past experience with winning. However, it is difficult to understand why localized advertising of this kind would be effective if not for a belief in the lucky store theory among potential consumers.

Our paper contributes to an understanding of how consumers perceive randomness and respond to stochastic shocks. Behavioral economists point to laboratory evidence to support a claim that individuals often fall prey to “overinference” and what Tversky and Kahneman (1971) referred to as the “law of small numbers,” which refers to the mistaken notion that small samples must resemble large samples overly closely (e.g. Bar-Hillel and Wagenaar (1991)). Outside the lab, the finance literature has focused on evidence that investors chase stochastically high-earning assets (e.g. Chevalier and Ellison (1997, 1999)).

Our evidence builds on these strands of the literature, and combines the best attributes of research designs from financial markets and the laboratory. As is the case with the evidence on financial investment decisions, the findings here reflect consumption choices that have real costs for consumers. In contrast to the laboratory studies, consumers buy lottery tickets in real markets.<sup>1</sup> If lottery consumers misperceive the probability of winning—because say they irrationally believe a store is “lucky”—they are in effect responding to the wrong price, and must forgo other consumption in exchange for the lottery tickets.<sup>2</sup> Furthermore, our findings suggest that sales increases

---

<sup>1</sup> We are not the first to use state lottery sales to examine perceptions of randomness; such work by Clotfelter and Cook (1993) and Terrell (1994) is described below.

<sup>2</sup> Kearney (forthcoming) provides evidence that lottery gambling displaces household expenditures on non-gambling items. Her analysis of household expenditure data from the Consumer Expenditure Survey finds

are most pronounced for winning retailers located in zip codes with relatively high proportions of high school dropouts, elderly individuals, and households receiving public assistance. Therefore, if the response does reflect a mistaken notion, economically-disadvantaged groups appear to be those most likely to espouse it.

The empirical finance evidence is subject to the criticism that investors may actually be correct. The signals they see may in fact predict future returns. For example, mutual funds that performed well in the past may have managers that are more skilled. In contrast, just as in controlled laboratory experiments, random assignment rules out most rational explanations for consumers responding to a store selling a winning lottery ticket. We can be sure that conditional on current ticket sales, the fact that a store sells a winning ticket has no effect on the likelihood that any given ticket it sells in the future will be a winner.

An additional advantage of using this setting to learn how consumers respond to and perceive randomness is that a broad population of individuals engages in lottery gambling. State lotteries constitute a \$40 billion dollar annual industry. According to the Gallup Lifestyle Poll conducted in December of 2003, state lottery gambling is by far the most common form of gambling in the United States. About half of Americans say they have bought a lottery ticket in the past 12 months. A 2000 nationally-representative survey conducted by GTECH found that 72 percent of adults in the United States bought a lottery ticket in the past year, with 28 percent reporting playing at least once a week. Put concisely, the population of lottery gamblers is not comprised of a small group of people, but rather, is quite representative of the adult U.S. population. Furthermore, given

---

that among households in the lowest income third, the introduction of a state lottery is associated with statistically significant declines in spending on necessity goods, including food eaten in the home.

the large role state governments play as monopoly providers of lottery goods, it should be of significant interest to both economists and policymakers whether consumers act rationally when buying lottery tickets and whether lottery gambling is addictive.<sup>3</sup>

## **II. Background: Rationality and Addiction**

In this paper we present evidence on two aspects of gambling consumption: whether consumers act as if they have rational beliefs about the probability of winning, and whether lottery consumption is addictive. In this section of the paper we describe two bodies of relevant theoretical literature.

### *A. Perceptions of randomness*

There is a large body of literature in both psychology and behavioral economics on perceptions of randomness. It has frequently been noted that people are often unable to generate or even recognize a random sequence of numbers. This bias manifests itself in both the “gambler’s fallacy” and the “hot hand” myth. The “gambler’s fallacy” is the mistaken notion that the second draw of a signal will be negatively correlated with the first draw. The “hot hand” myth refers to the belief by an observer that a positive signal is more likely to be followed by another positive signal than by a negative signal. A commonly cited example is the belief among sports fans that a basketball player’s chance of hitting a shot is greater following a hit than following a miss on the previous shot.

Both fallacies stem from what Tversky and Kahneman (1971) facetiously labeled the “law of small numbers”: the belief that even small samples will closely resemble the processes or populations that generated them. The label is a reference to the mathematical

---

<sup>3</sup> A comprehensive evaluation of the state’s role as a monopolist would also address both the pricing of lottery tickets and how the revenue generated by the lottery is used. For examples of such analyses, see

law of large numbers, which states that if  $X_1, \dots, X_n$  form a random sample from a distribution for which the mean is  $\mu$ , and  $\bar{X}_n$  is the sample mean, then

$p \lim(n \rightarrow \infty) \bar{X}_n = \mu$ . The error people commonly make is to exaggerate how likely it is that a small sample resembles the parent population from which it is drawn. In other words, they assume the law to hold for a small finite  $n$ .

Rabin (2002) models a belief in the law of small numbers and studies how people making this error differ from Bayesians in their inferences. His model shows that the law of small numbers leads to the tendency to over-infer from short sequences and to believe in nonexistent variation. Rabin's model describes an individual who observes a sequence of binary signals of some underlying quality. An example is a consumer who observes a sequence of good or bad investments made by a mutual-fund manager that signals his underlying level of competence. The consumer is a Bayesian and has correct probabilistic priors about the rate at which signals are generated. In reality, the signals are generated by an i.i.d. process. The consumer's bias derives from a mistaken belief that the signals are drawn *without replacement* from a size-  $N < \infty$  urn, which contains the two values of the signal in proportion to the true quality level. As  $N$  approaches infinity, the person becomes fully Bayesian; as  $N$  approaches one, the person is increasingly convinced of the law of small numbers and increasingly biased in his predictions.<sup>4</sup> It is easy to see that Rabin's formulation of the belief in the law of small numbers leads directly to the "gambler's fallacy".

---

Forrest, Gulley, and Simmons (2000) and Evans and Zhang (2003), respectively.

<sup>4</sup> Rabin offers the following example: assume an observer is positive that a particular fund manager invests successfully with probability 1/2. If  $N=4$ , the observer thinks the analyst has an "urn" of two good years and two bad years. Then if the analyst is successful in year one, the observer thinks there is only a 1/3 chance of success in year two. This captures the law of small numbers in that the person believes the proportion of signals must balance out to the population rate before  $N$  signals are observed.

Laboratory experiments have produced much evidence of the gambler's fallacy. Wagenaar (1972) finds that when people are asked to generate or identify random sequences, their sequences often have negative autocorrelation. Bar Hillel and Wagenaar (1991) provide a more recent review of similar laboratory evidence. While the "gambler's fallacy" might appear to contradict the belief in a "hot hand," both Rabin (2002) and Camerer (1995) point out that the "hot hand" fallacy may actually follow from the "gambler's fallacy". They posit that someone who expects negative autocorrelation – because of a mistaken belief in the law of small numbers – will be surprised by the number of streaks he observes. He will thus come to believe that a series with streaks has positive autocorrelation. This misconception is often discussed in the context of basketball fans who believe that a player has a "hot hand" if he makes a number of shots in a row, even though field data and experiments show that hits and misses appear to be statistically independent (Gilovich, Vallone, and Tversky 1985).

Empirical evidence of such misperception of randomness is often drawn from laboratory settings. One important exception is the set of papers showing that mutual fund flows chase high returns (e.g. Carhart (1997), Chevalier and Ellison (1997, 1999), Grinblatt and Keloharju (2001), Odean (1998)). One implication of these papers is that many investors erroneously attribute stochastically high returns to mutual fund managers.<sup>5</sup> Such a misperception can be classified as a belief in the "law of small numbers". A strength of these papers relative to much of the experimental evidence on the subject is that investors face real costs from their actions. A weakness is that high returns may be correlated with signs of future returns that are unobserved to the

---

<sup>5</sup> Grinblatt, Titman, and Wermers (1995) show evidence that mutual fund managers themselves chase high returns on stocks, suggesting the phenomenon is not isolated among uninformed investors.



econometrician. In other words, investors may in fact *correctly* attribute the high returns to a permanent characteristic of the fund.

Our empirical strategy attempts to draw on the strengths of each of these empirical literatures. Like the finance literature on investors, we examine lottery consumers outside of the laboratory. Consumers are faced with real costs from their consumption choices. If lottery consumers misperceive the probability of winning, they are in effect responding to the wrong price and must forgo other consumption in exchange for the lottery tickets. Like the laboratory, our research design makes use of random assignment. Since winning lottery numbers are chosen randomly, we can be sure that the probability of buying a winning ticket is independent of the location of its purchase. Thus, any response to past locations of winners must be attributed to something other than a change in the price of the ticket.

Two previous studies – Clotfelter and Cook (1993) and Terrell (1994) – provide evidence of the gambler’s fallacy among lottery players. Using data from the Maryland and New Jersey numbers games respectively, they find that the amount of money bet on a particular number falls sharply after the number is drawn and that it gradually returns to its former level after several months. It is an empirical question as to whether lottery players behave as though they correctly perceive the randomness of where a winning lottery ticket is sold. The probability that a vendor sells a winning lottery ticket depends only on the number of tickets (or more precisely, of unique numbers) sold at that store. If consumers understand this, the winning vendor should experience no change in demand after the sale of a winner. That is, consumers should be neither more nor less inclined to purchase lottery tickets from that particular vendor in the future.

One caveat to the above claim is that the sale of a winning ticket might have an informational or advertising component that induces a rational consumer response. Someone could fully comprehend the randomness of the lottery and yet respond to the sale of a winning ticket simply because the event of having a winner brings the lottery to the fore of his attention or because the hoopla and chatter surrounding this event increase his entertainment value of playing the lottery. Our empirical analysis attempts to distinguish between these two potential responses.

### *B. Addiction*

The random assignment of demand shocks for lottery tickets also allows us to present evidence on the extent to which lottery tickets are addictive. As early as Hicks (1965), economists have recognized that preferences for current consumption may be affected by past consumption levels. Pollack (1970) describes a formal model in which the marginal utility of current consumption depends on past consumption.<sup>6</sup> Consider a simple model of lifetime consumption with two goods, where  $Y_t$  is the numeraire and  $g_t$  is potentially addictive. Consumers maximize the discounted sum of period-specific

utility  $U = \sum_{t=0}^{\infty} (1 + \rho)^{-t} u_t$ , where  $u_t = u(y_t, g_t, g_{t-1})$ . Note that utility is a function of

both current and past values of  $g$ . We say that  $g$  is addictive if  $\frac{\partial^2 u_t}{\partial g_t \partial g_{t-1}} > 0$  and

$\frac{dg_t^*(p_t, I_t, g_{t-1})}{dg_{t-1}} > 0$ . In other words, good  $g$  is addictive if past consumption

---

<sup>6</sup> Pollack explicitly assumes that consumers do not foresee the effect current consumption will have on future preferences or choices. Ryder and Heal (1973), Stigler and Becker (1977), Boyer (1978), Iannaccone (1986), and most rigorously Becker and Murphy (1988) incorporate this forward-looking behavior into the model. Rational addicts, as they are called, take into account that current consumption will increase the

of  $g$  influences current consumption of  $g$  by increasing the marginal utility of current consumption.

This reduced-form definition of addiction is commonly used in the economics literature (e.g. Becker and Murphy (1988), Becker, Grossman and Murphy (1994), Gruber and Koszegi (2001)). It is distinguished from a physiological or psychological definition in that learning-by-consuming qualifies as addiction. We follow the economics literature and use this definition, though the distinction between learning about a product through consumption and physiological or mental addiction should be kept in mind when interpreting the empirical results. Note that the economics definition allows for both harmful ( $\partial u_t / \partial g_{t-1} < 0$ ) and beneficial ( $\partial u_t / \partial g_{t-1} > 0$ ) addictive goods. Addiction is determined by past consumption's effect on *marginal* utilities, not by its effect on the *level* of utility.

In some instances, the literature makes a distinction between addiction and habit formation. In models of addiction, lagged values of consumption affect current marginal utilities. In models of habit formation, consumers endogenously (or exogenously) develop “normal” levels of consumption. The marginal utility of consumption is then related to the distance from these normal levels. More general models of addiction, such as Becker and Murphy's (1988) model of rational addiction, allow the entire history of consumption to affect current utility in a fairly unrestricted way. These models describe both addiction and habit formation. Following these general models, our definition of addiction does not distinguish between addiction and habit formation.

---

marginal utility of consumption in future periods, and that when current consumption becomes past consumption it will have either a negative or positive effect on utility.

Empirical tests of addiction often hinge on observational correlations between past and present consumption. These tests have trouble distinguishing serial correlation from true state dependence.<sup>7</sup> In contrast to these studies, we estimate the demand response to the sale of a winning lottery ticket – which consumers may view as a change in the subjective probability of winning, and thus as a change in the price. Our empirical strategy exploits the random variation in winner location, which generates exogenous shocks to demand. To test for addiction, we look for persistence in this response. It is the exogenous shock to sales coming from the sale of a winning ticket that allows us to distinguish between serial correlation and persistence.

As we explain in the next section, we must make two key identifying assumptions. First, we assume that conditional on contemporaneous sales, the store that sells the winning ticket is randomly selected. This is a weak assumption and follows from the fact that each individual ticket has the same chance of winning in a given drawing. Because consumers cannot predict the location of future winners, the random assignment implies that conditional on contemporaneous sales neither the level nor the trend in store-level sales should be correlated with the sale of a winning ticket in the future. Any residual correlation between subsequent same-store lottery sales and the sale of a winning ticket must be causally related to sale of the winner.

Second, to interpret persistence as addiction, we must assume that the initial consumption response to the sale of the winner does not occur with a long lag. We aim to

---

<sup>7</sup> An exception is Mobilia (1993), which uses instrumental variables techniques to test for rational addictive behavior in the context of horse track betting. The main implication of the Becker-Murphy (1988) rational addiction model is that gambling consumption in consecutive periods is positively related. Following Becker, Grossman, and Murphy (1994), the author employs a two-stage least squares approach, with past and future prices serving as instrumental variables for past and future consumption. The empirical results of the test, based on panel data from 148 horse tracks from 1950 to 1987, are consistent with rational

identify the causal effect of current consumption on future consumption, as opposed to a persistent direct effect of the sale of the winning ticket. To be precise, say the sale of the winning ticket directly causes increases in same-store sales for  $D$  periods. This effect could be a result of an irrational belief that the winning store has become “lucky” or of an increase in advertising at the winning store. Further denote  $P$  the number of periods for which we estimate that sales at the winning store persistently remain above its pre-winner level. As long as  $P > D$ , we can interpret extended increases in sales as responses to lagged consumption. In other words, if sales remain high between period  $D$  and period  $P$ , the increase must be causally related to the increase in sales between period 0 and period  $D$ . According to the definition described above, this would be evidence of addiction.<sup>8</sup>

Note that the assumption that  $P > D$  along with conditional random assignment of winners are the necessary assumptions for the sale of a winning ticket to be a valid instrument in a regression of current lottery consumption on lagged lottery consumption.<sup>9</sup>

### **III. Data and Background on the Texas State Lottery**

To analyze the short- and long-run effects of the sale of a winning lottery ticket, we compile a dataset that includes weekly store-level sales of lottery tickets by game, the location and jackpot size of winning tickets in three lotto games, and zip-code-level

---

addiction. Gruber and Koszegi (2001) use an instrumental variables approach to test for rational addiction in the context of cigarette smoking.

<sup>8</sup> This conclusion is not dependent on the specific form of addiction described in the model above; it follows even if utility were a function of the full history of past consumption. Under the null hypothesis of no addiction, no past consumption has any causal effect on current consumption.

<sup>9</sup> Recent empirical work on addiction in the economics literature (e.g. Becker, Grossman and Murphy (1994), Gruber and Koszegi (2001)) has focused on the distinction between myopic and rational addiction. Myopic addicts ignore the future implications of current consumption on marginal utilities while rational addicts take these effects into account in a fully time-consistent manner. One test often used to distinguish these two models is whether anticipated future increases (decreases) in the price of an addictive good decrease (increase) current consumption. We are not able to distinguish these two models because the variation in winners, which we employ as a shock to demand, is unpredictable by consumers. Thus, neither myopic nor rational addicts should respond to future sales of winning tickets. If consumers’ initial response

demographics for each lottery retailer. The data span the period from January 2000 to June 2002 and cover every lottery retailer active in the state of Texas during the period under study. Weekly counts of store-level sales by game were obtained through Open Records agreements with the Texas Lottery Commission. Table 1 reports descriptive statistics. During the sample period, there are 24,400 active lottery retailers in Texas spread across 1,386 cities and 3,660 nine-digit zip codes. On average, there are 827 retailers per city and 30 per zip code.

We analyze the effects of the sale of winning tickets in the three large lotto games in the Texas Lottery system: Lotto Texas, Texas Two Step, and Texas Cash Five. As can be seen in the bottom panel of Table 1, these games typically offer very different prizes. Lotto Texas jackpots range from \$1.03 million to \$51 million; Texas Two Step jackpots range from \$200,000 to \$1.6 million; and Texas Cash Five jackpots range from \$8,888 to \$93,201. More details on each game are available in Appendix 2. We focus on the sale of winning tickets from *lotto* games as opposed to other types of state lottery games because lotto games have larger top prizes and fewer winners. For the same reason, we focus on the sale of winning *on-line* games, as opposed to *instant* games.<sup>10</sup>

---

to the sale of a winning ticket were irrational, this mistake would be magnified by either rational or myopic addiction.

<sup>10</sup> Instant game tickets contain a series of symbols or numbers that are covered by a secure "scratch-off" coating. Once the coating has been scratched off, the ticket reveals on-the-spot whether the player won or lost. Online lottery games, of which lotto is the most popular, are the most common lottery game in the US. They are played through sales terminals linked to central computers via communications networks. Players select a set of numbers from a larger pool, for example, six from forty-nine or five from thirty. Lotto jackpots can either be fixed or rolling, meaning that if no ticket contains the winning number in a given week, the jackpot "rolls over" and accumulates. Multi-state lotto games, such as PowerBall and The Big Game, allow players from multiple states to contribute to one prize pool, thus offering larger jackpot potential than a single-state lotto game. Numbers games constitute another common form of on-line lottery games. Players typically select a three or four-digit number and winning numbers are typically drawn every day. Players try to select the exact order or any combination of the drawn number. These games offer fixed prizes; typically, the top prize is \$2,500 for a four-digit game and \$250 for a three-digit game. A more recent addition to the lottery game product mix is five-minute keno, an online game that provides instant results to the player.

The Texas lottery brings in roughly \$2.8 billion in sales annually. Over the period we observe, lottery retailers in Texas averaged weekly sales of \$2,576. Retailers averaged weekly sales of \$733 on Lotto Texas, \$110 on Texas Two Step, and \$170 on Cash Five. Texas Lottery also offers an afternoon and evening three-digit numbers game and a variety of scratch-off games. Prior to Texas Two Step, Texas Lottery offered Texas Million, which only had five winners in 2000 and 2001. Ticket sales from all games are included in total sales amounts.

We link sales data to information about where and when winning tickets are sold. We observe the week in which a top prize is won on a particular game, the amount of the top prize, the retailer that sold the winning ticket, and the zip code and city in which the retailer is located. As shown in the middle panel of Table 1, our sample period includes 68 winning jackpots from Lotto Texas, 55 from Texas Two-Step, and 571 from Cash 5. These winners are spread across 669 retailers and 480 zip codes. This information is publicly available and posted on the Texas Lottery's website.

The detailed sales data linked to information on where winning tickets were sold allow us to track sales at stores after they sell tickets that win large jackpots. Because we know the address of stores, we are able to track the sales at nearby stores to test whether there is spatial substitution by consumers. Because we have sales of different lotto games, we can test whether there is substitution within stores across different games. We then link stores to zip-code-level demographic data from the 2000 U.S. Census, allowing us to test whether the consumption response following the sale of a winning ticket is more or less pronounced in zip-codes with different characteristics.

#### IV. Short-Run Sales Effects: Substitution, Lucky Stores and Advertising

In this section, we present evidence on the short-run effects of the sale of a winning lottery ticket on lottery ticket sales. We then garner additional evidence to aid in interpreting the economic implications of these responses. To uncover the effect on sales at the winning store one week after the winning ticket is sold we estimate the following specification separately for each game  $j=\{1,2,3\}$ :

$$[1] \quad g_{it} = \alpha_1 + \gamma_1 w_{i(t-1)} + \phi_1 g_{i(t-1)} + \mu_t + e_{1,it}$$

where  $i$  indexes stores,  $t$  indexes weeks,  $g$  is the log of the number of tickets sold,  $w$  is a dummy variable indicating that store  $i$  sold a winning lotto ticket for game  $j$  in week  $t-1$ ,  $\alpha$ ,  $\gamma$ , and  $\phi$  are parameters to be estimated and  $\mu$  is a fixed week effect, which among other things, captures week-to-week variation in the jackpot, which in practice is closely linked to sales. The estimated effect of selling a winner is thus the effect relative to other stores that week, controlling for the fact that all stores will sell more tickets when the jackpot is very high and fewer tickets when the jackpot is very low.

The more tickets a store sells, the more likely it is to sell a winning ticket. Since sales are serially correlated, it follows that

$$E[\varepsilon_{ijt} \mid w_{ij(t-1)} = 1] \neq E[\varepsilon_{ijt} \mid w_{ij(t-1)} = 0]$$

and therefore a simple comparison of average sales at stores that sold and did not sell winners one week ago does not recover the causal effect of the winning ticket sale. Fortunately, since each lottery ticket has the same chance of winning, the probability of a store selling a winning lottery ticket is a linear function of the number of tickets it sells in



a week.<sup>11</sup> Thus, conditional on the number of tickets sold in week  $t$ , each store has the same chance of selling a winning lottery ticket. Therefore,

$$E[\varepsilon_{ijt} \mid w_{ijt(t-1)} = 1, g_{i(t-1)}] = E[\varepsilon_{ijt} \mid w_{ijt(t-1)} = 0, g_{i(t-1)}].$$

Serial correlation is not a problem for the estimation of  $\gamma_1$  because any two stores with the same sales in week  $t-1$  have the same chance of selling a winning ticket in week  $t-1$ , regardless of whether sales have been high for a large number of weeks or if sales are only high for one week as a result of a temporary shock. Therefore, conditional on sales in week  $t-1$  ( $g_{it-1}$ ),  $w_{ijt-1}$  is randomly assigned, and a simple Ordinary Least Squares (OLS) estimate of  $\gamma_1$  will be unbiased. Some readers may be troubled by the inclusion of a serially correlated lagged dependent variable as a regressor. While it is true that the resulting estimate of  $\phi_1$  is not the causal effect of lagged sales on current sales, the logic above still ensures that the estimate of  $\gamma_1$  is unbiased.

This specification allows for the estimation of longer lags, by game, simply by estimating

$$[2] \quad g_{it} = \alpha_k + \gamma_k w_{i(t-k)} + \phi_k g_{i(t-k)} + \mu_t + e_{k,it}$$

where  $k$  is the number of weeks after the winner is sold for which the effect is to be estimated. In the following section, we estimate the dynamics of the sales effects up to

---

<sup>11</sup> This is not entirely accurate. The probability of selling a winning ticket is a linear function of the number of unique number combinations sold. If store A sells  $X$  tickets, representing  $\lambda X$  unique combinations, and store B sells  $X$  tickets, representing  $\lambda' X$  unique combinations, where  $\lambda' > \lambda$ , then store B has a higher probability of selling a winning ticket. We have no way of knowing in the data how many unique number combinations were sold, only how many tickets. Furthermore, there is no reason to suspect that the proportion of tickets that reflect unique combinations varies systematically across stores. We thus make the simplifying assumption that the number of unique number combinations a store sells is a fixed proportion of the number of tickets the store sells.

40 weeks after the sale of the winning ticket. To do this we estimate 40 different versions of the above specification.<sup>12</sup>

One nice feature of this specification is that it allows for a clean and direct test of the identifying assumptions. If these assumptions hold, estimates of lead effects (i.e. estimates of  $\gamma_k$  in the above equation with negative values of  $k$ ) should be indistinguishable from zero. Consider the one-week lead specification. Conditional on sales in week  $t+1$ , the fact that a store will sell a winning ticket in week  $t+1$  should be uncorrelated with sales in week  $t$ . In other words, the logic used above to argue that serial correlation was not a problem for estimating the  $\gamma_k$ 's implies that last week's sales should not predict whether a store will sell a winning ticket this week, once we have conditioned on this week's sales. Much in the spirit of an event-study, the  $\gamma_k$ 's should be zero in the weeks leading up to the sale of the winner (the lead effects). In the weeks following the sale of the winner the  $\gamma_k$ 's (the lag effects) are free to follow whatever path lottery consumers choose.<sup>13</sup>

---

<sup>12</sup> Alternatively, we could model lottery ticket sales at retailer  $i$  in week  $t$  ( $g_{it}$ ) as a function of store fixed effects ( $\alpha_i$ ), week fixed effects ( $u_t$ ), and a set of dummy variables indicating the sale of winning ticket in the

$j^{\text{th}}$  week before and after week  $t$ :  $g_{it} = \alpha_i + \mu_t + \sum_{k=1}^K \gamma_{t+k} w_{i(t+k)} + \sum_{k=1}^K \gamma_{t-k} w_{i(t-k)} + \varepsilon_{it}$ . We would

interpret the vector of estimates  $\{\gamma_{t-k}\}$  as the effect of selling a winning ticket in week  $t-k$  on sales in week  $t$ . The lead indicator variables,  $w_{i(t+k)}$ , could be used to test the identifying assumption. This specification has intuitive appeal, but does not mathematically capture the randomization process. In particular, this "fixed-effects specification" does not control for intertemporal variation in sales within stores. Stores that have particularly high sales in a specific week are more likely to sell a winning ticket. To the extent that this unobserved shock to sales is serially correlated, we would erroneously attribute increases in sales, possibly both before and after the winner is sold depending on the form of the serial correlation, to the sale of the winning ticket. We therefore prefer the specification presented in the text.

<sup>13</sup> An alternative strategy is to impose a constant parametric rate of persistence  $\pi$  such that  $\gamma_k = \pi^{k-1} \gamma_1$ .

We solve for  $\gamma_1$  and  $\pi$  using a Generalized Methods of Moments (GMM) technique. We use GMM to solve the following optimization problem

$$\min_{\gamma_1, \theta} \left\{ (\hat{\gamma}_k - f_k(\gamma_1, \pi))' \hat{V}^{-1} (\hat{\gamma}_k - f_k(\gamma_1, \pi)) \right\}$$

*A. Short-run effects: Same-game sales at the winning retailer*

Table 2 presents the separate lags results from OLS estimation of equation [2]. The equation is estimated separately for each of the three games. Each entry in the table is the estimated effect of selling a winning ticket from a separate lag or lead regression. The effects one-week subsequent to the sale of the winning ticket can be seen in the row labeled “1 week lag”. All regressions control for contemporaneous (i.e. measured the same week the winning ticket was sold) log sales, and week effects. The dependent variable is the log of store-level game-specific sales.

Within-game estimates suggest that same-store sales increase significantly the week after a store sells a winning ticket. The estimated sales increases are precise enough to rule out the absence of an effect and are large in magnitude, ranging from 12 log points to 32 log points. The size of the response is positively and monotonically related to the average size of the jackpot offered by the game. The largest increases are for Lotto Texas, which is the only game to offer multi-million dollar jackpots. Lotto Texas winning retailers experience a 32.0 log point, or 37.7 percent, increase in Lotto Texas sales the following week. The standard error of this estimate is 3.7 log points. In comparison, stores that sell winners of Texas Two-Step, a midsize game with typical jackpots between \$500,000 and \$700,000, experience a 19.5 log point, or 21.5 percent, increase in sales the following week. The standard error of this estimate is larger (6.5 log points) because there are fewer Texas Two-Step winners in our sample. And lastly, winners of the smallest jackpot game in our sample, Cash 5, experience a same-store

---

where  $f_k(\gamma_1, \pi) = \pi^{k-1} \gamma_1$  and the OLS-estimated coefficients  $\hat{\gamma}_k = \{\hat{\gamma}_1, \dots, \hat{\gamma}_k\}$  are treated as data. The weighting matrix  $\hat{V}^{-1}$  is the inverse of the Newey-West estimate of the variance-covariance matrix of the

sales increase of 11.7 log points, or 12.4 percent. Though smaller in magnitude, this estimate is statistically significant with a standard error of 1.8 log points.

We also estimate equation [2] with an additional term constructed by multiplying the indicator for a winner by the “excess” jackpot prize amount. The excess prize is defined as the jackpot prize above the minimum jackpot prize for that game during our sample. The minimum prize amounts are \$1,032,666 for Lotto Texas, \$200,000 for Texas Two-Step, and \$8,888 for Cash 5. Defining the interaction term this way yields the interpretation of the coefficient on the main effect as the effect of selling a winning ticket for the minimum jackpot prize. Table 3 presents the results. In order to focus on the initial consumer response, coefficient estimates are presented for the 1-week through 5-week effects. For Lotto Texas, relative to stores that did not sell a winner in the previous week, the winning retailer experiences a 24.4 log point increase in sales, with an additional 9.2 log point increase for every additional \$10 million in jackpot prize. This implies that the sale of a winning ticket on a \$10 million jackpot increases relative sales by 33.6 log points (39.9 percent) in the subsequent week. The data do not provide evidence that the response to the sale of a winning Texas Two-Step or Cash 5 ticket depends on within-game variation in the top prize amount. This is not entirely surprising given that these games offer substantially smaller jackpots.

### *B. Testing the identifying assumption*

Our model facilitates a direct test of the identifying assumption that stores are randomly selected conditional on contemporaneous sales. If this identifying assumption is satisfied, sales in week  $t$  should not predict whether a store sells a winner at a future

---

estimated  $\hat{\gamma}_k$ 's.

week  $t+k$ , controlling for sales in week  $t+k$ . In other words, contemporaneous sales should be a sufficient statistic for the probability that a store sells a winner. A direct test of this prediction is to estimate equation [2] defining the winning indicator over future weeks. These are the *lead* versions of the *lags* specifications, and the results are presented in each table along with the *separate lags* estimates. The top half of Table 2 presents the lead estimates for the game-specific specifications. Consistent with the prediction of the identifying assumption, none of the lead estimates are significantly different from zero. The point estimates vary in sign and hover around zero.

As can be seen in Figures 1-3, the data show little evidence of a trend in sales leading up to the winning week. There is a sharp discontinuous break in the estimates in the winning week. The figures indicate that, conditional on contemporaneous sales, winning and non-winning stores experience basically the same time-series of sales in the weeks leading up to the sale of the winning ticket. Immediately following the sale of the winner, the winning stores' sales of tickets for that game jump sharply and remain elevated for up to 40 weeks. We discuss the longer-run estimates below, but the comparison of the lead and lag estimates in Figures 1-3 shows clearly that same-store sales of the winning game respond to the sale of a winning ticket.

In summary, the data unambiguously indicate that same-store sales of tickets for the winning game increase the week after a store sells a winner. These increases are large in both economic and statistical terms. The average store sells an additional 276 tickets the week after selling a winning Lotto Texas ticket. This increase is about 11 percent of the week-to-week standard deviation in total retail-level lottery ticket sales; about 38 percent of the week-to-week standard deviation in retail-level Lotto Texas ticket

sales; and 50.2 percent of the week-to-week standard deviation in the change in retail-level Lotto Texas ticket sales.

### *C. Interaction with population demographics*

The sale of a winning ticket (conditional on contemporaneous sales) contains no information about the likelihood that a store will sell a winning ticket in the future. Nor does it contain new information about the odds of winning a lottery jackpot. Therefore, if lottery ticket consumers act as rational, informed investors in lottery gambling, there should be no response to the location of the winning sale. We might thus conclude that the observed sales increase reflects a “mistake” on the part of consumers. (We consider alternative rational explanations for the observed response below.) It is therefore of interest to know whether certain populations appear to be more likely to make this “mistake” than others.

We investigate whether the observed response to the sale of a winning lottery ticket is more pronounced for retail outlets located in zip codes with greater proportions of economically disadvantaged populations. We link 2000 United States Census data on Texas zip codes to our store-level sales data using address information provided by the Texas Lottery Commission.<sup>14</sup> This allows us to estimate the following equation for each game  $j=\{1,2,3\}$ :

$$[3] \quad g_{it} = \alpha_1^d + \gamma_1^{d0} w_{i(t-1)} + \gamma_1^{d1} (w_{i(t-1)} * X_z) + \phi_1^d g_{i(t-1)} + \beta^d X_z + \mu_t^d + \varepsilon_{1,it}^d,$$

where the  $d$  superscript indicates the parameters are from a specification with demographic interactions and  $X_z$  is the proportion zip code  $z$ 's population from a

---

<sup>14</sup> A zip-code is a nine-digit number used by the U.S. Postal Service to facilitate the delivery of mail. The 2000 Census contains demographic information on five-digit zip codes. The average five-digit zip-code in Texas in 2000 was 97.25 square miles. The average five-digit zip code in our sample had a population of

particular demographic group. The OLS estimate of  $\gamma_1^{d0}$  recovers the one-week lag effect on retail-game sales in a zip-code with  $X_z = 0$ . The OLS estimate of  $\gamma_1^{d1}$  recovers the amount by which economically disadvantaged zip-codes exhibit larger responses to the sale of a winning lottery ticket. The variable  $X_z$  is defined alternately as percent of the population 25 and over with less than a high school degree; percent of the population nonwhite; percent of the population 16 and over who are age 65 or above; percent of the population in poverty; and percent of households on public assistance. We also estimate equation [3] with multiple demographics included simultaneously.

Table 4 reports the one-week lagged results from estimating equation [3] for the Lotto Texas game. In the case of all five demographic groups, a greater proportion is associated with an increased sales response to the sale of a winning ticket. Column (1) displays the coefficient estimates when  $X$  is defined as percent of adults less than high school. The estimated  $\gamma_1^{d0}$  is 0.083 (standard error of 0.071) and the estimate of  $\gamma_1^{d1}$  is 1.014 (standard error of 0.260). At the mean proportion of high school drop-outs (26.2 percent), the estimated effect is an increase of 34.9 log points ( $0.083 + 0.262 * 1.014$ ), or 41.7 percent. These estimates imply that if a zip code has no adults with less than a high school degree, the effect is an 8.7 percent increase in sales. But if 90 percent of adults in the zip code have less than a high school degree, ticket sales at the winning retailer double.

Column (2) displays the coefficient estimates when  $X$  is defined as percent of the zip code population identified as nonwhite. The estimated  $\gamma_1^{d0}$  is 0.194 (standard error of

---

13,226. In our sample period, the average five-digit zip code had 25.2 retailers, compared to 6.7 lottery retailers in the average nine-digit zip code.

0.076) and the estimate of  $\gamma_1^{d1}$  is 0.499 (standard error of 0.269). However, the estimated coefficient on the interaction with percent nonwhite is not positive when the regression includes either the interaction with the percent less than high school or the interaction with the percent in poverty, as shown in columns (6) and (7). The data therefore suggest that percent nonwhite is serving as a proxy variable for other demographic characteristics, but does not itself have a direct positive effect on sales following the sale of a winning ticket.

The estimates presented in column (3) imply that if a zip code has no adults over the age 65, the effect of the sale of a winner is an 10.8 log point increase in sales. But if half of a zip code's adult population is over the age of 65, the effect is an 84.3 log point increase in sales. Additional results presented in columns (4) and (5) similarly suggest that the magnitude of the response to the sale of a winning ticket increases with the proportion of a zip code's population that is poor or participating in public assistance. Unfortunately there is not sufficient statistical power to separately identify the effects of these highly-correlated demographic characteristics, as reported in column (8). Nonetheless, these results provide unambiguous evidence that the response to the sale of a winning ticket is more pronounced in zip codes with larger proportions of economically-disadvantaged groups. The results in the final two columns corroborate this interpretation. They reveal that the sales response *decreases* in the proportion of the population with a college degree (column 9) and with the proportion of the population above 200 percent of the poverty line (column 10).

#### *D. Interpretation: Information v. Lucky Store*



The one-week results demonstrate that consumer demand for lottery tickets responds positively to the sale of a winning ticket. We discuss two classes of explanations for this phenomenon. The first class of theories concerns the information associated with a winning ticket. In practice, the sale of a winner is accompanied by advertising, both explicit and word-of-mouth. This advertising may lead to increased sales because it brings lottery tickets to the forefront of consumers' minds, or because consumers consider advertising and lottery tickets to be complementary goods.<sup>15</sup> Alternatively the information may lead consumers to update their subjective estimate of the probability of winning.

The first response may be a purely rational reaction to advertising, while the second response has a psychological element and is not rational because the probability that a ticket wins is not objectively related to the fact that a winner was recently sold nearby. We call both of these responses *information effects* because consumers update their consumption decisions in response to the receipt of information that a winning ticket was sold. Crucially, the only reason these advertising and information effects would not apply uniformly to all lottery retailers in the state is that information about the win flows more easily to those closer to the winning consumer and retail outlet.

The second explanation for the increase in same-retailer ticket sales is that consumers think the store that sells a winning ticket is "lucky". We call this explanation the *lucky store effect*. Just as in the second type of *information effect*, the consumer revises his estimate of the subjective probability of winning when he sees a winning ticket sold. The distinction is that the consumer attaches this probability change to the

---

<sup>15</sup> See e.g. Becker and Murphy (1993) for a model in which advertising and advertised goods are complements.

retailer rather than generally to the game. This distinction allows us to distinguish between the lucky store effect and the information effect based on patterns in the data.

As we will show shortly, the response to the sale of a winning ticket is larger at the winning store than at close-by stores. We can therefore rule out all but very particular information-based explanations. One might argue that the highly localized response—at the winning store and not at the store across the street—is a result of store-specific advertising. We argue that the fact that the particular advertising done by stores in response to the sale of the winning ticket is effective suggests a belief in the lucky store effect. A sign that advertises “Winning Ticket Sold Here” is effective *because* of a belief in the lucky store effect.

The lucky store effect might be considered “overinference”, a mistake that is often discussed in the literature on the law of small numbers. In particular, lottery players are so surprised by the rare event of a retailer selling a winning ticket that they infer variation in the degree of luck across lottery vendors. Rabin (2002) theorizes a similar revision of beliefs by customers of financial analysts. In his model, all analysts are of average ability. Rabin’s consumers, as believers in the law of small numbers, expect analysts to have bad years following good years. When they observe positive performance by an analyst in consecutive years, the customers are so surprised that they incorrectly infer there must be variation in the quality of analysts. While Rabin’s investors are surprised by the second positive signal of quality because of their belief in the law of small numbers, believers in the “lucky store” theory are surprised by the sale

of a winner at a nearby store because the event is so rare.<sup>16</sup> Both, however, over-infer about the underlying probability of an event from a signal that is surprising to them.

*E. Substitution? The estimated effect on aggregate ticket sales*

We now turn our attention to determining whether the increase in sales at the winning retailer reflects an aggregate increase in sales. We have two goals in doing so. First, we seek to determine whether the increase in sales reflect new sales or merely cannibalized sales from other games or retail outlets. As argued above, cannibalized sales from other retailers would unambiguously indicate a *lucky store* effect. Second, if it can be determined that the increase in retail-level sales reflects an aggregate increase in lottery gambling, then we have a unique opportunity to examine the persistence of a positive shock to lottery ticket consumption and test for addiction. In addition, the nature of the initial consumer response—whether the increase in store-specific sales is the result of consumers changing *where* they buy lottery tickets, *whether* they buy lottery tickets, or *how many* lottery tickets they buy—is directly relevant to determining the costs consumers are willing to incur to respond to the sale of a winning ticket. In the former case the consumer changes the location of his lottery purchases; in the latter two cases he reduces consumption of non-lottery goods.

The increase in game-specific sales after the sale of a winner could reflect an aggregate increase in retailer sales, or it could be completely offset by declines in sales of other games at the store. To investigate these possibilities we estimate equation [2] for the pooled sample of games defining the dependent variable as the log of total retailer

---

<sup>16</sup> Using data from Maryland and New Jersey, respectively, Clotfelter and Cook (1993) and Terrell (1994) find evidence that players of the numbers games are less likely to bet on numbers that have recently won. They interpret their results as evidence of the “gambler’s fallacy”. One interpretation of our result that

lottery sales, including sales from all instant and on-line games. All regressions control for lagged log sales separately for each lotto game to take account of the different probabilities of winning across games. The point estimates (not reported) suggest that retailers experience a 3.9 log point (4 percent) increase in lottery sales the week after selling a winning lotto ticket on any of the three games. With a standard error of 2.1 log points an effect of zero is at the margin of the 95-percent confidence interval. The magnitude of the point estimate is consistent with an aggregate increase in retailer sales commensurate with the game-specific point estimates.<sup>17</sup> Though the point estimates are consistent with a positive demand spillover to other games in the week following a winner, they are not sufficiently precise to rule out even complete substitution across games with a 95 percent confidence interval.

To analyze substitution across games more directly, we estimate equation [2] at the retailer level defining the dependent variable as sales of all other lottery games. In other words, the dependent variable is the log of total sales minus sales in game  $j$ , and the regressor of interest is a dummy for whether the store sold a winning ticket in game  $j$ . The results, presented in Table 5, indicate that the sale of a winning Lotto Texas ticket has a *positive* spillover on other games. The one-week estimate shows a 17.7 log point (19.5 percent) increase in sales of other lottery games at that retailer. This estimate is marginally significant with a standard error of 8.9 log points. Thus, there is no evidence that the increase in Texas Lotto sales shown in Table 2 can be accounted for by

---

same-store sales respond positively to the sale of a winner is that it contradicts the gambler's fallacy. A counterargument might be that selling a winning ticket is so rare that the gambler's fallacy does not apply.

<sup>17</sup> Lotto Texas, Cash 5 and Texas Two-Step respectively account for 25.6, 5.8 and 4.0 percent of lottery sales on average. Thus, if sales of non-winning games did not react to the sale of the winner, we would expect to see increases in retailer sales of 8.3, 1.0 and 1.0 percent. The average increase in retailer sales given the sample of winners would be 1.7 percent. Of the 694 winners in our sample, 68 are Lotto Texas, 571 are Cash 5, and 55 are Texas Two-Step.

substitution away from sales of other lottery games. Lead effects are all insignificant and are close to zero in the weeks leading up to the winner.

Analogous estimates for Texas Two-Step and Cash 5 are also positive, which is also inconsistent with *net* substitution across games. However, the estimates are too imprecise to rule out an estimate of zero effect on other games. The estimated effect on other games from the sale of a Cash 5 winner is small enough to rule out a negative effect on other games of more than 1 percent. Since Cash 5 accounts for only four percent of lottery sales, a one-percent decline in other lottery games could account for all of the 11.7 log point increase in Cash 5 sales. Taken together, however, the bulk of the evidence in Table 5 suggests that the increase in game-specific sales at the winning retailer *does* reflect an aggregate increase in retailer sales.

The preceding analysis investigated the possibility of substitution across games within the winning store. We next investigate whether there is evidence that consumers substitute away from nearby stores in favor of the winning store. In particular, we are interested in whether the increase in the winning retailer's sale of tickets on the winning game reflects an aggregate increase in zip code level sales of tickets on that game. In this analysis we are using the 9-digit zip code as a measure (albeit an imperfect one) of the local market. We define the dependent variable as the log of total game-specific ticket sales in the zip code in week  $t$ . The regressor of interest is an indicator for whether any retail outlet in zip code  $z$  sold a winning ticket on game  $j$  in week  $t-k$ . The estimating equation is the following:

$$[4] \quad g_{zt} = \alpha_k^z + \gamma_k^z w_{z(t-k)} + \phi_k^z g_{z(t-k)} + \mu_t^z + \varepsilon_{zkt}^z.$$

Table 6 presents the results separately by game. All zip codes are included in the analysis. The estimates indicate that zip codes with Lotto Texas winners experience a 13.2 log point increase (standard error of 1.9 log points), or 14.1 percent, in Lotto Texas sales. The Texas Two-Step results show similar initial increases. The one-week estimated effect is a 16.0 log point increase in game-specific sales (standard error of 2.4 log points) in the zip code. The Texas Two-Step lead estimates are fairly large and positive, though fairly imprecisely estimated. To be conservative, one might subtract the one-week lead point estimate from the lag effects. The one-week lag point estimate would then be a 10.7 log point increase in sales at the zip code level. The estimated effect on total zip code Cash 5 sales is 8.3 log points (standard error of 0.6 log points), though the Cash 5 lead estimates are consistently positive and precisely estimated. The lag estimates are all about 4 to 5 log points higher than the lead estimates.

To pinpoint the spillover effect on other stores in the zip code, we estimate own-store sales as a function of whether there was a winner in the zip code in week  $t-k$  and whether the particular store sold a winner in week  $t-k$ :

$$[5] \quad g_{it} = \alpha_k^z + \gamma_k^{z1} w_{i(t-k)} + \gamma_k^{z2} w_{z(t-k)} + \phi_k^{z1} g_{i(t-k)} + \phi_k^{z2} g_{z(t-k)} + \mu_t^z + \varepsilon_{izt}^z.$$

The coefficient  $\gamma_k^{z2}$  captures the effect of the sale of a winning ticket in the zip code; the sum of  $\gamma_k^{z1} + \gamma_k^{z2}$  captures the effect of the sale of a winning ticket at the particular store. The equation is estimated separately by game.

Table 7 reports the results. The estimated effects reveal that the sale of a Lotto Texas ticket has a positive effect on other stores' sales; in the week following the sale of a winner, stores in the winning zip code experience an average increase in sales of 5.5 log points (standard error 1.0). However, the winning store experiences an additional 26.6 log

point increase in sales. The sale of a winning Texas Two-Step or Cash 5 ticket does not appear to have a significant spillover effect. Notice also that in contrast to the specification reported in Table 6, the lead effects for all three specifications are close to zero and insignificant. Importantly, the fact that in all three cases the estimate of  $\gamma_k^{z1}$  is significantly positive confirms that the effect is highly localized, and particular to the winning store. This result rules out all but the most localized (i.e. store-specific) information stories, and is consistent with the lucky store effect.

One weakness of the zip-code level regressions is that there is a great deal of heterogeneity in the size of these “local markets”. One might be interested in whether there is more evidence of substitution from stores that are very close to the winning retailer. To this end, we estimate a specification very similar to [5] that defines a local market as a one-mile radius around a retailer. The regression includes one observation per week for each retailer. The dependent variable is the log of game-specific sales in the store in week  $t$ , and the regressors of interest are indicators for whether a winning ticket was sold in the store and whether a winning ticket was sold within one mile of the store in week  $t-k$ . Regressions control for log game-specific sales in week  $t-k$  at both the store and one-mile group level. Stores are dropped if there are no other retailers within one mile.<sup>18</sup>

The results confirm that the demand shock is significantly larger at the winning store than at nearby stores. The results for the first five weeks are reported in Table 8.

While there appeared to be a positive spillover to non-winning stores in the same zip

---

<sup>18</sup> One-mile groups are defined around an individual retailer and therefore, unlike zip codes, they are not mutually exclusive. As an individual retail outlet can belong to more than one group, spatial correlation among retail observations is potentially an issue. To account for this, we adjusted the standard errors for clustering at the nine-digit zip code level; the standard errors were virtually unchanged.

code, there is no evidence of a similar spillover to stores within one mile. For all games the estimated lag effects at non-winning stores are insignificantly different from zero, and for Texas Two-Step the point estimates are negative. The point estimates for the stores within a one-mile radius are all smaller in magnitude than those for stores within the same zip code. The one-mile results therefore suggest that there is indeed more substitution away from nearby stores than from other stores in the zip code. Just as the previous estimates showed, the one-mile estimates clearly indicate that the effect of a winning ticket is much larger at the winning store than at nearby stores. Again, these estimates are inconsistent with a general information story. Whatever information produced by the sale of the winning ticket would have to accrue to consumers who frequent the winning store and not to consumers who shop at stores within one mile of that store.

The only viable information-based explanation is that the winning store changed its advertising within one week of selling the winning ticket. The results of a 5-minute telephone survey of the retailers who sold winning Texas Lotto tickets suggests winning retailers post a “Winning Ticket Sold Here” sign that they receive from the Texas Lottery Commission (TLC). No winning retailers reported otherwise changing their advertising behavior. It is informative to ask why such advertising is effective. One possibility is that consumers learn that the lottery exists. However, all lottery retailers reported posting signs advertising the lottery in general prior to selling the winner.<sup>19</sup>

Alternatively, uninformed consumers might learn that the probability of winning is greater than zero. Since there is greater local media coverage, one would expect a “Winning Ticket Sold Here” sign to contain less marginal information when the jackpot



is larger. As indicated in Table 3, however, the response is largest for the game with the largest jackpot, and is increasing in the size of the jackpot for this game. This leaves the possibility that the “Winning Ticket Sold Here” sign is effective precisely because consumers believe in the lucky store effect.

Taken together, the results presented thus far suggest that the sale of a winning ticket leads to an aggregate increase in lottery sales. There is clear evidence that game-specific sales increase at the winning retailer. This increase could be accompanied by substitution across games within the store or by substitution across local stores. The retail level results suggest that the increase in game-specific sales reflects an aggregate increase in retail lottery sales; the estimates are not consistent with a *net* reduction in retailer sales of tickets for other lottery games. Point estimates indicate there is actually some *positive* spillover effect to other lottery games sold at the winning retailer. The zip-code-level results suggest that the increase in retail-level sales is associated with an aggregate increase in game sales in the zip code; the estimates are not consistent with a *net* reduction in the winning game’s sales at other retailers in the zip code, though imprecise estimates suggest there may be some such substitution by stores within one mile of the winning retailer. We therefore conclude that the sale of a winning lottery ticket leads to an initial aggregate increase in lottery sales.

## **V. Long-Run Sales Effects: Persistence, Habit-Formation and Addiction**

In this section, we turn to the longer-run effects of the sale of a winning ticket. First, we show that the short-run sales response exhibits some persistence. Under certain identifying assumptions, which we will specify shortly, we can causally link long-run

---

<sup>19</sup> The details of the survey and the full set of results are reported in Appendix 3.

increases in lottery sales to the short-run demand shock uncovered in the previous section. If the increase in lottery consumption 40 weeks later was *caused* by an increase in lottery consumption one week following the winner, we can conclude that lottery gambling is addictive. Clearly, to make this inference we must assume that the persistent effect is not caused by persistence in the underlying determinant of the short-run demand shock. For instance, the belief in the lucky store effect must not last for 40 weeks.

To estimate the persistence of the short-run sales responses discussed in the previous section, we estimate 40 versions of equation [2]. Each of the 40  $\gamma_k$ 's recovers the  $k$ -week lagged effect of the sale of a winning lottery ticket on same-store same-game lottery sales. The justification for the consistency of each of these 40 estimates is the same as was for the estimate of the one-week effect,  $\gamma_1$ : conditional on sales in week  $t-k$ ,  $w_{t-k}$  is randomly assigned.<sup>20</sup>

Once we have estimated these reduced-form effects, we will then interpret them within a simple model of the lifetime consumption of an addictive and a non-addictive good. This model makes clear the assumptions necessary to interpret these long-run reduced-form responses as evidence of addiction.

#### A. A model of addiction

---

<sup>20</sup> Recall that each  $\gamma_k$  is estimated from a separate OLS regression. Consistency is ensured by the fact that for all  $k$ ,  $E[\varepsilon_{ijt} \mid w_{ijt(t-k)} = 1, g_{i(t-k)}] = E[\varepsilon_{ijt} \mid w_{ijt(t-k)} = 0, g_{i(t-k)}]$ . It is by necessity that we estimate the  $\gamma_k$ 's with separate regressions since we must control for  $g_{ijt-k}$  when estimating  $\gamma_k$ . For example, we cannot control for  $g_{ijt-1}$  when estimating  $\gamma_2$  because  $g_{ijt-1}$  is an outcome of  $w_{ijt-2}$ . In fact, we cannot control for any  $g_{ijt-k}$  when trying to estimate the effect of any  $w_{ijt-k-m}$  where  $m > 0$ . Thus, we cannot estimate any two  $\gamma_k$  and  $\gamma_{k'}$ ,  $k \neq k'$ , in the same OLS regression.

The model is as follows. Consider a consumer whose current-period utility depends on current consumption of a numeraire good  $Y$ , current plus  $T$  lags of a potentially addictive good  $g$ , and unmeasured life-cycle variables, which are captured by  $e_t$ . We can write the discounted value of lifetime utility

$$U = \sum_{t=0}^{\infty} \beta^t u(y_t, g_t, g_{t-1}, g_{t-2}, \dots, g_{t-T}, e_t)$$

where  $\beta$  is the discount factor  $1/(1 + \rho)$ ,  $\rho$  being the rate of time preference. Individuals

maximize  $U$  subject to the lifetime budget constraint  $\sum_{t=0}^{\infty} \beta^t (y_t + p_t g_t) = I$ , where

$p_t$  is the price of the potentially addictive good,  $I$  is the present value of lifetime income,

and the interest rate is assumed to be the same as the rate of time preference.<sup>21</sup> For

simplicity, there is no labor supply in the model. Expanding on Becker, Grossman, and

Murphy (1994), we consider a utility function that is quadratic in  $y_t, e_t, g_t$ , and  $T$  lags

of  $g$ . Solving the utility-maximization problem for the first-order conditions for  $y_t$  and  $g_t$

yields the following relationship:

$$\begin{aligned} g_t = & \alpha + \theta_p p_t \\ & + \theta_{-1} g_{t-1} + \theta_{-2} g_{t-2} + \dots + \theta_{-T} g_{t-T} + \theta_1 g_{t+1} + \theta_2 g_{t+2} + \dots + \theta_T g_{t+T} \\ & + \theta_{e_0} e_t + \theta_{e_1} e_{t+1} + \dots + \theta_{e_T} e_{t+T}. \end{aligned}$$

Appendix 1 presents the details.

This equation is the basis for our empirical specification. Consider an observed exogenous shock to  $p_t$ , which would have a direct effect on  $g_t$  and would (as shown in the

---

<sup>21</sup> While the basis for our empirical specification is derived from an individual consumer's utility maximization problem, data on lottery consumption is only available at an aggregate level. To estimate this model empirically we therefore define  $g_t$  as retail-level weekly sales (and take the natural logarithm). For this specification to follow from the model, we must implicitly rely on a representative agent assumption.

appendix) identify the  $\theta_{-k}$  's in future periods. If the shock to  $p_t$  were anticipated, it would also identify the  $\theta_k$  's. We observe when a retailer sells a winning lottery ticket, an event that may heterogeneously affect consumers' subjective estimates of  $p_t$  (e.g., consumers may think the store is "lucky," thereby increasing the probability of buying a winning ticket there). The perceived decrease in  $p_t$  would cause an increase in  $g_t$ , which would become  $g_{t-1}$  the next period, and so on.

We do not estimate the  $\theta_{-k}$  's directly. Rather, we estimate the  $\gamma_k$  's in the way described above. The  $\theta_{-k}$  's can be thought of as the structural parameter of interest, where the  $\gamma_k$  's are reduced form estimates. To recover the  $\theta_{-k}$  's, we must make further assumptions described in below.

Assume that the shock to  $p$ , which is caused by the sale of the winning ticket ( $w$ ), lasts for  $\bar{k}$  periods. We can then be sure that any increase in sales causally related to  $w$  that occurs more than  $\bar{k}$  periods later must be caused by increases in lagged consumption, which were due to the initial shock to  $p$ . Thus, to interpret the persistent reduced-form sales responses as evidence of addiction—to be sure that they are causal responses to lagged lotto consumption—we must assume that the direct effect of the sale of the winning ticket is short-lived. In other words, the "lucky store effect", the advertising effect, or whatever drove the initial sales response documented in the previous section, must not itself be persistent. At the very least, we must assume that this direct effect, or the perceived reduction in  $p$ , does not last as long as the reduced form effects are persistent.

---

More realistically, we suspect that there are heterogeneous consumers and our empirical estimation is

*B. Longer-run effects: Assessing the evidence for addiction/habit-formation*

Figures 1-3 plot the 40 lead and 40 lag estimates of [2] along with 95-percent confidence bands. Table 9 shows the results from similar regressions that are pooled to the month level to improve precision. As can be seen in Table 9 and in Figures 1-3, game-specific same-store sales remain elevated up to 40 weeks after the sale of a winner. The initial 32 log point increase in Lotto Texas sales diminishes quickly in the first ten weeks and then remains constant between 7 and 14 log points. The longer lag effects are quite consistent; the 6, 7 and 8 month lag effects are significant at the 5-percent level, while the 9 and 10 month effects are significant at the 10-percent level. Texas Two-Step estimates show less of a decline, but are less precise. The lower bound of the 95-percent confidence interval of the week-level long-run estimates hovers around zero, and the point estimates range from 9.5 to 33.0 log points with no particular time-series trend. Estimates pooled to the month level are more precise, and can reject the null hypothesis of no effect in all 10 lag specifications.

The Cash 5 estimates are more precisely estimated and are similar to the Lotto Texas estimates. The initial 11.7 log point increase declines to an effect of 6.7 log points after 40 weeks. Because of the larger number of Cash 5 winners, even these smaller long-run estimates are precise enough to rule out the absence of an effect. In contrast to Lotto Texas however, the Cash 5 estimates decline fairly linearly. It is interesting to note

---

identifying the average marginal effect.

that the larger initial Lotto Texas estimate declines more rapidly than the smaller initial Cash 5 estimate.<sup>22</sup>

To determine whether there is an aggregate increase in lottery sales in the long run—which would be clear evidence of addiction given the assumptions described in the previous section—we must examine substitution patterns as we did with the short run results. We focus on substitution across retailers, and first turn to zip-code-level aggregate results. Table 10 shows the estimated effects of selling a winning ticket in game  $j$  in a zip code on subsequent aggregate sales of game  $j$  in that zip code. The data are pooled to the month level to increase precision. Ten months following the sale of a winning ticket, there is a marginally significant (at the 10-percent level) 4.5 log point increase in aggregate sales of Texas Lotto tickets. The nine-month estimate is insignificantly positive, though the earlier eight month estimates are significant and positive in sign. Similarly, the Texas Two-Step specifications show positive and significant sales increases for all 10 months. All Texas Lotto and Texas Two-Step lead estimates are insignificant. The ten Cash 5 estimates are significant and positive, but four of the five lead estimates are also significantly positive, raising the concern that the lag estimates are also picking up a spurious correlation.

In Table 11 we more directly investigate the question of whether the long-run increase in sales at the winning store is accounted for by substitution away from nearby stores or substitution away from non-lottery consumption. The table presents estimates of [5] for lags of up to ten months. Once again, the estimates are pooled by month to

---

<sup>22</sup> The implied week-to-week persistence rates using the estimator described in footnote 10 are 0.959 for Lotto Texas, 0.994 for Texas Two-Step, and 0.989 for Cash 5. This estimator imposes a constant persistence rate and implies that on average for Texas Lotto  $\gamma_k$  is 95.9 percent as large as  $\gamma_{k-1}$ .

increase precision. The results confirm the previous findings that winning stores experience persistent increases in sales. However, the estimate of  $\gamma_k^{z^2}$ , the effect of the sale of a winning ticket for non-winning stores in the zip code, is too imprecise to distinguish the two hypotheses. Point estimates for Texas Lotto are positive, which would be inconsistent with large-scale substitution across stores within the winning zip code, and would point towards fairly clear evidence of addiction. However, in all three specifications the estimates are too imprecise to rule out enough substitution from nearby stores to account for the increase in sales at the winner.

In Table 12, we present long run versions of the one-mile specification. The estimates suggest that non-winning stores within one mile of the winner experience significant decreases in sales of Texas Lotto and Texas Two-Step 6-10 months later. The contrast of these results with those in Table 11 suggest that there is more substitution from stores that are very close to the winner (within a one-mile radius) than from stores within the same zip code but more than one mile away.

Taken together the long-run results do not provide strong evidence that lottery tickets are addictive, though a naïve view of Figures 1-3 would indicate so. There is clear evidence of a persistent increase in game-specific sales at the winning retailer. In contrast, the longer-run estimates do not provide clear evidence for persistence of the increase in aggregate zip-code-level sales.<sup>23</sup> *The data suggest that in contrast to the*

---

<sup>23</sup> This finding is potentially consistent with recent work on the effectiveness of state lottery advertising. Zhang (2004) studies the sales response to legislated reductions in state lottery advertising in three states. If lottery gambling were addictive or habit-forming, then an initial increase in sales brought about by advertising would persist even after advertising terminated. However, Zhang finds that 90 percent of the cumulative effect of advertising depreciates within three to nine months. She interprets her results as potentially being driven by the form of advertising employed; a potential alternative explanation for the finding rests with the nature of lottery gambling itself.

*initial increase in game-specific sales at the winning retailer, which reflects in part an aggregate increase in sales, the long-run increase in game-specific sales at the winning store is potentially accounted for by permanent shifts in the location of consumers' purchases.*

In sum, we have determined that the increase in game-specific sales at the winning retailer reflects an increase in total ticket sales at the winning retailer and in game-specific sales in the zip code. This implies that the initial increase in retail-game sales is not driven entirely by cannibalized sales from other games or non-winning stores.

We have further shown evidence that the increase in sales at the winning vendor is persistent. To interpret this persistence as addiction, we must show that the initial aggregate increase in lottery sales causally leads to subsequent aggregate increases. The data do *not* show this clearly. The data are consistent with the long-run increase in sales at the winning vendor being completely offset by a permanent substitution away from other retailers selling tickets on this game. We therefore do not find strong evidence of addiction of lottery gambling in the long run. Our estimates are, however, too imprecise to rule addiction out.

It should also be pointed out that even a precisely estimated zero effect on sales 40 weeks later would not rule out addiction. The initial increase in gambling could causally increase gambling the following week but by a small enough amount that the effect would fully dissipate 40 weeks later. What we can say is that lottery gambling does not seem to be so addictive that a shock to demand causes permanent changes in behavior.



One explanation of the long-run findings is that some fraction of consumers permanently revises their subjective estimate that a lottery ticket purchased at the winning store will be a winner. This is a persistent lucky store effect. A second explanation of the long-run findings is that it is a rational response to the highly localized (and persistent) advertisement of a winner. As discussed earlier, one should ask why this advertising is effective if not because the consumer believes in the lucky store theory. These first two explanations for the long-run patterns in the data were precisely what we assumed away in order to identify gambling addiction.

A third explanation of the long-run results is that consumers initially shift their consumption of lottery tickets in response to the sale of the winner (a lucky store effect), and then get used to shopping at the new store. While this explanation implies that there is a causal effect of lagged consumption on some attribute of current consumption, it would seem more reasonable to label this behavior habit formation of shopping patterns than addiction of lottery gambling *per se*.

## **VI. Conclusion**

Detailed sales data from the Texas Lottery clearly indicate that the sale of a winning jackpot ticket leads to an increase in ticket sales at the winning retail outlet. In particular, the sale of a winning Lotto Texas ticket leads to a 38 percent increase in game-specific sales the following week, controlling for contemporaneous sales and week fixed effects. The effect is most pronounced for retail outlets located in zip codes with high proportions of economically-disadvantaged populations. Additional analyses suggest that this increase does not reflect a decrease in the winning retailer's sales of tickets on other

lottery games. Nor does this initial increase reflect a net decrease in sales of game tickets at other retail outlets near the winning store. We interpret these findings as evidence that the sale of a winning ticket leads initially to an aggregate increase in lottery sales.

There are two potential explanations for this consumer response. First, consumers could simply be responding to the event of a win at a particular store because the win generates advertising for the lottery; the response is not necessarily to the location of the winner, but rather to the advertising of the lotto game. To fit the patterns observed in the data, this explanation requires that the advertising and its effects are highly localized. In contrast to this purely rational response, the second explanation we put forth is behavioral: consumers believe that the winning store is lucky. The sale of a jackpot ticket by a particular retailer leads such a consumer to irrationally revise his subjective probability that a ticket purchased from that store will be a winner.

The initial increase in game-specific sales depreciates, but not completely. Some 40 weeks after the sale of a winner, the winning store continues to experience an increase in game-specific sales of approximately ten percent. However, there is weak evidence of persistence in the increase in game-specific sales at the zip-code level. And, more direct tests show evidence of substitution away from nearby stores in the long run. We interpret the whole of the evidence as suggesting that some of the initial consumer response is new ticket purchases (either from new consumers or current players who spend more dollars) and some of the response reflects ticket purchases that would have occurred at other retailers. In the long run, the data only clearly indicate an increase in sales at the winning vendor, not in the aggregate. We thus find little evidence that lottery gambling is addictive, though consumers appear to form habits of where they shop.

## References

- Bar-Hillel, Maya, and William Wagenaar (1991). "Perceptions of Randomness," *Advances of Applied Mathematics*, XII: 428-454.
- Becker, Gary S., Michael Grossman, and Kevin M. Murphy (1994). "An Empirical Analysis of Cigarette Addiction," *American Economic Review*, 84(3): 396-418.
- Becker, Gary S. and Kevin M. Murphy (1988). "A Theory of Rational Addiction," *Journal of Political Economy*, 96(4): 675-700.
- Becker, Gary S. and Kevin M. Murphy (1993). "A Simple Theory of Advertising as a Good or Bad," *Quarterly Journal of Economics*, 108(4): 941-964.
- Boyer (1978). "A Habit Forming Optimal Growth Model," *International Economic Review*, 19(3): 585-609.
- Camerer, Colin. (1995). "Individual Decision Making," in ed. John Hagel and Alvin E. Roth, *Handbook of Experimental Economics*.
- Carhart, Mark M. (1997). "On Persistence in Mutual Fund Performance," *Journal of Finance*, 52(1): 57-82.
- Chevalier, Judith and Glenn Ellison (1997). "Risk-Taking by Mutual Funds as a Response to Incentives," *Journal of Political Economy*, 105(6): 1167-1200.
- Chevalier, Judith and Glenn Ellison (1999). "Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance," *Journal of Finance*, 54(3): 875-899.
- Clotfelter, Charles, and Philip Cook (1993). "The 'Gambler's Fallacy' in Lottery Play," *Management Science*, XXXIX: 1521-1525.
- Evans, William and Ping Zhang (2003). "The Impact of Earmarked Lottery Revenue on State Educational Expenditures." University of Maryland working paper.
- Forrest, David O., David Gulley and Robert Simmons (2000). "Elasticity of Demand for U.K. National Lottery Tickets." *National Tax Journal* 53: 853-863.
- Gilovich, T., R. Vallone, and Amos Tversky (1985). "The Hot Hand in Basketball: On the Misperception of Random Sequences." *Cognitive Psychology* 17: 295-314.
- Grinblatt, Mark and Matti Keloharju (2001). "What Makes Investors Trade?" *Journal of Finance*, 61(2): 589-616.

- Gruber, Jon and Botond Koszegi (2001). "Is Addiction 'Rational'? Theory and Evidence," *Quarterly Journal of Economics*, 116(4): 1261-1303.
- Hicks, John. (1965). *Capital and Growth* (New York: Oxford University Press).
- Iannacone, Laurence R. (1986). "Addiction and Satiation," *Economics Letters*, 21: 95-99.
- Kearney, Melissa S. (forthcoming) "State Lotteries and Consumer Behavior" *Journal of Public Economics*.
- Mobilia, Pamela (1993). "Gambling as a Rational Addiction," *Journal of Gambling Studies* 9(2): 121-151.
- Odean, Terrance (1998). "Are Investors Reluctant to Realize Their Losses?" *Journal of Finance*, 53(5): 1775-1798.
- Pollack, Robert A. (1970). "Habit Formation and Dynamic Demand Functions," *Journal of Political Economy*, 78(4): 745-763.
- Rabin, Matthew. (2002) "Inference by Believers in the Law of Small Numbers," *Quarterly Journal of Economics* 117(3): 775-816.
- Ryder, Harl E. and Geoffrey M. Heal (1973). "Optimum Growth with Intertemporally Dependent Preferences," *Review of Economic Studies*, 40(1): 1-33.
- Stigler, George J. and Gary S. Becker (1977). "De Gustibus Non Est Disputandum," *American Economic Review*, 67(1): 76-90.
- Terrell, Dek (1994). "A Test of the Gambler's Fallacy – Evidence from Pari-Mutuel Games," *Journal of Risk and Uncertainty*, 8: 309-317.
- Tversky, Amos and Daniel Kahneman (1971). "Belief in the 'Law of Small Numbers'," *Psychological Bulletin*, 76: 105-110.
- Wagenaar, William (1972). "Generation of random sequences by human subjects: A Critical survey of the literature." *Psychological Bulletin* 77: 65-72.
- Zhang, Ping. (2004) "Over- or Under- Advertising by State Lotteries" in Ph.D. Dissertation. Department of Economics, University of Maryland.

Figure 1

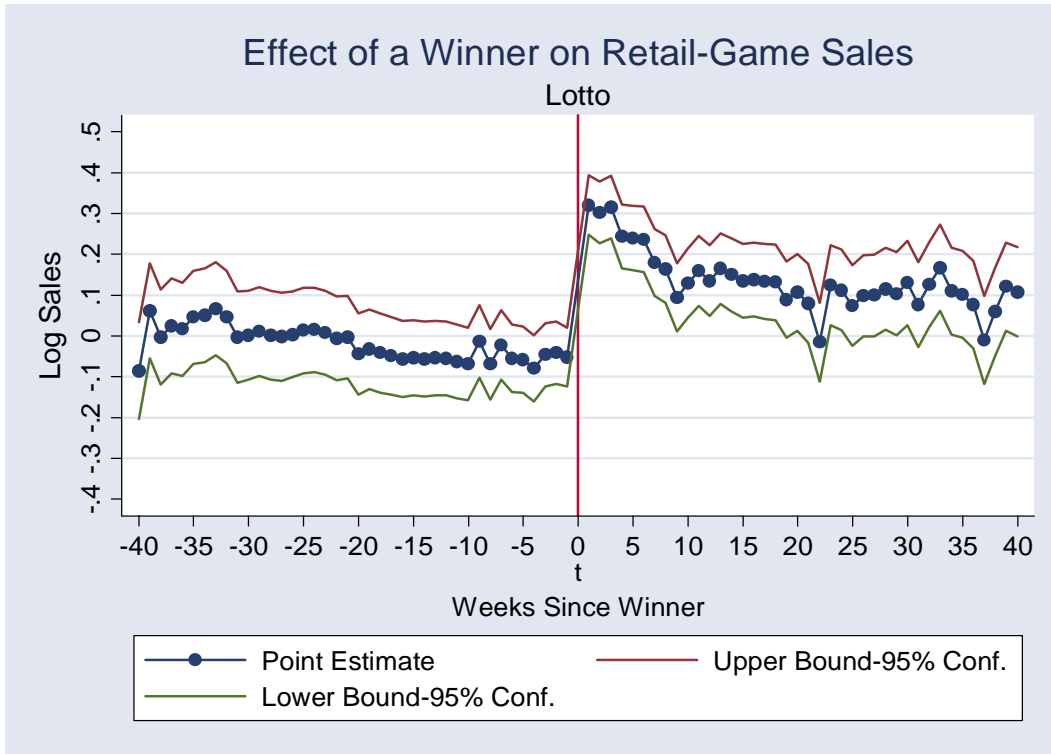
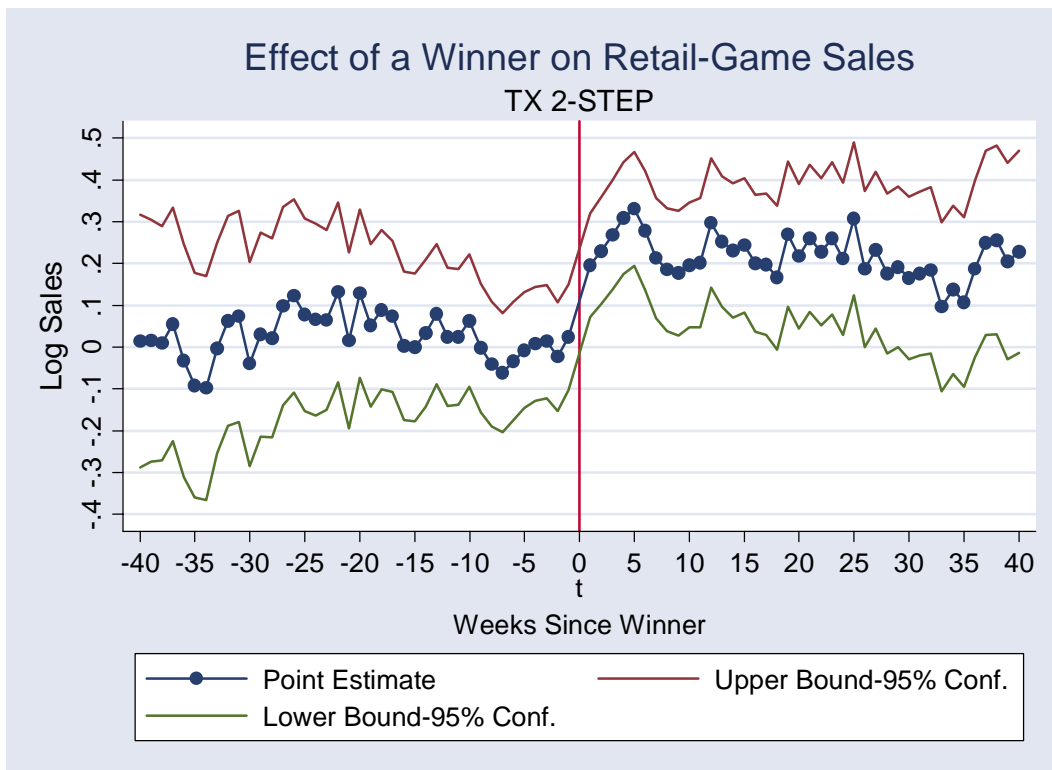
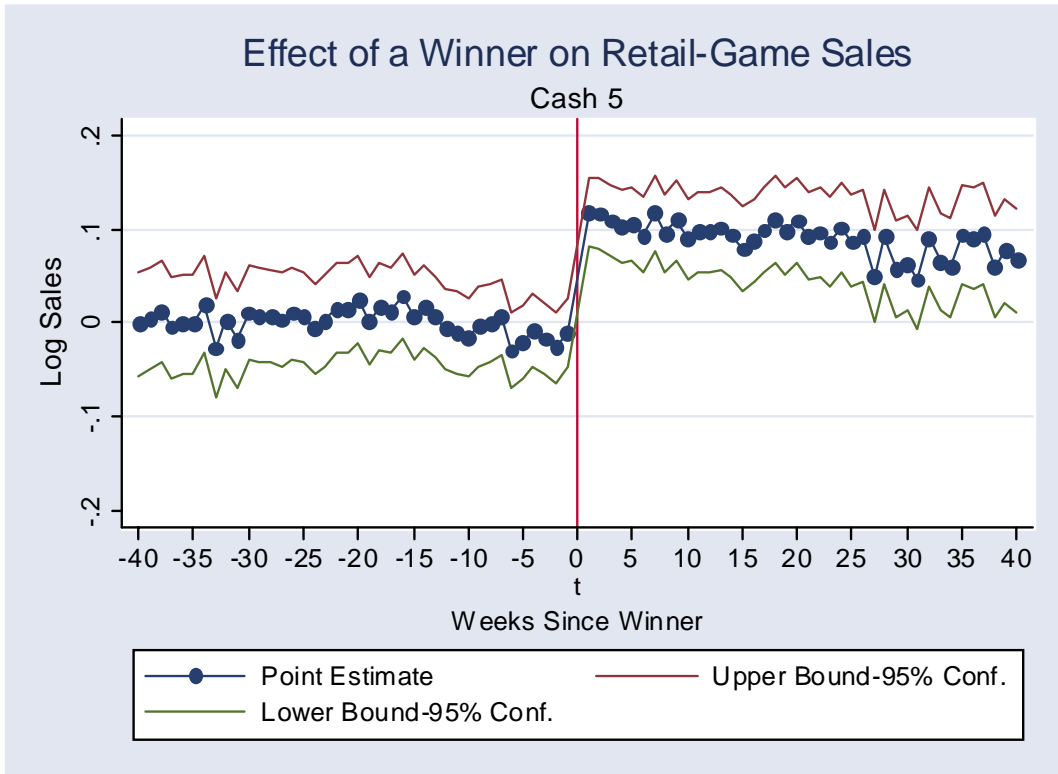


Figure 2



**Figure 3**



**Figure 4**

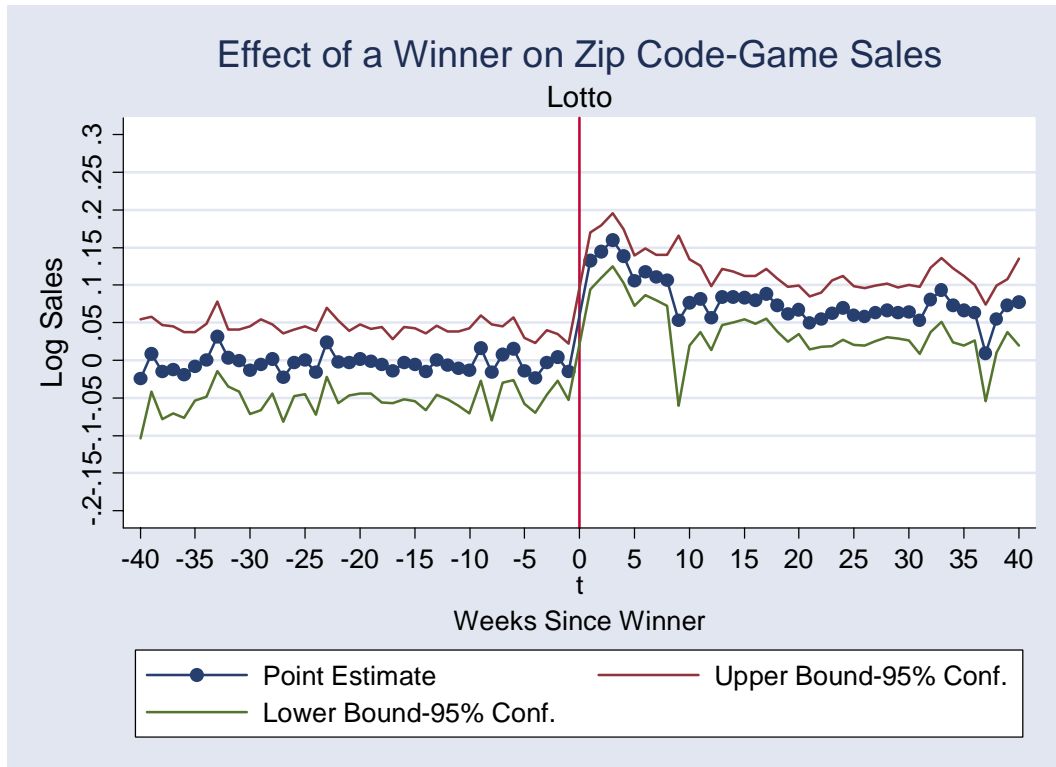


Figure 5

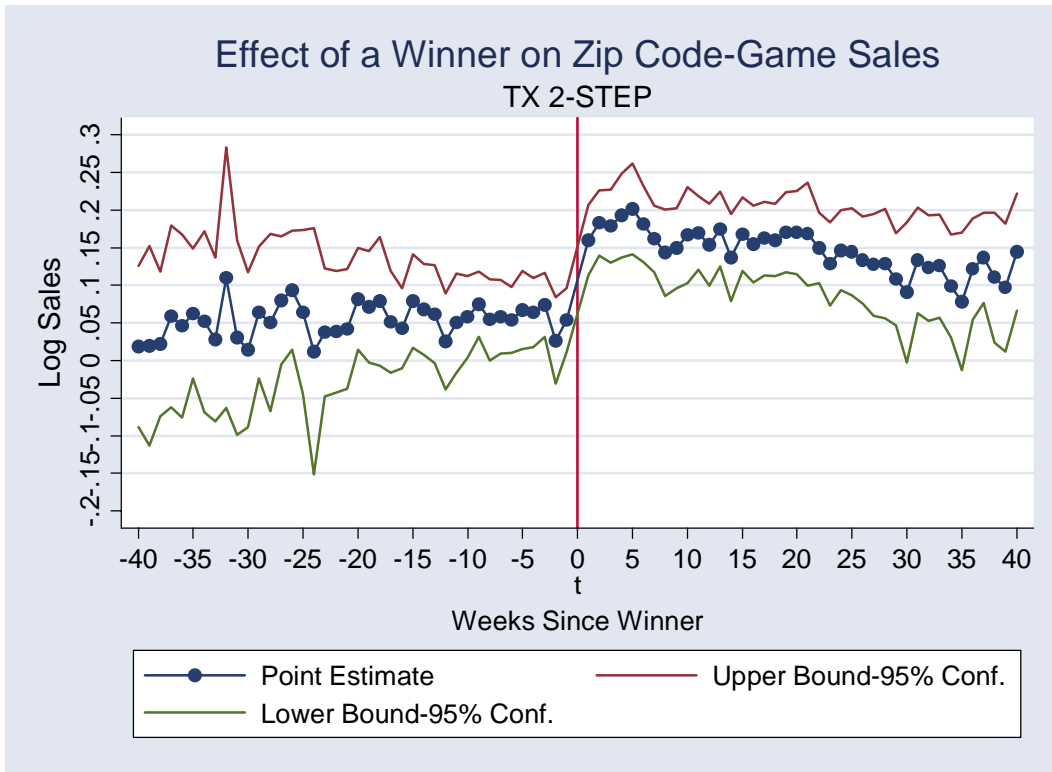
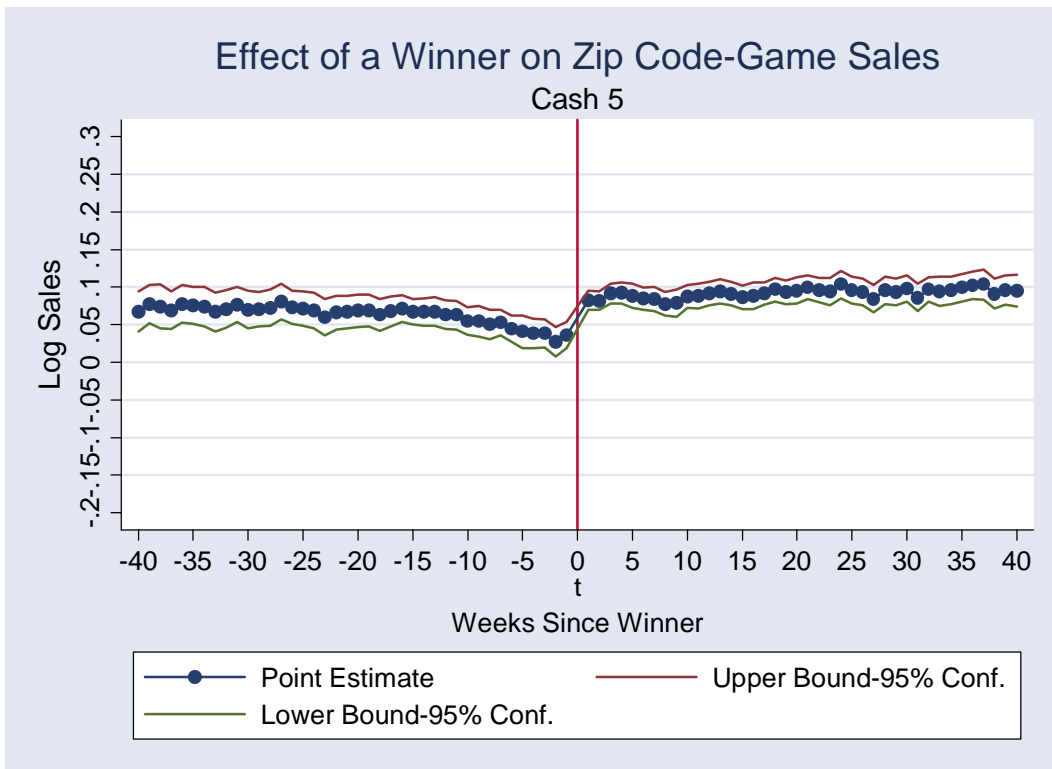


Figure 6



**Table 1: Texas Lottery**

<b>Descriptive Statistics</b>				
	<b>Pooled</b>	<b>2000</b>	<b>2001</b>	<b>2002 (thru 6/02)</b>
<i>Number of retailers</i>	24,400	19,357	22,823	24,278
<i>Number of zip codes</i>	3,660	3428	3571	3660
<i>Number of cities</i>	1,386	1,334	1,372	1,385
<i>Retailers per zip code</i>	6.7 <i>min: 1</i> <i>max: 94</i>	5.7 <i>min: 1</i> <i>max: 74</i>	6.4 <i>min: 1</i> <i>max: 89</i>	6.6 <i>min: 1</i> <i>max: 94</i>
<i>Retailers per city</i>	17.6 <i>min: 1</i> <i>max: 3,734</i>	14.5 <i>min: 1</i> <i>max: 2,952</i>	16.6 <i>min: 1</i> <i>max: 3,477</i>	17.5 <i>min: 1</i> <i>max: 3,721</i>
<i>Total Annual Sales</i>		\$2.81 billion	\$2.77 billion	\$1.33 billion
<u><i>Average Weekly Retailer Sales</i></u>	\$2,576	\$2,746	\$2,486	\$2,518
<i>Lotto</i>	\$733	\$759	\$753	\$630
<i>TX Two-Step</i>	\$110	-	\$118	\$98
<i>Cash 5</i>	\$170	\$194	\$159	\$141
<u><i>Number of Winners</i></u>				
<i>Lotto</i>	68	32	24	12
<i>TX Two-Step</i>	55	-	36	19
<i>Cash 5</i>	571	239	236	96
<u><i>Prize Amounts</i></u>				
<i>Lotto</i>		\$8.2 million <i>min: 1.03M</i> <i>max: 34M</i>	\$12 million <i>min: 1.1M</i> <i>max: 51M</i>	\$7.3 million <i>min: 2.2M</i> <i>max: 18.6M</i>
<i>TX Two-Step</i>		-	\$557,639 <i>min: 200,000</i> <i>max: 1.6M</i>	\$652,632 <i>min: 200,000</i> <i>max: 1.5M</i>
<i>Cash 5</i>		\$48,031 <i>min: 15,897</i> <i>max: 93,201</i>	\$38,409 <i>min: 8,888</i> <i>max: 82,244</i>	\$34,949 <i>min: 9,847</i> <i>max: 64,315</i>

*Sources:* Sales and winner data provided to the authors by the Texas Lottery Commission under an Open Records agreement. Counts of retailers, zip codes and cities in the pooled column refer to the total number ever in the sample. These numbers are larger than the individual year counts because of entry and exit.



**Table 2: The Initial Effect of a Winning Ticket on Retailer-Game Sales**

Coefficient on “a jackpot winner on this game at this retailer” dummy			
Dependent variable: $\ln(\text{weekly retailer sales on game } j)$			
	LOTTO TEXAS	TEXAS TWO-STEP	TEXAS CASH 5
<i>5 week lead</i>	-.058 (.041)	-.007 (.071)	-.015 (.020)
<i>4 week lead</i>	-.080 (.041)	.008 (.070)	-.007 (.020)
<i>3 week lead</i>	-.047 (.040)	.013 (.069)	-.017 (.019)
<i>2 week lead</i>	-.041 (.039)	-.024 (.066)	-.026 (.019)
<i>1 week lead</i>	-.052 (.037)	.024 (.064)	-.010 (.018)
<i>1 week lag</i>	.320 (.037)	.195 (.063)	.117 (.018)
<i>2 week lag</i>	.302 (.038)	.230 (.065)	.116 (.019)
<i>3 week lag</i>	.315 (.039)	.268 (.067)	.109 (.019)
<i>4 week lag</i>	.244 (.040)	.309 (.068)	.103 (.020)
<i>5 week lag</i>	.239 (.040)	.330 (.070)	.106 (.020)

*Notes:* Each cell corresponds to a unique regression estimating  $\ln(\text{sales})_{jit}$  as a function of  $\ln(\text{sales})_{ji(t-k)}$  and  $\text{win}_{ji(t-k)}$ , where  $j$  indexes lotto games,  $i$  indexes retailers,  $t$  indexes weeks, and  $k$  indexes the number of weeks lead or lag. Each row reports the results for a different  $k$  and each column for a different  $j$ . All regressions control for week fixed effects. The initial sample consists of 2,031,395 Lotto Texas observations, 887,632 Texas Two-Step observations, and 2,020,253 Cash 5 observations.

**Table 3: The Initial Effect of a Winning Ticket on Retailer-Game Sales by Jackpot Size**  
**Coefficients on “a jackpot winner on this game at this retailer” indicator and the**  
**interaction with prize amount less minimum prize amount**  
**Dependent variable: ln(weekly retailer sales on game  $j$ )**

	<u>LOTTO TEXAS</u>		<u>TX 2 STEP</u>		<u>CASH 5</u>	
	<i>Winner sold</i>	<i>Winner* (Prize – Min Prize)</i>	<i>Winner sold</i>	<i>Winner* (Prize – Min Prize)</i>	<i>Winner sold</i>	<i>Winner* (Prize – Min Prize)</i>
<i>5 week lead</i>	.023 (.058)	-.090 (.045)	.023 (.099)	-.007 (.017)	-.033 (.037)	.004 (.009)
<i>4 week lead</i>	.006 (.057)	-.095 (.045)	.088 (.099)	-.019 (.017)	-.036 (.036)	.009 (.009)
<i>3 week lead</i>	.025 (.054)	-.085 (.043)	.018 (.098)	-.001 (.016)	-.043 (.036)	.008 (.009)
<i>2 week lead</i>	-.024 (.052)	-.021 (.043)	-.122 (.094)	.023 (.016)	-.027 (.035)	.000 (.009)
<i>1 week lead</i>	-.012 (.050)	-.049 (.041)	.030 (.091)	-.002 (.016)	-.038 (.034)	.009 (.009)
<i>1 week lag</i>	.244 (.050)	.092 (.041)	.112 (.087)	.022 (.015)	.113 (.034)	.001 (.009)
<i>2 week lag</i>	.224 (.052)	.095 (.043)	.208 (.089)	.006 (.016)	.090 (.035)	.008 (.009)
<i>3 week lag</i>	.248 (.053)	.082 (.044)	.208 (.091)	.015 (.016)	.098 (.036)	.003 (.009)
<i>4 week lag</i>	.214 (.054)	.036 (.044)	.267 (.093)	.011 (.017)	.093 (.036)	.003 (.009)
<i>5 week lag</i>	.193 (.054)	.056 (.044)	.313 (.096)	.005 (.017)	.092 (.037)	.004 (.009)

Notes: Each row reports the results of three separate regressions, one for each game, estimating  $\ln(\text{sales})_{jit}$  as a function of  $\ln(\text{sales})_{ji(t-k)}$ ,  $\text{win}_{ji(t-k)}$ , and  $\text{win}_{ji(t-k)} * (\text{prize}_{ji(t-k)} - \text{minprize}_j)$ , where  $j$  indexes lotto games,  $i$  indexes retailers,  $t$  indexes weeks, and  $x$  indexes weeks lead or lag. Each row reports the results for a different  $k$  and each panel for a different  $j$ . All regressions control for week fixed effects. The prize amounts are scaled by 10 million for Lotto Texas; 100,000 for Texas Two-Step; and 10,000 for Cash 5. Minimum prize amount for Lotto Texas was \$1,032,666; \$200,000 for Texas Two-Step; and \$8,888 for Cash 5.

**Table 4: The Initial Effect of a Winning Lotto Texas Ticket on Retailer-Game Sales by Zip Code Demographics**

<b>Coefficients on “a Lotto Texas jackpot winner at this retailer” dummy and Interaction</b>										
<b>Dependent variable: ln(weekly retailer TX Lotto sales)</b>										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
<i>1 week lag</i>	.083 (.071)	.194 (.076)	.108 (.093)	.094 (.070)	.127 (.060)	.318 (.037)	.318 (.037)	.319 (.037)	.487 (.066)	.886 (.151)
<i>_*(% adults HS dropout)</i>	1.014 (.260)	-	-	-	-	.191 (.273)	-	-	-	-
<i>_*(% pop nonwhite)</i>	-	.499 (.269)	-	-	-	-.069 (.002)	-.036 (.001)	-	-	-
<i>_*(% adults over 65)</i>	-	-	1.47 (.598)	-	-	-	-	-	-	-
<i>_*(% pop in poverty)</i>	-	-	-	1.632 (.428)	-	-	.255 (.433)	.144 (.915)	-	-
<i>_*(% hholds on pub assistance)</i>	-	-	-	-	6.02 (1.48)	-	-	.257 (3.66)	-	-
<i>_*(% adults college plus)</i>	-	-	-	-	-	-	-	-	-.745 (.239)	-
<i>_*(% pop &gt; 200% poverty)</i>	-	-	-	-	-	-	-	-	-	-.861 (.222)
<i>Adj R<sup>2</sup></i>	.890	.890	.890	.890	.890	.890	.890	.890	.890	.890

*Notes:* Each column reports coefficients from a separate regression estimating  $\ln(\text{sales})_{it}$  as a function of  $\ln(\text{sales})_{i(t-k)}$ ,  $\text{win}_{i(t-k)}$ ,  $\text{win}_{i(t-k)} * (\text{percent}X)_z$ , and  $\text{percent}X_z$ , where  $i$  indexes retailers,  $t$  indexes weeks, and  $X$  refers to the particular demographic characteristic. All regressions control for week fixed effects. Population demographics at the five-digit zip code level are obtained from the 2000 U.S. census. The mean value of the percent of adults in a zipcode (weighted by the number of retailers in a zip code) with less than a high school degree is 26.2; population nonwhite is 21.6; adults over age 65 is 18.0; population in poverty is 15.1; and the households on public assistance is 3.1. The mean percent of adults in a zip code with a college degree or greater is 20.4 and the mean percent of the population above 200 percent of the poverty line is 16.1.

**Table 5: The Initial Effect of a Winning Ticket on Retailer Sales of Other Games**

<b>Coefficient on “a jackpot winner at this retailer on game <math>j</math>” dummy</b>			
<b>Dependent variable: <math>\ln(\text{weekly total retailer sales} \text{ minus sales on game } j)</math></b>			
	<b>Lotto Texas</b>	<b>TX 2 STEP</b>	<b>Cash 5</b>
<i>5 week lead</i>	.033 (.093)	-.058 (.076)	-.017 (.022)
<i>4 week lead</i>	-.087 (.092)	.031 (.075)	-.015 (.022)
<i>3 week lead</i>	.006 (.090)	.025 (.075)	.006 (.022)
<i>2 week lead</i>	-.137 (.089)	.033 (.074)	-.029 (.022)
<i>1 week lead</i>	-.004 (.091)	.040 (.075)	-.011 (.022)
<i>1 week lag</i>	.178 (.089)	.038 (.073)	.034 (.022)
<i>2 week lag</i>	.085 (.087)	.036 (.072)	.020 (.022)
<i>3 week lag</i>	.144 (.087)	.036 (.072)	.036 (.022)
<i>4 week lag</i>	.074 (.088)	-.019 (.073)	.034 (.022)
<i>5 week lag</i>	.116 (.088)	.058 (.074)	.013 (.022)

*Notes:* Each cell corresponds to a unique regression of  $\ln(\text{sales})_{i(-j)t}$  as a function of  $\ln(\text{sales})_{ji(t-k)}$ ,  $\ln(\text{sales})_{i(-j)(t-k)}$ , and  $\text{win}_{ij(t-k)}$ , where  $j$  indexes lotto games,  $i$  indexes retailers,  $t$  indexes weeks, and  $x$  indexes the number of weeks lead or lag. Each row reports the results for a different  $k$  and each column for a different  $j$ . where  $j$  indexes lotto games,  $-j$  refers to all games other than  $j$ ,  $i$  indexes retailers, and  $t$  indexes weeks. All regressions control for week fixed effects. The initial sample consists of 2,026,059 Lotto Texas observations, 886,572 Texas Two-Step observations, and 2,018,527 Cash 5 observations.

**Table 6: The Initial Effect of a Winning Ticket on Zip-code-Game Sales**

Coefficient on “jackpot winner on game $j$ in zip code” dummy			
Dependent variable: $\ln(\text{weekly zip code sales on game } j)$			
	<b>Lotto Texas</b>	<b>TX 2 STEP</b>	<b>Cash 5</b>
<i>5 week lead</i>	-.014 (.022)	.067 (.027)	.041 (.011)
<i>4 week lead</i>	-.023 (.024)	.064 (.023)	.038 (.010)
<i>3 week lead</i>	-.003 (.022)	.074 (.022)	.038 (.010)
<i>2 week lead</i>	.004 (.016)	.027 (.029)	.027 (.010)
<i>1 week lead</i>	-.015 (.019)	.053 (.022)	.036 (.009)
<i>1 week lag</i>	.132 (.019)	.160 (.024)	.083 (.006)
<i>2 week lag</i>	.145 (.018)	.183 (.022)	.082 (.006)
<i>3 week lag</i>	.160 (.018)	.178 (.025)	.092 (.007)
<i>4 week lag</i>	.139 (.018)	.193 (.028)	.092 (.007)
<i>5 weeks lag</i>	.106 (.017)	.201 (.031)	.089 (.008)

*Notes:* Each cell corresponds to a unique regression estimating  $\ln(\text{sales})_{jzt}$  as a function of  $\ln(\text{sales})_{jz(t-k)}$  and  $\text{win}_{zj(t-k)}$ , where  $j$  indexes lotto games,  $z$  indexes zip codes,  $t$  indexes weeks, and  $k$  indexes weeks lead or lag. Each row reports the results for a different  $k$  and each panel for a different  $j$ . All regressions control for week fixed effects. Standard errors are robust standard errors. The initial sample consists of 407,912 Lotto Texas observations, 175,590 Texas Two-Step observations, and 406,918 Cash 5 observations.

**Table 7: The Initial Effect of a Winning Ticket within Zip code on Own-Store Game Sales**  
**Coefficients on “any jackpot winner at this retailer on game  $j$ ” and “any jackpot winner in this zipcode on game  $j$ ”**  
**Dependent variable:  $\ln(\text{retailer sales on game } j)$**

	<b>Lotto Texas</b>		<b>TX 2 STEP</b>		<b>Cash 5</b>	
	winner at store	winner in zipcode	winner at store	winner in zipcode	winner at store	winner in zipcode
<b>5 weeks lead</b>	-.032 (.048)	.011 (.011)	-.074 (.085)	.027 (.019)	-.017 (.022)	.004 (.005)
<b>4 weeks lead</b>	-.059 (.046)	.024 (.010)	.012 (.082)	.028 (.019)	-.017 (.022)	.004 (.005)
<b>3 weeks lead</b>	-.021 (.044)	.037 (.010)	-.015 (.077)	.003 (.019)	-.009 (.021)	.006 (.005)
<b>2 weeks lead</b>	-.029 (.044)	.005 (.010)	.013 (.073)	.020 (.018)	-.015 (.020)	-.002 (.005)
<b>1 week lead</b>	-.020 (.040)	-.007 (.010)	-.002 (.070)	-.008 (.018)	-.007 (.019)	-.002 (.005)
<b>1 week lag</b>	.266 (.041)	.055 (.010)	.152 (.069)	.022 (.018)	.116 (.020)	-.001 (.005)
<b>2 week lag</b>	.064 (.042)	.064 (.010)	.166 (.072)	.041 (.019)	.115 (.020)	.002 (.005)
<b>3 week lag</b>	.234 (.043)	.085 (.011)	.251 (.073)	.028 (.019)	.100 (.021)	.007 (.005)
<b>4 week lag</b>	.120 (.044)	.075 (.011)	.263 (.075)	.035 (.019)	.096 (.021)	.008 (.005)
<b>5 weeks lag</b>	.194 (.044)	.063 (.011)	.285 (.077)	.041 (.019)	.107 (.021)	.004 (.005)

*Notes:* Results are from regressing  $\ln(\text{sales})_{ijt}$  as a function of  $\ln(\text{sales})_{ij(t-x)}$ ,  $\ln(\text{sales})_{zj(t-x)}$ ,  $\text{win}_{ij(t-x)}$ , and  $\text{win}_{zj(t-x)}$  where  $j$  indexes lotto games,  $i$  indexes retailers,  $z$  indexes 9-digit zip code, and  $t$  indexes weeks. Regressions control for week fixed effects. The sample is limited to stores located in a zip code with at least one other retailer. The initial sample for Texas Lotto has 1,812,093 observations; for Texas Two-Step 792,211 observations; and for Cash 5, 1,801,852 observations.

**Table 8: The Initial Effect of a Winning Ticket within 1-Mile on Own-Store Game Sales**

Coefficients on “any jackpot winner at this retailer on game  $j$ ” and “any jackpot winner in this zipcode on game  $j$ ”  
 Dependent variable:  $\ln(\text{retailer sales on game } j)$

	Lotto Texas		TX 2 STEP		Cash 5	
	winner at store	winner within 1-mile	winner at store	winner within 1-mile	winner at store	winner within 1-mile
<i>5 weeks lead</i>	-.113 (.076)	.004 (.085)	-.056 (.092)	-.263 (.143)	-.009 (.031)	.033 (.034)
<i>4 weeks lead</i>	-.119 (.075)	-.109 (.084)	-.030 (.090)	-.046 (.142)	-.011 (.031)	.034 (.034)
<i>3 weeks lead</i>	-.090 (.074)	-.140 (.084)	.005 (.089)	-.197 (.141)	-.023 (.030)	.040 (.033)
<i>2 weeks lead</i>	-.068 (.071)	.036 (.078)	.015 (.085)	-.106 (.138)	.003 (.030)	.005 (.032)
<i>1 week lead</i>	-.113 (.067)	-.175 (.077)	.036 (.083)	-.165 (.136)	-.005 (.030)	.002 (.031)
<i>1 week lag</i>	.333 (.068)	.041 (.080)	.141 (.082)	-.028 (.130)	.173 (.029)	.020 (.031)
<i>2 week lag</i>	.348 (.070)	-.035 (.080)	.185 (.085)	-.024 (.133)	.162 (.030)	.048 (.032)
<i>3 week lag</i>	.332 (.071)	.049 (.084)	.261 (.087)	-.153 (.136)	.132 (.031)	.003 (.033)
<i>4 week lag</i>	.206 (.072)	.052 (.085)	.272 (.089)	.072 (.137)	.101 (.031)	.033 (.033)
<i>5 weeks lag</i>	.223 (.072)	.146 (.086)	.284 (.089)	-.202 (.138)	.130 (.032)	-.005 (.034)

Notes: Results are from regressing  $\ln(\text{sales})_{ijt}$  as a function of  $\ln(\text{sales})_{ij(t-x)}$ ,  $\ln(\text{sales})_{gj(t-x)}$ ,  $\text{win}_{ij(t-x)}$ , and  $\text{win}_{gj(t-x)}$  where  $j$  indexes lotto games,  $i$  indexes retailers,  $g$  indexes “1 mile group”, and  $t$  indexes weeks. Regressions control for week fixed effects. Standard errors are robust standard errors. The sample is limited to stores that have at least one other retailer within 1-mile.

**Table 9: The Persistence of the Effect of a Winning Ticket on Retailer-Game Sales**

<b>Coefficient on “a jackpot winner on this game at this retailer” dummy</b>			
<b>Dependent variable: ln(monthly retailer sales on game <math>j</math>)</b>			
	<b>Lotto Texas</b>	<b>Texas Two-Step</b>	<b>Cash 5</b>
<i>5 month lead</i>	-.101 (.078)	.001 (.063)	-.002 (.017)
<i>4 month lead</i>	-.082 (.049)	-.018 (.057)	.006 (.014)
<i>3 month lead</i>	-.086 (.056)	.011 (.045)	-.019 (.016)
<i>2 month lead</i>	-.048 (.038)	-.120 (.060)	-.028 (.014)
<i>1 month lead</i>	-.070 (.034)	-.056 (.040)	-.044 (.015)
<i>1 month lag</i>	.204 (.049)	.207 (.035)	.085 (.013)
<i>2 month lag</i>	.112 (.054)	.157 (.038)	.090 (.013)
<i>3 month lag</i>	.106 (.031)	.155 (.040)	.075 (.017)
<i>4 month lag</i>	.119 (.032)	.147 (.040)	.067 (.018)
<i>5 month lag</i>	.025 (.055)	.199 (.044)	.072 (.016)
<i>6 month lag</i>	.097 (.036)	.212 (.050)	.066 (.016)
<i>7 month lag</i>	.102 (.034)	.123 (.054)	.048 (.019)
<i>8 month lag</i>	.096 (.035)	.096 (.049)	.061 (.017)
<i>9 month lag</i>	.067 (.039)	.181 (.057)	.074 (.017)
<i>10 month lag</i>	.077 (.044)	.140 (.062)	.053 (.023)

*Notes:* Each cell corresponds to a unique regression estimating  $\ln(\text{sales})_{jit}$  as a function of  $\ln(\text{sales})_{ji(t-k)}$  and  $\text{win}_{ji(t-k)}$ , where  $j$  indexes lotto games,  $i$  indexes retailers,  $t$  indexes month, and  $k$  indexes months lead or lag. Each row reports the results for a different  $k$  and each column for a different  $j$ . All regressions control for year-month fixed effects. Standard errors are robust standard errors. The initial sample consists of 464,875 Lotto Texas observations, 211,337 Texas Two-Step observations, and 464,086 Cash 5 observations.



**Table 10: The Persistence of the Effect of a Winning Ticket on Zip-Code-Game Sales**

Coefficient on “jackpot winner on game $j$ in zip code” dummy			
Dependent variable: $\ln(\text{monthly zip code sales on game } j)$			
	<b>Lotto Texas</b>	<b>TX 2 STEP</b>	<b>Cash 5</b>
<i>5 month lead</i>	-.018 (.025)	-.024 (.032)	.040 (.009)
<i>4 month lead</i>	-.034 (.027)	-.031 (.032)	.038 (.008)
<i>3 month lead</i>	-.030 (.025)	-.008 (.023)	.035 (.007)
<i>2 month lead</i>	-.010 (.021)	-.016 (.020)	.024 (.007)
<i>1 month lead</i>	-.021 (.019)	-.015 (.017)	-.004 (.011)
<i>1 month lag</i>	.110 (.015)	.111 (.018)	.053 (.005)
<i>2 month lag</i>	.069 (.024)	.092 (.020)	.054 (.006)
<i>3 month lag</i>	.054 (.016)	.089 (.017)	.060 (.007)
<i>4 month lag</i>	.071 (.019)	.107 (.018)	.065 (.007)
<i>5 month lag</i>	.046 (.019)	.092 (.022)	.061 (.006)
<i>6 month lag</i>	.054 (.021)	.082 (.022)	.064 (.009)
<i>7 month lag</i>	.038 (.019)	.063 (.030)	.062 (.008)
<i>8 month lag</i>	.045 (.021)	.055 (.027)	.068 (.007)
<i>9 month lag</i>	.032 (.024)	.084 (.031)	.072 (.008)
<i>10 month lag</i>	.045 (.027)	.065 (.027)	.071 (.009)

*Notes:* Each cell corresponds to a unique regression estimating  $\ln(\text{sales})_{jzt}$  as a function of  $\ln(\text{sales})_{jz(t-k)}$  and  $\text{win}_{jz(t-k)}$ , where  $j$  indexes lotto games,  $z$  indexes zip codes,  $t$  indexes months, and  $k$  indexes months lead or lag. Each row reports the results for a different  $k$  and each column for a different  $j$ . All regressions control for year-month fixed effects. Standard errors are robust standard errors. The initial sample consists of 93,206 Lotto Texas observations, 41,680 Texas Two-Step observations, and 93,149 Cash 5 observations;

**Table 11: The Effect of a Winning Ticket within Zip Code on Retailer Game Sales**

Coefficient on “any jackpot winner at this retailer on game $j$ ” and “any jackpot winner in this zipcode on game $j$ ” Dependent variable: $\ln(\text{retailer sales on game } j)$						
	TEXAS LOTTO		TEXAS TWO-STEP		TEXAS CASH 5	
	winner at store	winner in zipcode	winner at store	winner in zipcode	winner at store	winner in zipcode
<i>5 month lead</i>	-.048 (.079)	.006 (.010)	-.133 (.083)	.038 (.026)	.002 (.022)	-.002 (.005)
<i>4 month lead</i>	-.027 (.057)	.004 (.009)	-.003 (.068)	.007 (.019)	-.006 (.018)	-.002 (.005)
<i>3 month lead</i>	-.050 (.073)	.004 (.009)	.010 (.066)	.033 (.018)	-.003 (.016)	-.002 (.004)
<i>2 month lead</i>	-.004 (.043)	.017 (.010)	-.102 (.065)	.024 (.015)	-.036 (.015)	-.003 (.004)
<i>1 month lead</i>	.017 (.010)	.027 (.011)	-.067 (.047)	.010 (.013)	-.035 (.014)	-.00005 (.004)
<i>1 month lag</i>	.159 (.038)	.066 (.010)	.176 (.050)	.022 (.014)	.076 (.013)	.003 (.004)
<i>2 month lag</i>	.123 (.047)	.046 (.013)	.161 (.047)	.011 (.015)	.078 (.014)	.003 (.004)
<i>3 month lag</i>	.102 (.032)	.031 (.010)	.116 (.051)	.014 (.016)	.060 (.017)	.007 (.004)
<i>4 month lag</i>	.123 (.040)	.043 (.011)	.172 (.059)	-.011 (.018)	.059 (.018)	.005 (.004)
<i>5 month lag</i>	.067 (.046)	.027 (.010)	.225 (.066)	-.015 (.019)	.056 (.018)	.009 (.005)
<i>6 month lag</i>	.104 (.042)	.022 (.010)	.250 (.070)	-.043 (.022)	.055 (.017)	.009 (.005)
<i>7 month lag</i>	.146 (.035)	.030 (.011)	.126 (.061)	-.044 (.022)	.043 (.019)	.006 (.005)
<i>8 month lag</i>	.123 (.038)	.042 (.011)	.059 (.059)	.006 (.023)	.040 (.020)	.013 (.005)
<i>9 month lag</i>	.112 (.046)	.012 (.012)	.159 (.071)	.023 (.025)	.056 (.021)	.010 (.006)
<i>10 month lag</i>	.122 (.049)	.012 (.014)	.137 (.083)	-.019 (.029)	.048 (.023)	.006 (.006)

*Notes:* Results are from regressing  $\ln(\text{sales})_{ij,t}$  as a function of  $\ln(\text{sales})_{ij,t-x}$ ,  $\ln(\text{sales})_{zj,t-x}$ ,  $\text{win}_{ij,t-x}$ , and  $\text{win}_{zj,t-x}$  where  $j$  indexes lotto games,  $i$  indexes retailers,  $z$  indexes 9-digit zip code, and  $t$  indexes weeks. Regressions control for week effects. The sample is limited to stores located in a zip code with at least one other retailer. Standard errors are robust standard errors. The initial sample has 417,785 observations for Texas Lotto; 190,357 observations for Texas Two-Step; and 417,048 for Cash 5.

**Table 12: The Effect of a Winning Ticket within 1-mile on Game Sales**

Coefficients on “any jackpot winner at this retailer on game  $j$ ” and “any jackpot winner in this group on game  $j$ ”  
 Dependent variable:  $\ln(\text{retailer sales on game } j)$

	Lotto Texas		TX 2 STEP		Cash 5	
	winner at store	winner in group	winner at store	winner in group	winner at store	winner in group
<i>5 month lead</i>	-.033 (.048)	.021 (.082)	.026 (.078)	-.222 (.199)	-.012 (.025)	-.011 (.029)
<i>4 month lead</i>	-.086 (.046)	-.041 (.069)	.003 (.072)	-.030 (.207)	-.001 (.020)	-.018 (.031)
<i>3 month lead</i>	-.068 (.049)	-.128 (.086)	.043 (.055)	-.092 (.132)	-.040 (.024)	-.001 (.028)
<i>2 month lead</i>	-.054 (.051)	-.105 (.077)	-.121 (.080)	-.209 (.245)	-.045 (.020)	.026 (.021)
<i>1 month lead</i>	-.081 (.047)	-.067 (.041)	-.046 (.046)	-.121 (.079)	-.043 (.019)	.029 (.023)
<i>1 month lag</i>	.215 (.043)	.072 (.043)	.192 (.037)	.038 (.062)	.078 (.017)	.022 (.027)
<i>2 month lag</i>	.043 (.115)	-.001 (.064)	.139 (.043)	.0003 (.086)	.081 (.016)	.009 (.029)
<i>3 month lag</i>	.065 (.059)	-.043 (.042)	.162 (.048)	-.008 (.149)	.053 (.026)	-.005 (.034)
<i>4 month lag</i>	.127 (.051)	-.030 (.045)	.149 (.053)	.251 (.047)	.031 (.028)	.007 (.036)
<i>5 month lag</i>	.082 (.043)	-.038 (.061)	.230 (.061)	.176 (.066)	.060 (.021)	.002 (.034)
<i>6 month lag</i>	.104 (.054)	-.110 (.073)	.236 (.060)	.306 (.062)	.063 (.021)	.043 (.032)
<i>7 month lag</i>	.079 (.054)	-.198 (.014)	.100 (.077)	.234 (.064)	.004 (.032)	.016 (.043)
<i>8 month lag</i>	.061 (.046)	-.200 (.019)	.115 (.065)	.330 (.060)	.024 (.027)	.005 (.050)
<i>9 month lag</i>	.036 (.058)	-.051 (.075)	.150 (.083)	.483 (.066)	.035 (.026)	.040 (.037)
<i>10 month lag</i>	.033 (.065)	-.138 (.015)	.176 (.096)	-.100 (.029)	.039 (.027)	.032 (.039)

Notes: Results are from regressing  $\ln(\text{sales})_{ijt}$  as a function of  $\ln(\text{sales})_{ij(t-x)}$ ,  $\ln(\text{sales})_{gi(t-x)}$ ,  $\text{win}_{ij(t-x)}$ , and  $\text{win}_{gi(t-x)}$  where  $j$  indexes lotto games,  $i$  indexes retailers,  $g$  indexes “1 mile group”, and  $t$  indexes weeks. Regressions control for week effects. Standard errors are robust standard errors. The sample is limited to stores that have at least one other retailer within 1-mile. The initial sample has 245,300 observations for Texas Lotto, 122,592 observations for Texas Two-Step, and 245,743 observations for Cash 5.

## Appendix 1: Model of Addiction

Our empirical specification follows from a simple model of lifetime consumption. We extend the basic model presented in the previous section. Consider a consumer whose current-period utility depends on current consumption of a numeraire good  $Y$ , current plus  $T$  lags of a potentially addictive good  $g$ , and unmeasured life-cycle variables, which are captured by  $e_t$ . We can write the discounted value of lifetime utility

$$U = \sum_{t=0}^{\infty} \beta^t u(y_t, g_t, g_{t-1}, g_{t-2}, \dots, g_{t-T}, e_t)$$

where  $\beta$  is the discount factor  $1/(1 + \rho)$ ,  $\rho$  being the rate of time preference. Individuals maximize  $U$  subject to the lifetime budget constraint

$$\sum_{t=0}^{\infty} \beta^t (y_t + p_t g_t) = I$$

where  $p_t$  is the price of the potentially addictive good,  $I$  is the present value of lifetime income, and the interest rate is assumed to be the same as the rate of time preference. For simplicity, there is no labor supply in the model.

Following Becker, Grossman, and Murphy (1994), we consider a utility function that is quadratic in  $y_t, e_t, g_t$ , and  $T$  lags of  $g$ .<sup>24</sup>

$$\begin{aligned} U = & u_y y_t + \frac{1}{2} u_{yy} y_t^2 + u_e e_t + \frac{1}{2} u_{ee} e_t^2 + u_0 g_t + \frac{1}{2} u_{00} g_t^2 \\ & + u_1 g_{t-1} + \frac{1}{2} u_{11} g_{t-1}^2 + u_2 g_{t-2} + \frac{1}{2} u_{22} g_{t-2}^2 + \dots + u_T g_{t-T} + \frac{1}{2} u_{TT} g_{t-T}^2 \\ & + u_{ye} y_t e_t + u_{y0} y_t g_t + u_{y1} y_t g_{t-1} + u_{y2} y_t g_{t-2} + \dots + u_{yT} y_t g_{t-T} \\ & + u_{e0} e_t g_t + u_{e1} e_t g_{t-1} + u_{e2} e_t g_{t-2} + \dots + u_{eT} e_t g_{t-T} \\ & + \sum_{i=0}^T \sum_{j=i+1}^T u_{ij} g_{t-i} g_{t-j} \end{aligned}$$

The first-order conditions for  $y_t$  and  $g_t$  are, respectively

$$u_y + u_{yy} y_t + u_{ye} e_t + \sum_{s=0}^T u_{sy} g_{t-s} = \lambda \quad [\text{A1}]$$

and

<sup>24</sup> Becker, Grossman, and Murphy (1994) consider the utility function with only one lag of  $g$  (i.e. the special case where  $T = 1$ ). We simply extend this model to include more lags to be less restrictive. The economic conclusions of the model are unchanged.

$$\begin{aligned}
& u_0 + u_{00}g_t + u_{0y}y_t + u_{0e}e_t + \sum_{s=1}^T u_{0s}g_{t-s} \\
& + \sum_{m=1}^T \beta^m \left( u_m + u_{my}y_{t+m} + u_{me}e_{t+m} + \sum_{s=0}^T u_{ms}g_{t+m-s} \right) = \lambda p_t
\end{aligned} \tag{A2}$$

where the  $u_i$ 's and  $u_{ii}$ 's are, respectively, the coefficients on the first- and second-order elements of the utility function, the  $u_{ij}$ 's are the coefficient on the interaction terms,  $u_0$  is the coefficient on  $g_t$ ,  $u_1$  is the coefficient on  $g_{t-1}$ , and so on.

The first-order condition for  $y_t$  is the usual condition that the marginal utility of consumption is equal to the marginal cost (in utility terms) of consumption. Note, however, that the first-order condition for  $g_t$  contains an extra term for each lag of  $g$  that affects utility. Specifically, the consumer recognizes that changes in current consumption will enter the utility function as lagged consumption in the future in two ways. Lagged consumption has a direct effect on utility, as seen in the terms  $u_m$  and  $u_{mm}$  ( $m \in 1, 2, \dots, T$ ). Lagged consumption also affects the marginal utility of current consumption, as represented by the utility coefficients  $u_{01}, u_{02}, \dots, u_{0T}$ . The former coefficients determine the harmful or beneficial nature of the addiction, while the latter coefficients determine whether the good is addictive.

Solving [A1] for  $y_t$  and substituting into [A2] we get

$$\begin{aligned}
g_t &= \alpha + \theta_p p_t \\
&+ \theta_{-1}g_{t-1} + \theta_{-2}g_{t-2} + \dots + \theta_{-T}g_{t-T} + \theta_1g_{t+1} + \theta_2g_{t+2} + \dots + \theta_Tg_{t+T} \\
&+ \theta_{e_0}e_t + \theta_{e_1}e_{t+1} + \dots + \theta_{e_T}e_{t+T}
\end{aligned} \tag{A3}$$

where

$$\begin{aligned}
\theta_{-i} &= \frac{-\sum_{s=0}^{T-i} \beta^s (u_{s,s+i}u_{yy} - u_{sy}u_{s+i,y})}{\pi}; \quad \theta_i = \frac{\beta^i \theta_{-i}}{\pi}; \quad \theta_{e_i} = \frac{-\beta^i (u_{ie}u_{yy} - u_{iy}u_{ye})}{\pi}; \quad \theta_p = \frac{\lambda u_{yy}}{\pi}; \\
\alpha &= \sum_{s=0}^T \beta^s \left( u_s \frac{u_{sy}}{u_{yy}} (\lambda - u_y) \right); \quad \text{and} \quad \pi = \sum_{s=0}^T \beta^s (u_{ss}u_{yy} - u_{sy}^2).
\end{aligned}$$

Equation [A3] is the basis for our empirical specification.

## **Appendix 2: The Lotto Games**

### **Lotto Texas**

Lotto Texas offers multi-million dollar jackpots, which winners can choose to receive either as 25 annual payments or as one (present-discounted) cash payment. From its inception in 1992 until it was changed in mid-July 2000, Lotto Texas was played by choosing six numbers out of a field of 50, yielding odds of 15,890,700 to one of matching all six numbers. The field was later expanded to 54, yielding odds of 25,827,165 to one. The prize pool for each Lotto Texas drawing is comprised of 55 percent of sales for that drawing. Of this amount, 68 percent is allocated to the jackpot prize, plus any amount carried over from previous drawings. If no ticket bet matches the winning six numbers in a given week, the amount allocated for a top-prize winner is rolled over to the next draw. Portions of the prize pool are reserved each week to pay pari-mutuel prizes for five-of-six and four-of-six winners. A fixed prize of five dollars is paid to players who match three of the six numbers. Lotto Texas drawings occur twice a week. Players can purchase bets up to 10 drawings in advance, paying one dollar per drawings.

### **Texas Two Step**

Texas Two Step offers jackpots that start at \$200,000 and accumulate until someone wins the top prize, which is paid in a single lump-sum payment. The jackpot is won by correctly choosing four numbers from 35 and one Bonus Ball number from one to 35; the probability of doing so is 1 in 1,832,600. Similarly to Lotto Texas, the game includes lower prize tiers. The total prize pool comprises 50 percent of sales. Texas Two Step drawings occur twice per week. Players can play their numbers for up to ten drawings by marking their play slip accordingly; each play costs \$1.

### **Texas Cash Five**

Cash Five pays a pari-mutuel lump-sum top prize to players who correctly pick five numbers out of 39, with odds of 575,757 to one. (The playing field was changed to 37 numbers after our sample period.) If no one correctly matches all five numbers, the money set aside to pay the top prize rolls down to the four-of-five prize. A pari-mutuel prize is also awarded for matching three-of-five numbers and a fixed prize of two dollars is paid for matching two-of-five. The Cash Five prize pool equals 50 percent of sales. Drawings are held six evenings a week. Players can play their Cash Five numbers for up to 12 drawings by marking their play slip accordingly. They can also choose their numbers in advance and indicate what day they want to start playing.

### **Appendix 3: Survey of Retailers that Sold a Winning Lotto Texas Ticket**

#### *Methodology*

We conducted a five-minute telephone survey of the retail outlets that sold winning Lotto Texas tickets during the time period we observe. The survey asked various questions about the store's advertising practices before and after the sale of a winning ticket. We attempted to contact the full sample of 66 stores that sold the 67 winning Lotto Texas tickets in our data. The surveyor initially asked to speak to the manager at each store. In the event that the manager was not available, the surveyor offered to call back at a time when the manager might be available. In the event that the manager had not been working at the store at the time of the winning ticket sale, the surveyor asked to speak to a store employee who had been there at the time. Of the 66 winning stores, the contact information for 8 was not current, 14 refused to take the survey, and 14 were unable to complete the survey because none of the current employees had worked at the store at the time of the winning ticket sale. Of the remaining 30 stores, 17 completed the phone survey. The surveyor recorded responses directly onto the paper copy of the survey. The survey was mailed to the remaining 13 stores with an addressed, stamped return envelope; three completed surveys were returned.

#### *Results*

Of the 17 winning stores that completed the survey -

- 14 report having received some type of media attention for having sold a winning lottery ticket, with local newspapers and television stations being the most common.
- 14 report hanging the "Winning Ticket Sold Here" sign provided by the Texas Lottery Commission (TLC), with all 14 reporting that they displayed the sign immediately following the sale of the ticket or as soon as the sign arrived.
- 17 report displaying the point of sale materials provided by the TLC, such as game decals, game posters and display tickets prior to the sale of the winning ticket.
- 0 report changing their general advertising strategies after the sale of a winning ticket
- 15 report a perceived increase in lottery ticket sales following the winning ticket sale
- 9 report a perceived increase in sales of other products following the winning ticket sale