TILBURG ◆ ◆ UNIVERSITY

Tilburg University

Advertising as a Reminder

He, Chen; Klein, Tobias

Publication date:
2018

Document Version
Early version, also known as pre-print

Link to publication in Tilburg University Research Portal

Discussion paper

Advertising as a Reminder: Evidence from
the Dutch State Lottery

By Chen He, Tobias Klein

May, 2018

TILBURG • UNIVERSITY

Understanding Society

Advertising as a reminder: Evidence from the Dutch State Lottery[*]

Chen He and Tobias J. Klein[†]

May 2018

Abstract

We use high frequency data on TV and radio advertising together with data on online sales for lottery tickets to measure the short run effects of advertising. We find them to be strong and to last for up to about 4 hours. They are the bigger the less time there is until the draw. We develop the argument that this finding is consistent with the idea that advertisements remind consumers to buy a ticket and that consumers value this. Then, we point out that in terms of timing the interests of the firm and the consumers are aligned: consumers wish to be reminded in a way that makes them most likely to consider buying a lottery ticket. We present direct evidence that this does not only affect the timing of purchases, but leads to market expansion. Then, we develop a tractable dynamic structural model of consumer behavior, estimate the parameters of this model and simulate the effects of a number of counterfactual dynamic advertising strategies. We find that relative to the actual schedule it would be valued by the consumers and profitable for the firm to spread advertising less over time and move it to the last days before the draw.

Key words: Dynamic demand, limited attention, reminder advertising, adoption model.

JEL-classification: M37, D12, D83.

[†]He: CentER, Tilburg University; e-mail: C.He@uvt.nl. Klein: C.E.P.R., TILEC, CentER, Tilburg University; e-mail: T.J.Klein@uvt.nl.

1 Introduction

In 2016, global advertising spending amounted to 493 billion US dollars.[1] Yet, it remains a challenge to measure the effects of advertising and characterize the underlying mechanism through which it affects consumer behavior (Lewis and Rao, 2015).

The consensus that has emerged in the literature is that conceptually, firms pursue a combination of goals when advertising: they either aim at conveying information about the existence, characteristics and prices of products; or they wish to positively influence the inclination of consumers who already know about their products to buy them.[2] In this paper, we present novel empirical evidence that is in line with the view that in addition to these well-received ways of working, advertising can act as a reminder. The underlying idea is that consumers have limited attention and may therefore value being reminded.

Specifically, we use high frequency data on TV and radio advertising together with online sales data for lottery tickets to measure the short run effects of advertising. The high frequency nature of our data allows us to credibly identify advertising effects. The exact timing of advertisements is beyond the control of the firm and therefore, the thought experiment we can undertake is to compare sales just before the advertisement was aired to sales right after this. Our setup is well-suited to study reminder effects, because there are recurrent deadlines within a year, consumers are well-aware of the product and its characteristics, and there is no other closely competing product that is offered to them.

We find the short run effects of advertising to be sizable. They last up to about 4 hours and are the bigger the less time there is until the draw, consistent with our interpretation that advertisements indeed remind consumers to buy a ticket and that consumers value this. The underlying idea is that consumers enjoy the benefits of buying a ticket mainly at a later point in time, on the day of the draw, which is why they prefer to be reminded later and then react stronger to it. This argument is developed in Section 4.3.

An important related question is whether advertising only leads to purchase acceleration (individuals buying earlier rather than later) or also to market expansion (more people buying in total). In order to provide model-free evidence on this, we point out that if advertising has a short-run effect until the end of the period in which tickets can be bought, then it must be the case that it also leads to market expansion.[3] We find this to be the case.

After presenting this model-free evidence, we point out that in terms of timing the interest of the firm that is advertising and the consumers are aligned: consumers wish to be reminded in a way that makes them most likely to consider buying a lottery ticket. The tradeoff they face is that on the one hand, if the firm allocates all the advertising very late, then it may not reach certain consumers, for instance because they will not watch TV on these days; on the other hand, if it spreads advertising expenditures out over time in order to reach more consumers,

---

[1]Taken from a report by Letang and Stillman (2016).

[2]See Bagwell (2007) for an excellent survey on the economics of advertising.

[3]We would like to thank Martin Peitz for this suggestion.

then advertising effects may be smaller because consumers do not want to buy too early. This means that total sales crucially depend on the dynamic advertising strategy and therefore, it would be valuable to quantify the dependence of sales on counterfactual advertising strategies.

For this, we then develop a tractable structural model of consumer behavior that we estimate. Our counterfactual simulations suggest that relative to the actual schedule it would be profitable for the firm and also valued by the consumer if advertising would be shifted towards the last days before the draw.

Besides, the model is useful to formalize the idea that advertisements act as a reminder. We think of consumers, at a given point in time, as either buying a ticket for the following draw, or postponing the decision to do so to a later point in time, with the possibility that they either forget to buy a ticket or consciously decide not to do so. In each period, there is a considerations stage. The probability to consider buying a ticket, i.e. to compare the value from buying the ticket and the value of waiting, depends on an advertising goodwill stock that depreciates over time. When estimating the model we pay particular attention to the fact that at a given point in time goodwill stocks will be heterogeneous across *ex ante* identical consumers, because they depend on heterogeneous but unobserved viewing behavior.

Our paper relates to several strands of the literature. The overarching theme is that quantifying the effects of advertising and shedding light on the exact mechanism through which it affects consumer behavior remains challenging, even using large-scale field experiments (Lewis and Rao, 2015); but novel data sources and innovative empirical designs have allowed researchers to measure advertising effects in credible ways and also shed additional light on the underlying mechanism. Overall, given the importance of the advertising industry, there is relatively little empirical work on the topic in economics.

The idea that advertising may serve as a reminder has appeared in the context of a debate on the optimal number of times a consumer should be reached. Krugman (1972) argued that consumers need to first understand the nature of the stimulus, then evaluate the personal relevance, and finally are reminded to buy when they are in a position to do so. So, he concluded that they should be reached at least three times. The underlying way of thinking about consumers is related to models of limited attention that, according to Kahneman (1973) and others, may originate in limits of information processing power and therefore may lead to forgetting. We add to this by presenting novel empirical evidence that is in line with that view.

Ackerberg (2001, 2003) also focuses on the mechanism by which advertising influences consumer behavior. His aim is to empirically distinguish between advertisements being effective because they are informative *vis-à-vis* them being effective because they increase the valuation for the brand. He finds mainly support of former. We complement this by presenting evidence that is in line with the view that advertisements act as a reminder.

Lodish et al. (1995) study the effectiveness of TV advertising and document a combination of no and positive effects. Hu et al. (2007) find that the effects have increased in later years. We contribute to this literature by using high frequency data to show that advertising has significant

effects on online sales and that it leads to market expansion.

A number of recent studies shed light on the relationship between TV advertising and behavior online. Joo et al. (2015) find that there is a significant effect of TV advertising on consumers' tendency to search online. Lewis and Reiley (2013) study the effects of Super Bowl advertising on online search behavior. They find that large spikes in search behavior related to the advertiser or product within 15 seconds following the conclusion of the TV commercial. Stephens-Davidowitz et al. (2017) exploit a natural experiment and find that advertising has a positive effect on searches and on the demand for movie tickets on the opening weekend. Du et al. (2017) characterize how the effects of advertising on online searches depend on advertisement content, media-contextual factors, and brand. Both Liaukonyte et al. (2015) and our paper complement these papers with evidence from high-frequency advertising and sales (as opposed to search or low-frequency sales) data.

The model we estimate features an advertising goodwill stock as in the model by Dubé et al. (2005). They estimate a static model and focus on interesting dynamics on the supply side. Another difference is that in our model, advertising affects the probability of considering to buy a ticket at a given point in time and not the flow utility associated with buying. Sovinsky Goeree (2008) and Draganska and Klapper (2011) estimate static models with a consideration stage using aggregate level data.

Our model is dynamic and consumers decide at each point in time whether to buy a ticket or wait. Melnikov (2013) and De Groote and Verboven (2016) estimate similar models. Their respective models do however not feature a consideration stage in which advertising has an effect. So, our modeling contribution lies in proposing a model in which advertising can naturally be thought of as acting as a reminder because it affects the probability to consider buying through an advertising goodwill stock in a dynamic decision context.

The rest of this paper is structured as follows. Section 2 gives a brief overview over the market for lottery tickets in the Netherlands. Section 3 describes the data and provides descriptive statistics. Section 4 shows reduced-form evidence on the effect of advertising on sales. Section 5 develops our model of lottery ticket demand with advertising effects. Section 6 presents the results. Section 7 performs counterfactual experiments for the supply side, and Section 8 concludes by pointing towards other situations in which our model could be used, including public policy. The (intended) Online Appendix is attached at the very end. Appendix A provides details on the structural estimation procedure, Appendix B contains robustness checks for the structural analysis, and Appendix C contains additional tables and figures.

2   The market for lottery tickets in the Netherlands

The market for lottery tickets in the Netherlands is very concentrated, with three organizations conducting different types of lotteries. First, the Stichting Exploitatie Nederlandse Staatsloterij, from which we received the data, offers lottery tickets for The Dutch State Lottery (in Dutch:

Staatsloterij) and the Millions Game (Miljoenenspel). Staatsloterij has a history going back to the year 1726 and is run by the government. It is by far the biggest of its kind in the Netherlands. The second player is the Stichting Exploitatie Nederlandse Staatsloterij. It offers the Lotto Game (Lottospel), which is comparable but much smaller in size, next to other games such as Eurojackpot and Scratch Tickets (Krasloten) and sports betting. In 2016, these two organizations merged. The third player is Nationale Goede Doelen Loterijen offering a ZIP Code Lottery (Postcodeloterij), whose main purpose it is to donate money to charity. For that reason, it is not directly comparable to the other two lotteries.[4]

The lottery run by Staatsloterij is classical. A ticket has a combination of numbers and Arabic letters and a consumer can choose some of them. The size of the prize depends then on how many numbers and letters of a ticket match with the ones of the winning combination. On top of that, there is a jackpot whose size varies over time. For all draws but the very last one in a year, consumers can choose between a full ticket that costs 15 euros and multiples of one fifth of a ticket. For the last draw, the price of a ticket is 15 euros and consumers can buy multiples of one half of a ticket. Winning amounts are then scaled accordingly. The tickets can be purchased in two ways: they can either be purchased online via the official website of Staatsloterij, or offline, for example, in a supermarket or a gas station. Most of the sales are offline, but nevertheless the online business is considered important.

There are 16 draws in a calendar year. 12 of them are regular draws and 4 of them are special draws. Regular draws take place on the 10th of every month. The dates of 4 additional special draws vary slightly from year to year. In 2014 (the year for which we have data), the 4 special draws were on April 26 (King's day in the Netherlands), on June 24, October 1 and on December 31 (the new year's eve draw). All draws but the last in a year take place at 8pm (Central European Time). From 6pm onward, no more tickets can be bought for that draw.

3 Data and descriptive statistics

3.1 Overview

Our data are for 2014 and consist of 3 parts: online transactions, TV and radio advertising, and jackpot sizes. The transaction data are collected at the minute level. We observe the number of lottery tickets sold online.[5] The advertising data consists of minute-level measurements of gross

---

[4]In 2014, Staatsloterij had a turnover of 738 million euros with 579 million euros related to its lottery and De Lotto of 322 million euros with 144 million euros related to its lottery (https://over.nederlandseloterij.nl/over-ons/publicaties, accessed May 2018). The turnover of Nationale Goede Doelen Loterijen was 847 million euros in total with 624 million euros related to its charity ZIP code lottery (https://view.publitas.com/nationale-postcode-loterij-nv/npl-jaarverslag-2014/page/58-59, accessed February 2016).

[5]These data have been collected using Google Analytics. In particular, visits to the "exit page" confirming payment have been recorded. This means that we do not observe what type of ticket a consumer has bought. Advertising also affects offline sales, and therefore, ideally, we would also like to observe the number of lottery tickets sold offline. However, offline transactions are not observed in the dataset. At the same time, it generally takes longer until an offline sale takes place after an individual listens to a radio advertisement or sees a TV

Figure 1: Cumulative sales for selected draws
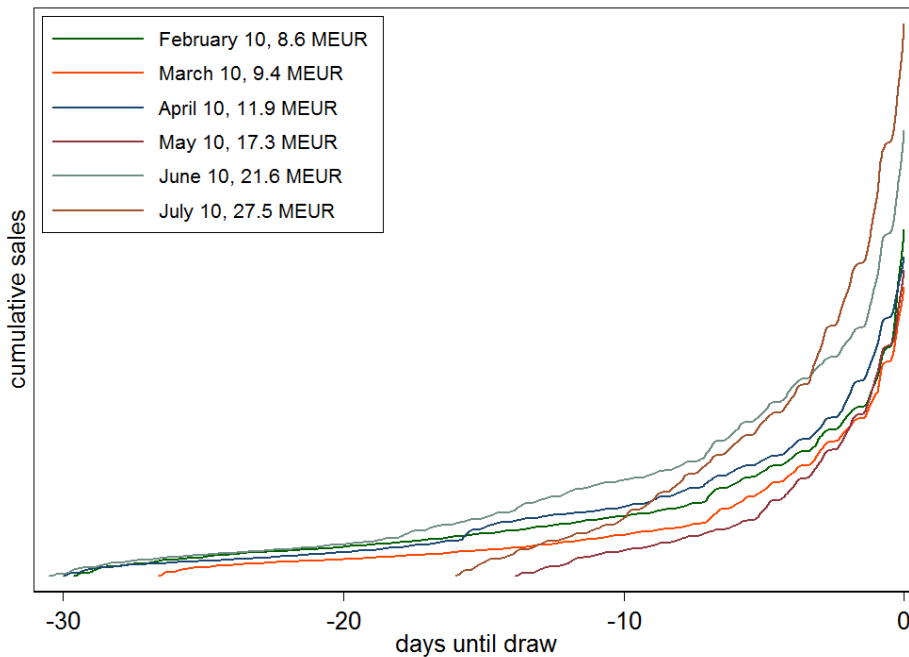
Notes: This figure shows cumulative sales for 6 selected regular draws. The respective jackpot amounts are given in the legend. See Figure 10 in the Online Appendix for the remaining draws.

rating points (GRP's), separately for TV and radio advertising. GRP's measure impressions as a percentage of the target population at a given point in time. For example, 5 GRP's in our data mean that in that minute 5 percent of the target population (in our case the general population) are exposed to an advertisement. This is a standard measure in the advertising industry.

Besides, we observe the jackpot size for the 12 regular draws in 2014. There is no jackpot size for the 4 special draws, as more involved rules apply to them. For example, on the drawing day, every 15 minutes consumers can win an additional 100,000 euros. In the empirical analysis, we will capture differences across draws in a flexible way.

We are not allowed to report levels of sales and advertising. Therefore, we will only present relative numbers and (semi-) elasticities in the tables and figures below and some vertical axis will have no units of measurements. Of course, we will still use these data when conducting the analysis.

Figure 2: Advertising and sales during the day
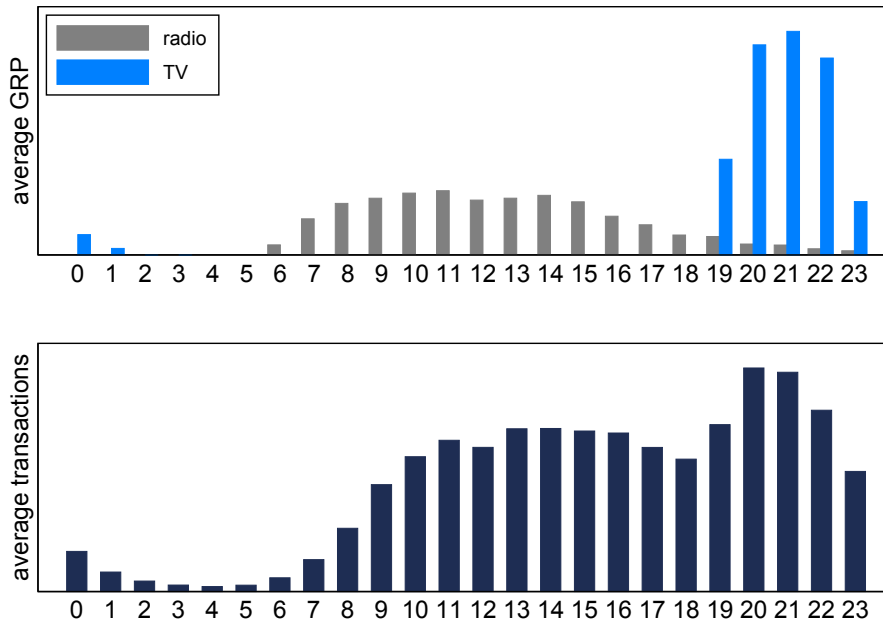
Notes: This figure shows average GRP's and sales for different times of the day. To produce this figure we first aggregate sales at the hourly level and then average over days and draws. We exclude the respective day of the draw because tickets can only be bought until 6pm on that day and there is a lot of advertising activity just before this deadline. See Figure 11 for the pattern on the day of the draw.

3.2 Descriptive evidence

Figure 1 shows cumulative sales for 6 selected regular draws against the time until the draw, together with the respective jackpot size.[6] Some of the draws take place one full month after the previous draw, while others will take place after less than a month. For example, the draw on July 10 follows on the one of June 24 and therefore the line for the draw on July 10 is only from June 24 (6:00 pm) to July 10 (5:59 pm). We do not expect this to have big effects, however, because most tickets are sold in the week before the draw.[7]

The figure shows that across draws there is a positive relationship between jackpot size and sales (that is, cumulative sales on the day of the draw). The draw on July 10 has the largest total sales of the 6 draws. It also has the largest jackpot size. The second largest sales for the draw on June 10, which also has the second largest jackpot size. However, in general, it is not true that larger jackpot size always implies larger total sales.

We further explore differences across draws by regressing the log of the total number of

---

advertisement. At the minimum, this will be the time it takes between listening to a radio commercial in the car and buying a ticket in a shop. Therefore, it will be much more challenging to measure advertising effects in offline data—a challenge we try to overcome with our high frequency online sales data. For the interpretation of our results below we focus on online sales.

[6]Patterns for the other draws are similar. See Figure 10 in the Online Appendix for the remaining draws.

[7]We nevertheless take this into account in our analysis.

Figure 3: GRP's at the minute-level for a regular draw
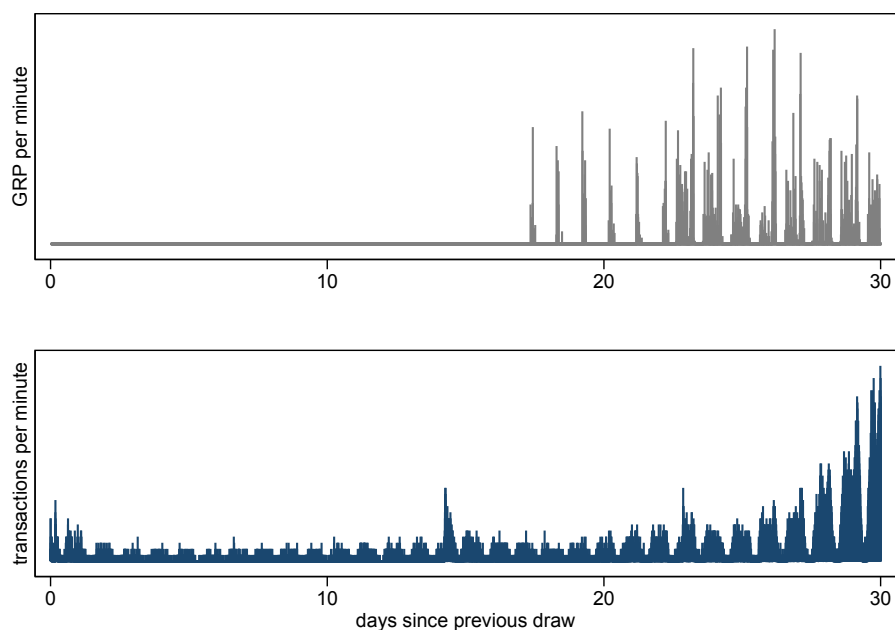
Notes: This figure shows GRP's and sales at the minute level, for the regular draw on April 10, 2014. Tickets for the next draw can be bought from 6pm on the day of the previous draw, which is depicted as 0 days since the previous draw.

tickets sold online on the log of the jackpot size and the total number of days between the date of the previous and current draw.[8] Obviously, we only have 16 observations and jackpot size only varies among the 12 regular draws. Nevertheless, we find a significant relationship between jackpot size and sales. We estimate the effect of a 1 percent increase in the jackpot size to be a 0.4 percent increase in total sales. We find no significant effects of lagged variables on sales.

Figure 2 shows the pattern of sales and GRP's across different hours of a day. We average over all days in 2014 except for the days of the draw. The reason for this is that the time until which tickets can be bought is 6pm and we observe that a large amount of sales occurs during the hours before 6pm. At the same time, we observe that sales are unusually low in the first several hours after 6pm on the day of the draw, as one would expect. So, by excluding those 16 drawing days, we can get a cleaner picture on how sales and GRP's are distributed over time during a typical day.

We distinguish between radio and TV advertisements. TV advertisements are concentrated during evening and night hours, while radio advertisements are more likely to be aired in the morning and in the afternoon. This clear separation is due to the fact that in the Netherlands TV advertisements related to gambling must not be aired during the day time, until 7pm.[9]

---

[8]See Table 5 in Appendix C for details.

[9]We have tried to exploit this regression discontinuity design to produce estimates of advertising effects. However, it turns out to be difficult to distinguish the discontinuity in the total number of GRP's from a flexible time trend. The reason is that number of GRP's increases in a continuous manner between 7pm and 9pm and did not

Figure 4: GRP's at the minute-level for a short time window
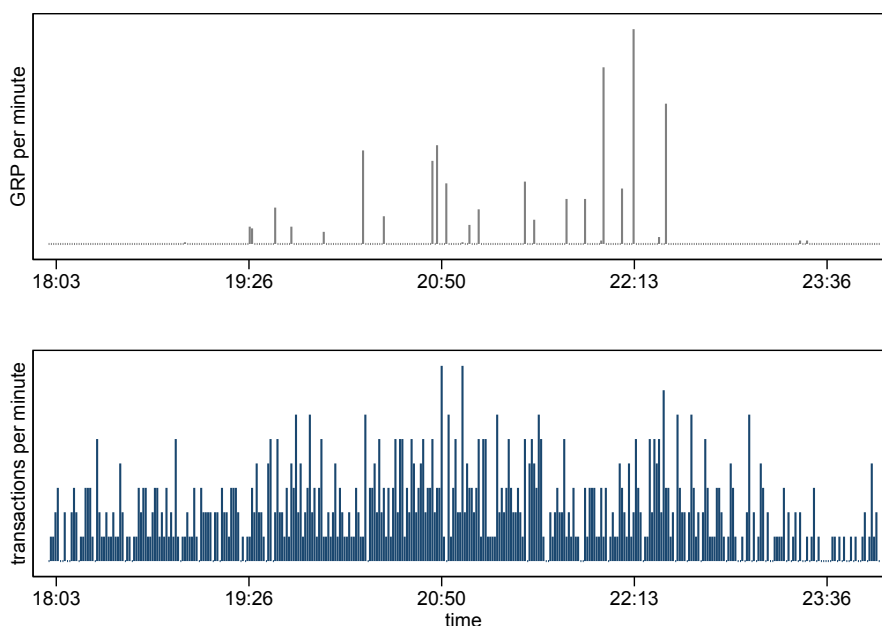
Notes: This figure shows GRP's and sales at the minute level for a short time window on April 3, 2014.

Figure 2 shows that GRP's are positively correlated with sales. During the hours in which sales are high, GRP's are also high. However, this does not necessarily mean that advertising has positive effects, because GRP's have not been assigned randomly. For instance, it could be that consumers have more time in the evening and are therefore more likely to buy a lottery ticket anyway.[10]

Next, Figure 3 shows GRP's and sales at the minute level for one regular draw.[11] We see that the firm starts advertising on the 17th day after the last regular draw, while sales only increase in the last days before the draw. This is already a first indication that advertising effects are low before those last days.

Finally, Figure 4 zooms in further and shows the pattern for one of the days in Figure 3. Related to our identification strategy described below, it is interesting to notice that the raw data presented in Figure 4 already show some evidence of short run sales responses to advertising. For example, there are some spikes of GRP's just before 20:50, followed by spikes of sales several minutes later. In the following section, we investigate this more systematically and show the dependence of advertising effects on the time until the draw.

sharply jump to a high level right after 7pm.

[10]This is a well-known challenge for the analysis of advertising effects. Our identification strategy for measuring the effects of advertising is akin to a regression discontinuity design and described in Section 4 below.

[11]Figure 12 in the Online Appendix shows GRP's and sales for the special draw on April 26. Patterns are similar.

4   Evidence on the effect of advertising

In this section, we empirically characterize the short term effects of advertising how they depend on the time until the draw.

In general, a challenge for the estimation of advertising effects is that sales and advertising are recorded at a low frequency, such as a week or a month. For that reason, they may be confounded by factors unobserved to the econometrician. This then leads to a positive correlation between the two even if advertising effects are zero. Consequently, a regression of sales on the amount of advertising will lead upward-biased estimates of advertising effects even if one controls for month or week dummies. Here, we overcome this challenge by exploiting the high frequency nature of our data.

There are two sources of exogenous variation. The first is related to the fact that advertising buying takes place several weeks in advance. The company specifies, among other things, a time window that is at least several hours long and a target amount of advertising during that time window.[12] This means that the exact timing of advertising is not controlled by the firm. The second source of exogenous variation is that for a given time in the future, it is uncertain how many viewers will be reached, as viewership demand depends on many factors other than the TV schedule, for instance the weather. This means that the target quantity bought by the firm is allocated to multiple spots, until the amount of advertising that was actually bought has been provided (see also Dubé et al., 2005). Consequently, once we control for all factors that drive advertising buying from an *ex ante* perspective we can estimate advertising effects by regressing sales on (lags of) advertising exposure. In practice this amounts to controlling for draw, days to the draw, and hour-of-day dummies We can also control for these confounding factors by means of a fixed effect for each time interval around an advertisement. The idea is then that the variation in advertising within this time window is random.

This identification strategy is akin to the one in a regression discontinuity design: average sales just before the advertisement can be interpreted as a baseline. The average difference between actual sales after the advertisement has been aired and those sales can therefore be interpreted as an estimate of the average effect of the advertisement.

Below we use a variety of different specifications that are all variants of this strategy. We first use data at the minute level to present direct evidence for a selected set of advertisements. Thereafter, we estimate a distributed lag model at the minute level, controlling for time effects in a very flexible way. Then, we aggregate the data to the hourly level to verify that estimated effects are similar. Finally, we provide evidence on the dependence of advertising effects on the time until the draw. Our key finding is that advertising effects are stronger the later an advertisement is aired. In Section 6.2 we develop the argument that this finding is in line with the idea that advertisements act as a reminder.

---

[12]It is in principle possible for the firm to buy specific spots. However, Staatsloterij did generally not do so because the price for those is higher.

Figure 5: The effect of advertising on sales for big advertisements
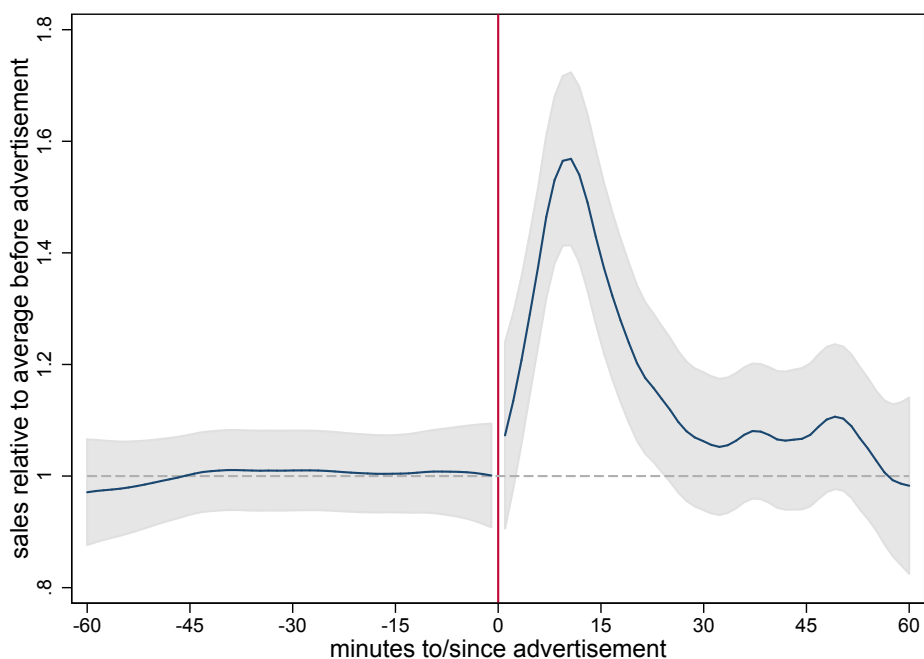
Notes: This figure shows the effect of advertising on sales, relative to average sales in the hour before the advertisement. Obtained using separate local polynomial regressions for the time to and since the advertisement was aired, respectively. We used a fourth-order polynomial and the rule-of-thumb bandwidth. The shaded area depicts pointwise 95 percent confidence intervals. See text for additional details.

4.1 Direct evidence for big advertisements

In our data, there are a number of relatively small advertisements. This means that there is often only a short amount of time between advertisements, which means that providing direct evidence on the effect of advertisements is challenging as advertising effects may overlay each other. Our first approach to overcome this challenge is to *select* advertisements with at least 9 GRP and then only keep the ones out of these advertisements for which we do not see another big advertisement in the hour before and after.[13] Figure 13 in Appendix C shows which advertisements were used.

Then, we regress sales divided by the average number of sales in the hour before the advertisement was aired on time to and since the advertisement, respectively, using two separate local polynomial regressions.

Figure 5 shows the resulting plot of relative sales against the time to and since the advertisement was aired. Notice that sales are flat in the 60 minutes before the advertisement was aired, in line with the idea these constitute a baseline that can be extrapolated. The dashed line denotes average sales before the advertisement was aired. Assuming that this is indeed the

---

[13]Results were similar when we only kept advertisements of sizes bigger than 9 GRP. However, this results in even more selected samples. We also experimented with smaller advertisements but found that effects for those are not measurable in this direct way.

baseline against which sales have to be compared after the advertisement was aired we find that the effect of a big advertisement is an increase in sales that lasts for about 30 minutes. The effect is fairly immediate and dies out relatively quickly. It is as high as 60 percent after a few minutes and overall leads to an increase of sales by 17 percent in the hour after it is aired.[14]

To provide more systematic evidence without selecting advertisements, we next estimate a distributed lag model, still using minute-level data.

4.2   Evidence from a distributed lag model

A distributed lag model is a model in which we regress sales on lagged amounts of advertising. We control for draw, time of the day and days until the draw fixed effects. It is important to control for these time effects in a flexible way, because they would otherwise confound advertising and sales: there are periods in which sales are naturally higher and during which the firm also advertises more, for instance in the last days before the draw. This means that we control for systematic variation in advertising and sales. We would expect an upward bias in our estimates if we did not control for this variation, because we expect the amount of advertising to be higher at the times at which sales are high anyway. After controlling for those time effects, as explained above, we assume that the remaining variation in the amount of advertising is random, which allows us to give our estimates a causal interpretation.[15]

Table 1 shows the results when we use the log of one plus sales as the dependent variable.[16] Column (1) is for our baseline specification. We find that the effect of advertising increases until 10 to 14 minutes after the advertisement was aired and then decreases. The main effect is observed in the first hour, but there are effects thereafter. The maximal effect is an increase in sales of about 3.7 percent for each additional GRP of advertising, between 10 and 14 minutes after the advertisement was aired. The total effect of advertising is an increase of sales by about 2 percent of the baseline sales in one hour.[17]

Moving to column (2) and (3), we find *percentage* increases in sales to be higher before the last week. This is measured from a lower baseline: sales are 8.1 times higher in the last week

---

[14]Note, however, that this is a highly selected set of very big advertisements. The effect of an "average" advertisement is expected to be much lower. We have been told that an effect of an increase in sales by 1 or 2 percent for a typical advertisement is already considered big in the industry.

[15]Note that the empirical strategy we use here is similar, but slightly different from the one we used in Section 4.1 above. The two strategies have in common that we assume that the exact timing of advertising is random conditional on time effects. In Section 4.1 we control for time effects by dividing by the respective number of sales before the advertisements were aired. This is akin to an approach with multiplicative fixed effects in levels or additive fixed effects in logs.

[16]We have also experimented with the pure level of sales. However, we found that the effect of advertising is better captured by this specification (in an R^2 sense). We use the log of one plus sales because there are hours in which sales are zero. We will interpret results as being (approximately) percentage changes. This is slightly worse an approximation as for the case of the pure natural logarithm. To see this denote sales without the advertisement by $sales_t^0$ and with an advertisement by $sales_t^1$. Then, we have that if, say, $\log(1 + sales_t^1) - \log(1 + sales_t^0) = 0.4$, then one can calculate that the increase in sales is about 50 percent provided that sales are above 2.

[17]This can be calculated as the weighted average of the reported coefficients, where the weights are proportional to the length of the captured time interval.

Table 1: The effect of advertising on sales

	(1) baseline	(2) before last week	(3) last week	(4) no controls
GRP between 0 and 4 minutes ago	0.0152***	0.00888***	0.00319**	0.0485***
	(0.000982)	(0.00147)	(0.00100)	(0.00200)
5 and 9 minutes	0.0350***	0.0343***	0.0193***	0.0677***
	(0.00101)	(0.00176)	(0.00103)	(0.00195)
10 and 14 minutes	0.0369***	0.0378***	0.0205***	0.0685***
	(0.000881)	(0.00151)	(0.000955)	(0.00184)
15 and 19 minutes	0.0272***	0.0250***	0.0124***	0.0578***
	(0.000911)	(0.00144)	(0.000982)	(0.00187)
20 and 24 minutes	0.0224***	0.0176***	0.00926***	0.0517***
	(0.000908)	(0.00140)	(0.000924)	(0.00185)
25 and 29 minutes	0.0194***	0.0124***	0.00748***	0.0468***
	(0.000978)	(0.00148)	(0.00103)	(0.00187)
0.5 and 1 hour	0.0152***	0.00992***	0.00217***	0.0375***
	(0.000394)	(0.000579)	(0.000441)	(0.000719)
1 and 1.5 hours	0.0106***	0.00737***	0.00102*	0.0242***
	(0.000389)	(0.000573)	(0.000433)	(0.000720)
1.5 and 2 hours	0.00830***	0.00469***	0.00117**	0.0192***
	(0.000397)	(0.000568)	(0.000446)	(0.000698)
2 and 2.5 hours	0.00214***	0.000657	0.000364	0.00962***
	(0.000350)	(0.000563)	(0.000423)	(0.000608)
2.5 and 3 hours	-0.00178***	-0.00330***	-0.000181	0.00720***
	(0.000334)	(0.000549)	(0.000400)	(0.000584)
3 and 3.5 hours	-0.00648***	-0.00458***	0.000340	0.00411***
	(0.000312)	(0.000519)	(0.000387)	(0.000527)
3.5 and 4 hours	-0.0105***	-0.00744***	-0.000290	0.00476***
	(0.000310)	(0.000487)	(0.000390)	(0.000485)
draw dummies	Yes	Yes	Yes	No
days to draw dummies	Yes	Yes	Yes	No
hour dummies	Yes	Yes	Yes	No
Observations	515205	361070	154135	515205
R^2	0.624	0.320	0.713	0.192

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows the results of regressions of the log of one plus sales on GRP's of advertising and lags thereof. Regressions were carried out at the minute level and standard errors are robust to heteroskedasticity. Regressions separately for TV and radio advertising are shown in Table 6 in Appendix C.

(see also Figure 1 that shows cumulative sales). Using this, we find that the *absolute* effect of advertising in the first hour after the advertisement was aired is about 3.5 times higher in the last week before the draw. This dependence of advertising effects on the time until the draw is closely related to advertisements acting as a reminder. In Section 4.3, we will characterize it in more detail.

Finally, in the last column, we carry out the same regression as in column (1), except that we do not control for time. As explained before, this should lead to upward-biased estimates as we then face an endogeneity problem as sales and advertising are jointly determined. And indeed, coefficient estimates are much higher.

Table 6 in Appendix C shows results from a specification in which we distinguish between TV and radio advertisements. We find the effects of TV advertising to be stronger, but to die out faster. Radio advertisements have a longer lasting effect. This could be related to the fact that it takes time to actually purchase a ticket. When seeing a TV advertisement, an individual may buy directly using his smart phone, sitting in front of the TV. To the contrary, when listening to a radio advertisement she could be driving her car or be occupied with something else and buy the ticket at the next occasion after having finished her ride or her task. Notice that the overall effects are nevertheless similar in terms of size and therefore we generally don't distinguish between TV and radio advertisements in this paper.

We have also estimated similar models using data aggregated to the hourly level. Table 7 in Appendix C shows the results. We find that advertising has a similar effect in the hour in which it is aired as it has in the following hour: on average, one GRP of advertising leads to about a 1.2 percent increase in the amount of tickets sold. The effect is about one third of this two hours after the advertisement was aired and not significantly different from zero (at the 5 percent level) after 3 hours. Comparing Table 1 to Table 7 shows that aggregating the data to the hourly level does not seem to have a big impact on the effect of advertising that we estimate: focusing on baseline specification, we find that the effect for the first hour that we estimate in the minute-level regressions is about 2 percent on average, and about 1 percent in the second hour. This is important, because it would not be feasible to estimate a structural model at the minute level.

4.3 The dependence of the effect of advertising on time

So far, we have shown that advertising effects are measurable using our data. We have presented first evidence pointing towards them being stronger the less time there is until the draw. This dependence of advertising effects on the time until the draw is related to advertisements acting as a reminder. We now develop this argument more systematically and then characterize the relationship between advertising effects and the time until the draw in more detail.

The way we think about consumers is that they are exposed to limits of information processing power and attention, which may lead to forgetting. Once they think about buying a lottery

Figure 6: Effect of timing
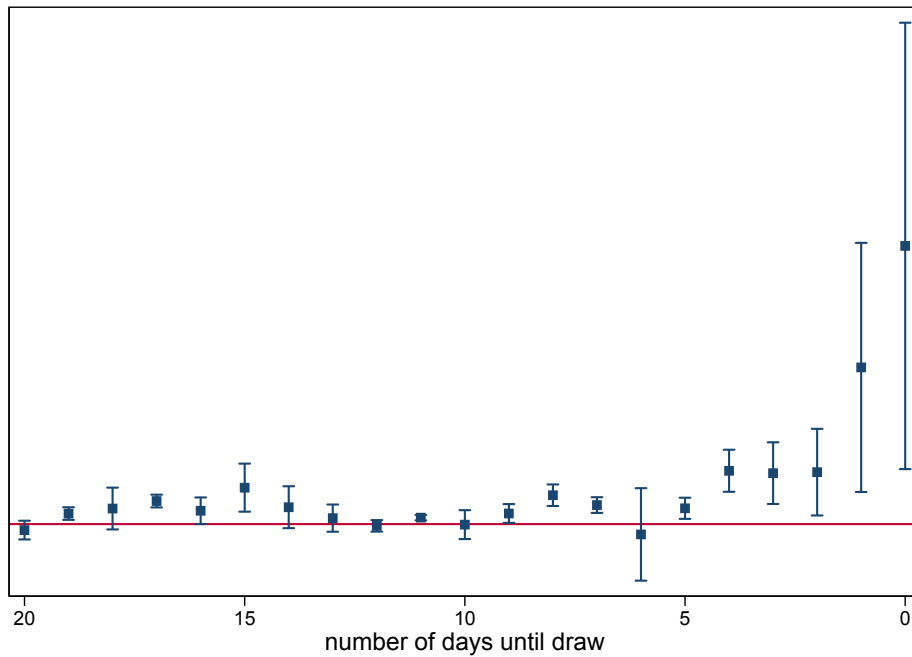
ticket, they weigh the costs of doing so at that point in time against the benefits. Costs here can be both, monetary and non-monetary, and may also include effort costs. Benefits are delayed, because the draw will only take place in the future, while costs are immediate. For that reason, consumers value to be reminded to buy a ticket as late as possible. In addition, when reminded early and therefore considering to buy a ticket early, they will be more reluctant to do so the less likely it is that they will consider buying a ticket in the future. After all, they are still able to buy in the future. This means that they will be more likely to buy, once reminded, when there is a substantial risk that they will not consider buying a ticket in the future. This in turn means that advertising effects will be strongest right before the deadline, because that is the last time at which they can buy a ticket if they have not done so already. Moreover, advertising effects will tend to decrease in the time until the deadline.

This way of reasoning is fully compatible with the structural model we propose in Section 5. The model can be seen as a formal version of the above argument. We estimate the structural parameters of this model and then use it to predict sales for alternative counterfactual advertising strategies. One of the model properties is that advertising effects depend on the time until the draw (see Figure 6 below).

To characterize the relationship between advertising effects and the time until the draw in more detail, ideally, we would estimate a different response curve for every day, but this is not feasible. Therefore, we instead estimate the immediate absolute effect of advertising on ticket sales and relate it to the number of days that are left until the draw. For this, we aggregate data

15

to the hourly level and take first differences to control for patterns in baseline sales. We specify

$$sales_t - sales_{t-1} = \beta_0 + \beta_1 \cdot (grp_t - grp_{t-1}) + \varepsilon_t \tag{1}$$

where $sales_t$ is the number of tickets sold in hour t and grp_t is the number of GRP's of advertising in t. We set GRP's to 0 if they are below 3, in order to single out advertisements that are big enough to have a measurable impact.[18] Moreover, guided by the finding in the previous subsection that the advertising effect lasts for about 4 hours, we drop observations where we see more than 3 GRP's of advertising in any of the four hours prior to that, which means that also $grp_{t-1} = 0$ in (1). Thereby, we ensure that advertising effects of previous instances have died out. Hence, the coefficient β_1 that we are estimating is the immediate increase in sales in response to increasing the GRP's by one, relative to sales before when there was no advertising. We run a separate regression for each day until the draw. We also control for draw and day of the week fixed effects to allow for differences in time trends across those.[19]

In Figure 6, we plot the estimated effects and the corresponding 95% confidence intervals against the number of days until the draw.[20] Towards the time of the draw, the effects increase, in line with the idea that advertisements act as a reminder.

We can use this empirical setup to make an additional observation. In general, if advertisements have an effect, then it could either be that consumers are motivated to buy earlier, but would have bought anyway (purchase acceleration). Or it means that consumers that would otherwise not have bought decided to buy (market expansion). Usually, it is challenging to empirically tell these apart from one another. To a large extent, this is the case because typically, consumers always have the possibility to buy a product later. However, in our case, there is a fixed ending time up to which lottery tickets can be bought. This provides us with the opportunity to study whether advertising also has an effect until shortly before the draw, which is what we find. This is direct evidence suggesting that advertising does not only lead to purchase acceleration but also to market expansion.

To summarize, exploiting the high frequency nature of our data, we have shown that advertising leads to economically sizable direct effects on sales in the order of a 2 percent increase. The absolute effect of advertising on sales is higher the less time there is until the draw, in line with the idea that advertisements act as a reminder. Our results also show that advertising does not only lead to purchase acceleration, but also to market expansion.

---

[18] There are many very small advertisements. Those small advertisements will lead to small increases in sales that we ignore. For that reason, the specification we use here is conservative because the estimated effects are lower bounds.

[19] Recall that the dependent variable is the difference in sales over time.

[20] We have also tried to "zoom in" and show that the effect is there in the very last hours before the draw, but we only have data for 16 draws, with a limited number of advertisements in the last hours before the draw.

5 A model of lottery ticket demand

Informed by the model-free evidence, we now spell out our dynamic structural model of ticket sales. The model is useful to rigorously describe the idea that advertisements act as a reminder. Moreover, while the data are informative about short run effects of advertising, inferring medium run effects directly is challenging because the necessary exogenous variation is typically missing. A model can be used to quantify these effects. Obviously, assumptions have to be made for this. Related to this, once the structural parameters that allow us to capture both, short and medium run effects, are estimated, we can predict the total amount of sales for any counterfactual dynamic advertising strategy. This is not possible without estimating a model, even if exogenous variation is present.

As pointed out before, our model has elements of the rational adoption models by Melnikov (2013) and De Groote and Verboven (2016). In an adoption model, consumers decide when to buy a product. The way taste shocks affect dynamic decision making is modeled as in Rust (1987). We augment this model by advertising affecting consumer choice through an advertising goodwill stock that increases the probability that a consumer will consider buying a ticket at a given point in time.

An important generalization relative to other models with an advertising goodwill stock is that the advertising goodwill stock differs across consumers. At a given point in time, some of them are reached—the percentage is known and given by the number of GRP's—while others are not. We implement this by simulating whether or not advertising reaches each member of a number of simulated consumers whom we follow over time. These simulated consumers therefore have heterogeneous advertising goodwill stocks.[21]

We now first describe the building blocks of our model. Then we describe how to solve it and take it to the data. The robustness to making alternative assumptions on the market size and viewership behavior is assessed in Appendix B.

5.1 General structure

There are N expected discounted utility-maximizing consumers. Choice is independent across draws. Time $t = 1, 2, \ldots, T$ is discrete and finite and measured at the hourly level. T is the hour of the draw and the last moment at which consumers can buy a ticket. Each individual can buy at most one ticket.

In every hour, each individual decides whether or not to buy a lottery ticket. If she does, then she receives a one-off flow of utility and cannot make any decisions anymore.[22] Otherwise, she continues in the next period and has the option of buying a ticket there.

---

[21] See Section 5.6 below for details.

[22] This is isomorphic to a model in which she pays a price today and expects to receive a flow utility in the future, provided that she cannot make any decisions in the meantime.

5.2 Consideration

In our model, advertising affects the likelihood that a consumer considers buying a ticket through an advertising goodwill stock.[23] This goodwill stock increases if the individual is exposed to an advertisement, but from an *ex ante* perspective it is uncertain for the consumer whether she will be exposed to an advertisement.

The number of GRP's in our data are informative about how many consumers are reached at a given point in time and we use it to simulate a number of goodwill stocks for different consumers. This is similar to, but also extends the specification of Dubé et al. (2005), where the goodwill stock is the same for all individuals. In our model, the goodwill stock is not the same for all consumers, but the probability to see an advertisement in a given period is the same. That is, consumers are identical *ex ante*, but we simulate how they differ *ex post*.

Denote the goodwill stock of individual i at the beginning of period t by g_{it}. We will refer to the goodwill stock after the time at which the individual can be reached by an advertisement as the augmented goodwill stock. It is denoted by g_{it}^a. The augmented goodwill stock affects consumer choice and depreciates exponentially over time. Let λ denote the depreciation rate and assume that the initial goodwill stock is 0. The law of motion for the (augmented) goodwill stock is

$$g_{it}^a = \begin{cases} g_{it} & \text{if } i \text{ did not see an advertisement in } t \\ g_{it} + 1 & \text{if } i \text{ saw an advertisement in } t \end{cases}$$

with initial condition

$$g_{i0} = 0$$

and

$$g_{it+1} = (1 - \lambda) \cdot g_{it}^a.$$

The augmented advertising goodwill stock then affects the probability to consider buying a ticket. We specify this probability as

$$P_{it}(\text{consider}) = \frac{1}{1 + \exp\left(-(\gamma_0 + \gamma_1 g_{it}^a)\right)}.$$

---

[23]Advertising can also have brand building effects across draws. This will be captured by draw fixed effects that also capture all the other across-draw effects. We will abstract from across-draw advertising effects in our counterfactual simulations because they are not separately identified from all other differences across draws without making strong assumptions.

5.3 Purchase decision

In the consideration stage, a consumer decides whether or not to buy a lottery ticket. Buying a ticket yields flow utility

$$u_{it} = -p + \delta^{T-t}\psi + \sigma\varepsilon_{i1t},$$

where p is the price of the ticket, δ is the hourly discount factor, ψ is the value of holding a ticket at the time of the draw, and ε_{it} is a type 1 extreme value distributed taste shock (recentered, so that it is mean zero). The coefficient on the price is normalized to be minus 1, which means that flow utility is measured in terms of money. Specifying flow utility to depend on $-p + \delta^{T-t}\psi$ means that a consumer has a taste for buying the ticket as late as possible because she has to pay for it immediately but only receives a discounted benefit from this. This feature of our model is meant to capture the empirical pattern that most sales occur in the last days before the draw (see Figure 1).

If a consumer chooses not to buy before the last period, she gets the continuation value $\delta\mathbb{E}[V_{t+1}(g^a_{it+1})|g^a_{it}] + \sigma\varepsilon_{i0t}$, where again ε_{i0t} is a type 1 extreme value distributed taste shock and $V_{t+1}(\cdot)$ is the value function tomorrow that is a function of advertising goodwill stock tomorrow. The expectation here is taken over whether or not the consumer will consider buying a ticket, whether she is reached by an advertisement, and future realizations of the taste shocks. We provide more details below in Section 5.5. If she does not buy in the last period, then the terminal value is $\sigma\varepsilon_{i0t}$.

In our model, as explained above, there is a cost to buying earlier. The benefit is that consumers won't forget to buy a ticket later, if they want to do so in principle, as they can't be sure to consider doing so in the future. Hence, they may forget.

5.4 Expectations

In our model, expectations about future advertising play an important role, as advertising reminds consumers to buy a ticket by increasing the probability that a consumer will consider doing so. The scalar state variable g^a_{it} summarizes all relevant information on consumer i's advertising exposure in the past. In addition, the value function in Section 5.3 is indexed by t because the consumer problem is a finite horizon one and because the probabilities to see advertisements in the future change over time. If, for example, the consumer knows that there is a large probability that she will see an advertisement tomorrow (or shortly before the draw), then she may be more likely to delay her purchase to tomorrow because she will likely be reminded to buy a ticket.

There are two ways in which we could proceed regarding these expectations when solving and structurally estimating the model. We could either solve a game between the consumers and the firm and then use the implied beliefs. This, however, may not be promising because there could be multiple equilibria, and it may be hard to solve that game in the first place. Moreover,

19

we would have to do this within every iteration of our estimation procedure, which would be computationally challenging (if not infeasible). And most importantly, we would have to make the strong assumption that the advertising strategy of the firm that we observe was actually optimal. Instead, we estimate this probability from our GRP data. The specification we use for this is

$$grp_t/100 = x_{it}'\beta + \varepsilon_t, \tag{2}$$

where x_t includes a constant term and a full set of hour, day, and draw dummies. The fitted value is then the probability to see an advertisement in t, which we denote by P_t. Figure (14) in Appendix C shows this probability together with the ones we use in our counterfactual experiments (discussed below). We take the these expectations $\{P_t\}_{t=1}^T$ about advertising activities as known.

5.5 Solving the model

We now describe how we solve the model for given values of the parameters, which we then vary in the outer loop of our estimation procedure. Recall that one time unit is equal to one hour. Also, observe that the time of the day does not enter the model directly. Instead, we count the time between midnight and 7am as 1 hour. This choice is guided by Figure 2 where one can see there are little sales during those hours.[24]

The state variables are time, whether or not a consumer has already bought a ticket, and the advertising goodwill stock g_{it}^a. The first two state variables are discrete, while the advertising goodwill stock is non-negative real-valued. The time horizon is finite. We solve this model recursively on a grid for the advertising stock, using interpolation to compute continuation values. We use an equally spaced grid with $G = 2000$ grid points. Denote the set of grid points by $\mathscr{G}$. We use the same grid points in each time period.

The structure of the adoption model simplifies the computation considerably, as individuals can buy at most once and the value to buying consists only of the flow utility. The main task is to compute the value to not buying, for every t and on the grid for the advertising stock. Another simplifying factor is that individuals will either see an advertisement in the next hour or not, with a known probability. This means that we can write down an expression for the corresponding expectation over this event and don't have to use simulation or numerical integration. The assumption that the taste shocks are distributed type 1 extreme value allows us to also find an

---

[24]One could in principle model the flow utility to depend on the time of the day and also on the day of the week. However, this would come at the cost of substantially increasing the computational burden. In the estimation procedure (described in more detail in Appendix A below), we solve the model each time we evaluate the objective function. However, it is unlikely that we will suffer from the same omitted variables bias as we would when estimating a distributed lag model as in Section 4.2 without controlling for hour of day effects. The reason for this is that the model structure imposes a lot of smoothness in the sense that sales in adjacent hours are predicted to be very similar to one another. Most of the time there are no advertisements and therefore parameters capturing the evolution of baseline sales net of hour-of-day and day-of-the-week effects will not be biased. Given this advertising effects will also be unbiased.

analytic expression for the value to not buying in period t, given the value function in $t+1$, as in Rust (1987). For that reason, we can solve the model relatively fast and on a grid with many grid points.

We solve the model recursively. For each time period t and grid point $\tilde{g}_{it}^a \in \mathscr{G}$, we calculate the expected value function in the next period, $\mathbb{E}\left[\max V_{it+1} | \tilde{g}_{it}^a\right]$; the value when considering to buy in the current period, $V_{it}^c(\tilde{g}_{it}^a)$; and the value in the current period $V_{it}(\tilde{g}_{it}^a)$. Next, we provide more details.

First consider the case in which an individual has not bought before the last period $t = T$ and the goodwill stock takes on the value $\tilde{g}_{it}^a \in \mathscr{G}$ on the grid. Then, the value to not buying is 0 because there is no future period. The value when considering in the last period is

$$
V_{iT}^c = \sigma \cdot \log \left[\exp \left(\frac{\delta \cdot 0}{\sigma} \right) + \exp \left(\frac{-p + \psi}{\sigma} \right) \right],
$$

where $\delta \cdot 0$ is the discounted value of not buying, which is zero because the individual cannot buy in the future, and $-p + \psi$ is the mean utility associated with buying. From this it follows that the value in the last period is

$$
V_{iT}(\tilde{g}_{iT}^a) = P_{iT}(\text{consider}) \cdot V_{iT}^c + (1 - P_{iT}(\text{consider})) \cdot \delta \cdot 0,
$$

where, again, $\delta \cdot 0$ is the value associated with not buying.

Now turn to the case in which an individual has not bought before $t = T - 1$, the second to last period. The expected value function in the next period, $\mathbb{E}\left[\max V_{it+1} | \tilde{g}_{it}^a\right]$, is

$$
\mathbb{E}\left[\max V_{iT} | \tilde{g}_{iT-1}^a \right] = P_T \cdot V_{iT}(\tilde{g}_{iT}^{a+}) + (1 - P_T) \cdot V_{iT}(\tilde{g}_{iT}^{a-}).
$$

Here, the expectation is taken over the advertising goodwill stock and the taste shocks. The goodwill stock $\tilde{g}_{iT-1}^a$ either changes to $\tilde{g}_{iT}^{a+}$ when the individual sees an advertisement in T, which will be the case with probability P_T, or it changes to $\tilde{g}_{iT}^{a-}$ if not, with probability $1 - P_T$ (see also Section 5.2). The two values $V_{iT}(\tilde{g}_{iT}^{a+})$ and $V_{iT}(\tilde{g}_{iT}^{a-})$ are obtained using interpolation. From this, we get that the value when considering in the second to last period is

$$
V_{iT-1}^c(\tilde{g}_{iT-1}^a) = \sigma \cdot \log \left[\exp \left(\frac{\delta \cdot \mathbb{E}_{T-1}\left[\max V_{iT} | \tilde{g}_{iT-1}^a \right]}{\sigma} \right) + \exp \left(\frac{-p + \delta \psi}{\sigma} \right) \right]
$$

and the value function is

$$
V_{iT-1}(\tilde{g}_{iT-1}^a) = P_{iT-1}(\text{consider}) \cdot V_{iT-1}^c + (1 - P_{iT-1}(\text{consider})) \cdot \delta \cdot \mathbb{E}\left[\max V_{iT} | \tilde{g}_{iT-1}^a \right].
$$

We proceed in a similar manner for the remaining time periods up to $t = 1$. This results in values $V_{it}(\tilde{g}_{it}^a)$ for all t and all $\tilde{g}_{it}^a \in \mathscr{G}$. From those, we can calculate the probability of buying

given consideration as

$$P_{it}(buy|consider) = \frac{\exp\left(\frac{-p+\delta^{T-t}\psi}{\sigma}\right)}{\exp\left(\frac{-p+\delta^{T-t}\psi}{\sigma}\right) + \exp\left(\frac{\delta \cdot \mathbb{E}_t\left[\max V_{it+1}|\tilde{g}_{it}^a\right]}{\sigma}\right)}$$

and the unconditional probability of buying as

$$P_{it}(buy) = P_{it}(\text{consider}) \cdot P_{it}(\text{buy|consider}).$$

5.6   Empirical implementation

In the first stage, we estimate the probability P_t to see an advertisement at any given point in time, as described in Section 5.3 above. In the second stage, we take these probabilities as given and estimate the parameters of the structural model. There is an inner and an outer loop. In the inner loop, we simulate consumer choice for given values of the parameters and compute the value of a method of simulated moment (MSM) objective function. In the outer loop we then estimate the parameters. The moments we use are related to sales at a given point in time given the advertising activity before that, and the evolution of cumulative sales.

We assume the market size for Dutch online lottery tickets market is 250,000 and we simulate choices of 1,000 consumers.[25] Thus each simulated consumer represents 250 real consumers. This is again a trade-off between computational burden and how realistic the model is. To implement this, we take aggregate sales and divide them by 250. The thought experiment that underlies our approach is that we match simulated sales to the expectation thereof, across 250,000 actual consumers, which is given by our data.

In our aggregate data, we only observe that a consumer has bought a ticket, but not which ticket. We assume that the price of the tickets bought is 3 euros. The key assumption we make here is that everybody buys the same ticket.[26]

In our estimation procedure we pay particular attention to the fact that different consumers have different advertising stocks at a given point in time, as it is random whether or not they are exposed to advertisements in the periods before that. Tentatively, there will be dynamic selection in the short run, because those consumers with higher advertising goodwill stocks will

---

[25]This is considerably more than the maximum number of tickets that was sold in each month in our data. We experimented with different market sizes and found that results of the counterfactual simulations are not very sensitive to it. In Section B we also present results when we assume that the market size is twice as high.

[26]3 euros is the price for the smallest ticket one can buy. See Section 2 for details. See also footnote 5. Assuming a different price will only re-scale the parameters, but will not change the results of counterfactual experiments. To see this, suppose we double the price and double at the same time Ψ and σ. Then, it follows from the expressions above that V_{iT}^c, $V_{iT}(\tilde{g}_{iT}^a)$ and $\mathbb{E}\left[\max V_{iT}|\tilde{g}_{iT-1}^a\right]$ will double. Consequently, $V_{iT-1}^c(\tilde{g}_{iT-1}^a)$ and $V_{iT-1}(\tilde{g}_{iT-1}^a)$ will double. But importantly, $P_{it}(buy|consider)$ and $P_{it}(buy)$ will stay exactly the same. This shows that both models are observationally equivalent. Consequently, simulated sales under counterfactual advertising strategies will be the same. This means that given that we assume that everybody buys the same ticket, setting the price to a particular value is a normalization.

be more likely to buy, so that those with lower advertising goodwill stocks remain. Our strategy allows and controls for that. For an example, think of 250,000 individuals who may in principle buy a ticket (the market size we assume). Suppose that there are 3 GRP's of advertising in a given hour and that there have not been any advertisements before that. Then, in expectation, 7,500 individuals will be reached. Now suppose that there are 4 GRP's of advertising in the hour after this. This reaches in total 10,000 individuals. Some of those individuals were among the 7,500 who have already seen an advertisement before and some of those will not. We assume that it is independent over time who is reached and therefore 300 individuals will see both advertisements.

After solving the model and simulating advertising goodwill stocks, we follow the simulated consumers with heterogeneous advertising goodwill stocks, at which we evaluate the value functions that we have already solved for, and then combine them with random draws u_{it} from the standard uniform distribution for each consumer at each point in time to generate simulated choices. To be precise, we calculate the probability of buying at the simulated goodwill stock $\tilde{g}_{it}^a$ for all time periods and compare them to u_{it}. If $P_{it}(\text{buy}) \geq u_{it}$, then the simulated choice $\hat{d}_{it}$ is one and otherwise zero. A consumer can buy at most one ticket and therefore we set this variable to zero after a consumer has bought for the first time. Aggregating gives simulated aggregate demand, which we match to (rescaled, as described above) actual aggregated demand. Further details are provided in Appendix A.

6 Results

6.1 Parameter estimates and fit

In this section, we present our estimation results and assess the fit of the model. Table 2 shows the estimated parameters. The effect of advertising on sales depreciates quickly, at an hourly rate of about 33.4 percent. The baseline probability of considering is estimated to be $1/(1 + exp(-\gamma_0)) \approx 0.39$. This means that 39 percent of the consumers will consider buying a ticket in the absence of advertising. The effect of the goodwill stock on flow utility (γ_1) is estimated to be 1.637, which means that a one unit increase in the goodwill stock from zero to one, driven by seeing an advertisement, will increase the probability of considering to $1/(1 + exp(-(\gamma_0 + \gamma_1))) \approx 0.77$. One hour later, the goodwill stock is $1 - 0.334 = 0.666$ and the probability to consider buying is 0.66 if no advertisement reaches the consumer. Yet another hour later it is 0.444 and the probability to consider is 0.57. If the consumer is instead reached by another advertisement in the hour after she was first reached, then the augmented goodwill stock becomes 1.666 and the probability to consider buying is 0.91 in the second period. And when she is reached again one hour later, it is 0.95. This form of concavity in the goodwill stock is the reason why consumers prefer it when advertisements are spread over time. Then the expected number of periods in which they consider buying is maximized.

23

Table 2: Parameter estimates

parameter	estimate	ste.
depreciation rate goodwill stock (λ)	0.334	0.197
effect of goodwill stock on probability of considering (γ_1)	1.637	0.731
intercept of goodwill stock on probability of considering (γ_0)	-0.430	0.329
hourly discount factor (δ)	0.994	0.000
multiplying factor taste shock (σ)	0.318	0.010
value to having a ticket on the day of the draw		
10 January, 2014	1.477	0.056
10 February, 2014	1.669	0.039
10 March, 2014	1.493	0.053
10 April, 2014	1.448	0.067
26 April, 2014 (King's Day)	1.906	0.052
10 May, 2014	1.620	0.046
10 June, 2014	1.711	0.049
24 June, 2014 (Orange draw)	1.728	0.046
10 July, 2014	1.887	0.058
10 August, 2014	1.410	0.060
10 September, 2014	1.644	0.053
1 October, 2014 (special 1 October draw)	1.608	0.048
10 October, 2014	1.541	0.057
10 November, 2014	1.510	0.051
10 December, 2014	1.749	0.045
31 December, 2014 (New year's eve draw)	2.336	0.066

Notes: Structural estimates. Obtained using the method of simulated moments. See Section 5.6 and Appendix A for details on the estimation procedure. The probability to consider buying is specified as $P_{it}(\text{consider}) = 1/\left(1 + \exp\left(-\left(\gamma_0 + \gamma_1 g_{it}^a\right)\right)\right)$.

Figure 7: Model fit
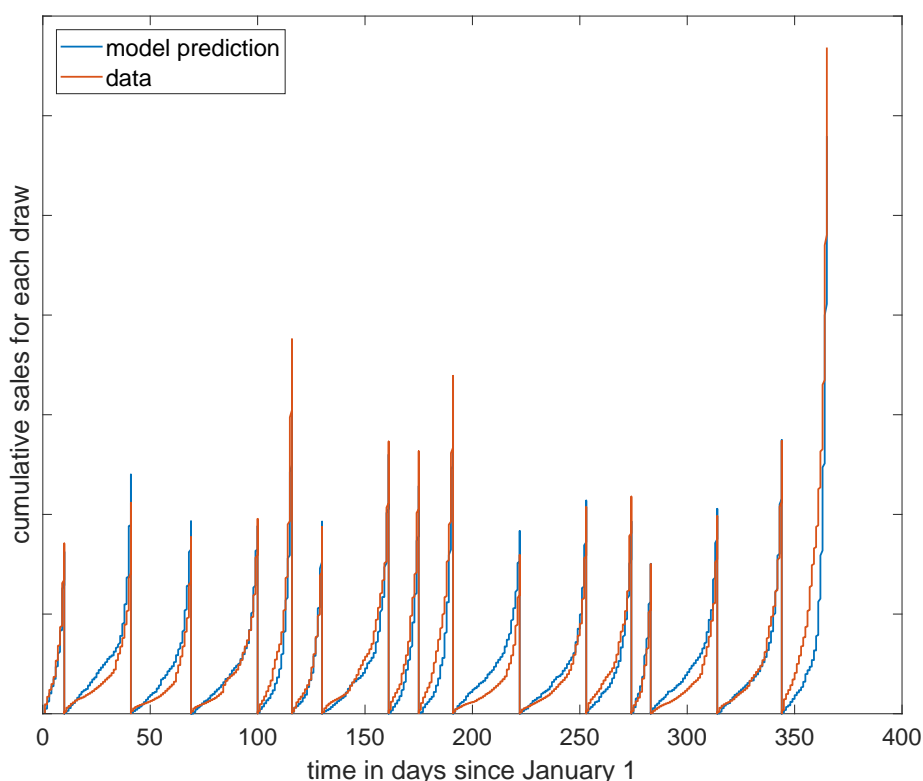
Notes: This figure shows actual cumulative sales and cumulative sales predicted by our structural model using the estimated parameters reported in Table 2.

The hourly discount factor is estimated to be 0.994. This means that one month before a draw, the value consumers attach to a ticket is only 3.9% of the value on the day of the draw. For that reason, consumers will value buying tickets late and being reminded at later points in time. Together with the desire to spread advertisements over time this gives rise to an interesting tradeoff that we explore further in our counterfactual experiments.

The estimated standard deviation of the taste shock is $0.318 \cdot \sqrt{\pi^2/6} \approx 0.41$ euros ($\sqrt{\pi^2/6}$ is the standard deviation of a type 1 extreme value random variable). Finally, the 16 estimates of the draw fixed effects are in line with expectations and positively related to the size of the jackpot, mirroring the pattern in Figure 1.[27]

Figure 7 shows the model fit. Arguably, with only a few parameters, the model fits the overall patterns in the data relatively well.

---

[27]Observe that they are smaller than the price, 3 euros. This is expected because only a relatively small fraction of individuals actually buys a ticket, driven by favorable draws of the taste shocks. For that reason, the value to holding a ticket is higher than the mean utility for those who buy a ticket.

Figure 8: Dependence of predicted effect of advertising on timing
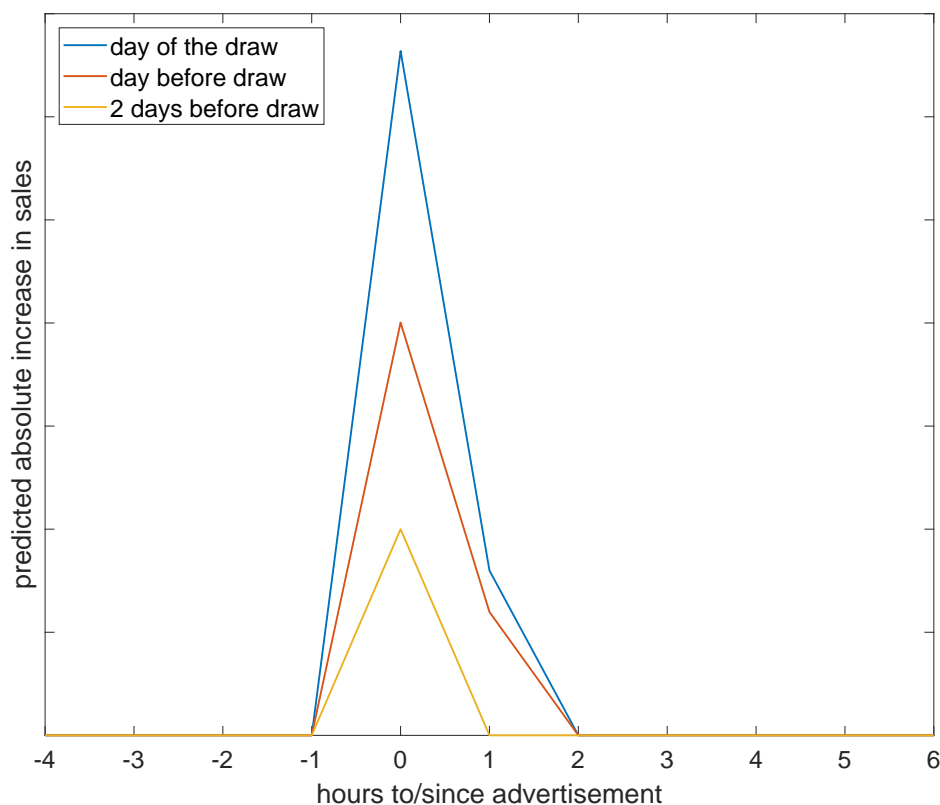
Notes: This figure shows how the predicted absolute increase in sales that is due to seeing an advertisement depends on the time until the draw. Obtained from our structural model using the estimated parameters reported in Table 2.

6.2 The dependence of advertising effects on time

A key quantity the model predicts is the immediate effect of advertising on sales and how this effect depends on the time until the draw. Figure 8 shows, for our structural estimates, how the probability of buying a ticket, if the individual has not done so yet, changes when she is exposed to an advertisement. There are three lines for three different time periods, just before the draw and 1 and 2 days before that. As one can see, the closer the time of the advertisement is to the deadline, the more effective is the advertisement—in line with the model-free evidence that we have presented in Section 4. The figure shows that our model can generate this effect.

7 Counterfactual experiments

Having estimated the model, we turn to the supply side. We do not have access to data on the profitability of an additional sold ticket, and also not on the cost of one GRP. It is, however, not unreasonable to assume as an approximation that the price of one GRP does not vary over time. Therefore, it is meaningful to study whether a given (monetary) budget could be allocated better over time, by asking the question whether it is possible to sell more tickets when one allocates

the same number of GRP's in a different way.

In the following, we use the model to generate counterfactual predictions about the total number of tickets sold. We consider 8 alternative strategies and compare the number of tickets sold to the simulated one for the original GRP schedule in the data. The first alternative strategy is to remove all advertising. In the second, we allocate all advertising to the last two days before the draw and distribute it equally over all hours on those two days. In the third, we spread all advertisements equally in the last 4 days before the draw. Then, there are three pulsing strategies. Studying the effect of those is interesting, as the model is non-linear in advertising exposure and its history, and therefore reacts to such patterns (Dubé et al., 2005). In the first pulsing strategy, the firm advertises in the last hour before the draw, but not in the second to last hour, again in the third to last hour, and so on, for the last 4 days. The amount of advertising, when the firm does so, is always the same. The second pulsing strategy proceeds similarly, but in blocks of two hours. The third pulsing strategy always allocates twice the amount in one hour and then pauses for three, and also lasts for 4 days. The last two counterfactual strategies take, respectively, the schedule as it is in the third week and move it to the fourth week, and *vice versa*.

When simulating the impact of those strategies, we distinguish between two cases. In the first case, we assume that expectations individuals have about the likelihood to be reminded in the future, by seeing an advertisement, remain unchanged (and in line with what we have used to estimate the model) even though we change the advertising strategy. The second case is one in which consumers' expectations are rational in the sense that they reflect the changes advertising strategy (we assumed this when we estimated the model). Making this distinction is interesting because it allows us to quantify the relative importance of changes in expectations. One can think of this as a second order effect, with the first order effect being the change in the actual advertising strategy.

Table 3 shows the result. We first focus on the last column, for rational expectations. Not advertising at all leads to 94% of the original sales. Generally, allocating advertising to later points in time increases sales. Doing the opposite leads to a decrease in sales. Moving it all to the last two days before the draw has the biggest effect, but may be infeasible in practice. Shifting advertising from the third week before draw to the fourth week may be feasible and leads to an increase of sales by 3 percent. The most successful pulsing strategy is the one with blocks of one hour.

Figure 9 shows the underlying dynamics. We plot the difference between the cumulative sales for a given strategy and the baseline strategy. As an illustration, consider the strategy of shifting all the GRP's from the fourth week to the third week. As expected, sales increase in the third week and decrease in the fourth week. Overall, fewer tickets are sold, which is reflected in the lower end point.

The results also show that expectations of consumers matter, but are quantitatively not of first-order importance. Qualitatively, when expectations are rational, then effects become

Table 3: Effect of various advertising strategies

strategy	expectations	
	unchanged	rational
data (reference point)	100%	100%
no advertising at all	94%	94%
all advertising in the last 2 days before the draw	109%	108%
spreading advertisements equally in the last 4 days before the draw	105%	104%
pulsing strategy in the last 4 days before draw (1 hour blocks)	105%	104%
pulsing strategy in the last 4 days before draw (2 hour blocks)	104%	103%
pulsing strategy in the last 4 days before draw (1 hour double, 3 hour none)	104%	103%
shift advertising from third week before draw to fourth week	103%	103%
shift advertising from fourth week before draw to third week	95%	95%

Notes: This table shows the effect of using alternative dynamic advertising strategies for the February draw. See text for a description of these strategies. In the column labeled "unchanged" consumer expectations are consistent with the advertising data we used to estimate the model and not with the changed advertising strategy. In the last column, we adjust expectations to reflect the change in the policy. Simulations are based on the parameter estimates reported in Table 2.

smaller. The intuition for this is that when consumers wrongly expect advertising activity to be lower at later points in time, then they already buy earlier and therefore the effect of changing the advertising strategy is bigger because they are reminded more often than they expected. Conversely, when there is no advertising anymore while consumers still expect to be reminded by it, then sales are lower. The same holds true when we shift advertising from the fourth to the third week.

To summarize, the results show that shifting advertising to later points in time leads to higher sales. Given the model structure and the fixed GRP budget, this is desirable for both, consumers and the firm.

8   Summary and concluding remarks

This paper uses high frequency advertising and sales data to measure the short run effects of advertising. The thought experiment we undertake for this is akin to a regression discontinuity design: we compare sales just before the advertisements are aired to sales thereafter. We find short-term advertising effects to be sizable and to last for about 4 hours. Besides, we make use of the fact that there is a given purchase cycle with a fixed deadline until which consumers can buy a ticket. Exploiting this we find that advertising does not only lead to purchase acceleration, but also to market expansion. Furthermore, advertising effects depend on the time until that deadline. The later the firm advertises, the higher the short term effect on sales. We develop the argument that this is novel evidence in favor of the view that advertising reminds consumers to consider buying a ticket. We then spell out a structural adoption model that can generate this dependence. We estimate the parameters of this model and simulate the effects of counterfactual dynamic advertising strategies on sales. Based on this we conclude that it is indeed likely that

Figure 9: Effect of different advertising strategies
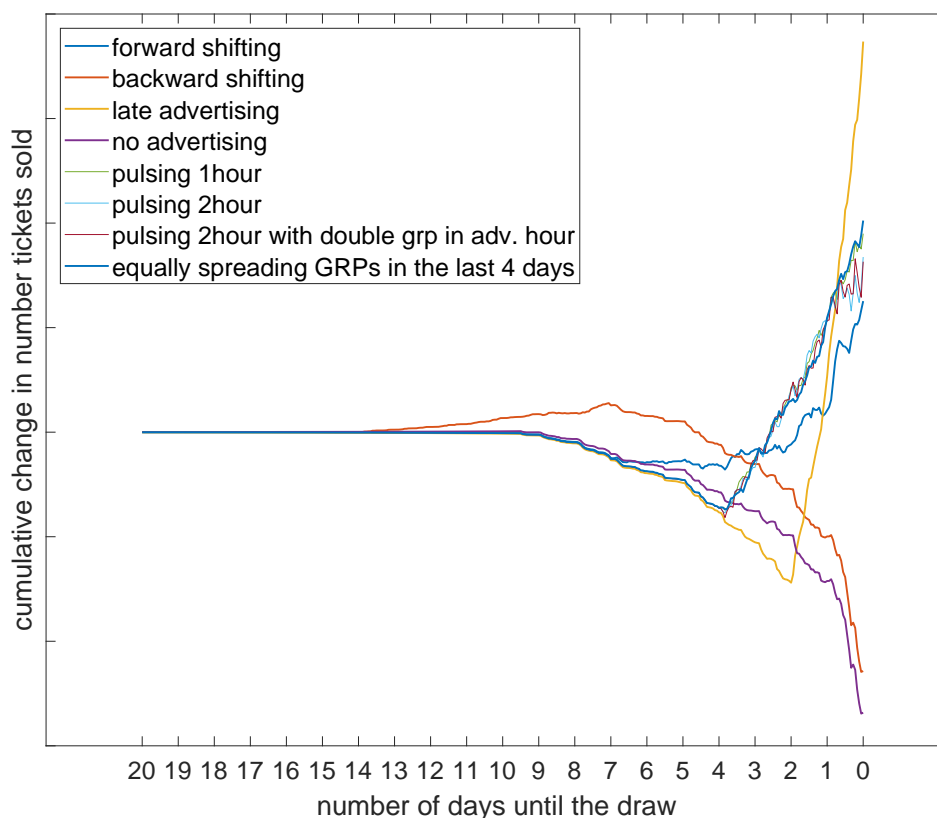
Notes: This figures shows the difference between the cumulative number of tickets sold at each point in time for a counterfactual advertising strategy and the cumulative number of tickets sold given the actual advertising schedule for the February draw. Consumer expectations are consistent with the respective actual or changed GRP schedule (column "rational" in Table 3). Based on parameter estimates reported in Table 2.

starting from the actual advertising schedule in the data and shifting advertising to later points in time has positive effects on sales.

The context of our study is the sale of lottery tickets for an upcoming draw. This context is particularly helpful for obtaining model-free evidence on the effects of advertising and structural estimates for the key model parameters, but advertisements can of course remind consumers also in other contexts. Examples include the purchase of durable goods and the adoption of technologies, in particular when there are natural deadlines such as Christmas or the end of the year, or deadlines set by the government. For instance, De Groote and Verboven (2016) study the adoption of solar panels. There is a deadline until which households have to buy in order to still be eligible for a tax subsidy. Our model could be used in such a situation to study how dynamic pricing and advertising strategies interact in a competitive environment, or which advertising strategy by the government complementing the subsidy scheme could be most effective. Other deadlines set by the government concern the decision of households to enroll into a savings plan or to change health insurance. Our model can be used to design an advertising strategy that most effectively reminds individuals to do so. One could also generalize it to study

29

the supply side interactions between firms to shed light on the question whether the possibility to remind consumers leads to increased competition and lower prices. Finally, our model could be extended to study the effects of present-bias and how they could most effectively be counteracted using information campaigns.

In the context of these examples one can imagine that advertisements that act as a reminder may be beneficial for consumers, for instance because they help individuals to do what they actually want to do or because they lead to increased competition through higher levels of awareness and consideration. This would be very much in line with Stigler's (1961) original point that the possibility to provide information by means of advertisements can lead to welfare improvements. His argument was based on the idea that consumers have limited information about the existence of products, their characteristics, or prices. We have presented novel evidence in line with the idea that advertisements can also help consumers who suffer from limited attention and forgetting.

References

Ackerberg, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 316–333.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination*. *International Economic Review 44*(3), 1007–1040.

Bagwell, K. (2007). The economic analysis of advertising. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, Volume 3, pp. 1701–1844. Elsevier.

De Groote, O. and F. Verboven (2016). Subsidies and myopia in technology adoption: Evidence from solar photovoltaic systems. Working Paper, KU Leuven, Belgium.

Draganska, M. and D. Klapper (2011). Choice set heterogeneity and the role of advertising: An analysis with micro and macro data. *Journal of Marketing Research 48*(4), 653–669.

Du, R. Y., L. Xu, and K. C. Wilbur (2017). Should TV advertisers maximize immediate online response? SSRN Working Paper 3037734.

Dubé, J.-P., G. J. Hitsch, and P. Manchanda (2005). An empirical model of advertising dynamics. *Quantitative Mareting and Economics 3*(2), 107–144.

Hu, Y., L. M. Lodish, and A. M. Krieger (2007). An analysis of real world TV advertising tests: A 15-year update. *Journal of Advertising Research 47*(3), 341.

Joo, M., K. C. Wilbur, and Y. Zhu (2015). Effects of TV advertising on keyword search. *International Journal of Research in Marketing*.

Kahneman, D. (1973). *Attention and effort*, Volume 1063. Prentice-Hall Englewood Cliffs, NJ.

Krugman, H. E. (1972). Why three exposures may be enough. *Journal of Adverting Research 12*(6), 11–14.

Letang, V. and L. Stillman (2016). MAGNA global advertising forecast. https://www.magnaglobal.com/wp-content/uploads/2016/12/MAGNA-December-Global-Forecast-Update-Press-Release-1.pdf.

Lewis, R. A. and J. M. Rao (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics 130*(4), 1941–1973.

Lewis, R. A. and D. H. Reiley (2013). Down-to-the-minute effects of super bowl advertising on online search behavior. In *Proceedings of the fourteenth ACM conference on Electronic commerce*, pp. 639–656. ACM.

Liaukonyte, J., T. Teixeira, and K. C. Wilbur (2015). Television advertising and online shopping. *Marketing Science 34*(3), 311–330.

Lodish, L. M., M. Abraham, S. Kalmenson, J. Livelsberger, B. Lubetkin, B. Richardson, and M. E. Stevens (1995). How TV advertising works: A meta-analysis of 389 real world split cable tv advertising experiments. *Journal of Marketing Research*, 125–139.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica 57*(5), 995–1026.

Melnikov, O. (2013). Demand for differentiated durable products: The case of the us computer printer market. *Economic Inquiry 51*(2), 1277–1298.

Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2112–2245. Amsterdam: North Holland.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*(5), 999–1033.

Sovinsky Goeree, M. (2008). Limited information and advertising in the US personal computer industry. *Econometrica 76*(5), 1017–1074.

Stephens-Davidowitz, S., H. Varian, and M. D. Smith (2017). Super returns to super bowl ads? *Quantitative Marketing and Economics 15*(1), 1–28.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy 69*(3), 213–225.

Online Appendix

A Details on the econometric implementation

In Section 5.6, we have given an overview over the estimation procedure. In this section, we provide further details.

A.1 Empirical setup

The data contain information on ticket sales and advertising activities for 16 draws. Since we collapse these data during the night, every day in the model has 18 hours. The starting period is 00:00-00:59 on Jan 1 and the last period is 17:00-17:59 on Dec 31. Thus, the total number of periods is $\tau = 6,564$ (τ is not to be confused with T, which we have defined in the context of our model). We divide them up into sub-periods, one for each draw. We account for the fact that they differ with respect to the total number of hours (T in the model) and the value to holding a ticket (ψ in the model), and of course with respect to the realized advertising activity.[1] The ticket price is constant over time and across draws.

A.2 Method of simulated moments

The set of structural parameters that do not change across draws is $\{\lambda, \sigma, \delta, \gamma\}$. In addition, we estimate 16 values $\psi_1, \ldots, \psi_{16}$ to holding a ticket at the time of the draw. Thus the full set of structural parameters to be estimated is $\theta \equiv \{\lambda, \sigma, \delta, \gamma, \psi_1, \ldots, \psi_{16}\}$.

Recall that we only have access to aggregate data. Let $\hat{u}_t(\theta) \equiv q_t - \tilde{q}_t(\theta)$ be the difference between actual aggregate demand q_t in the data, divided by 250, and the model prediction $\tilde{q}_t(\theta)$ as described in Section 5.6. In Section A.3 below we will specify a set of moments $\mathbb{E}[m(z_t, \hat{u}_t(\theta))] = 0$, where z_t is a vector of exogenous variables constructed from the data so that the left hand side is a column vector and the right hand side is a vector of zeros and the expectation is taken over hours. The (technical) condition for identification is that they hold if, and only if, we evaluate the function m at the true parameters θ (see for instance Newey and McFadden, 1994).

Let $\bar{m}(\tilde{\theta})$ be the average of $m(z_t, \hat{u}_t(\theta))$, over time in hours across all draws (thus over τ time periods), evaluated at any candidate parameter vector $\tilde{\theta}$. The MSM estimator is

$$\hat{\theta} = \arg\min_{\tilde{\theta}} \bar{m}(\tilde{\theta})' W \bar{m}(\tilde{\theta}),$$

where W is a positive definite weighting matrix.

Under the assumption mentioned above, $\hat{\theta}$ is consistent. An estimator of the variance-covariance matrix is given by (Newey and McFadden, 1994)

---

[1]This means that T and ψ need to indexed by the draw, because they differ across draws. For the ease of the exposition, in Section 5, we have described the model only for one draw. Within each draw, t runs from 1 to the draw-specific T.

$$\widehat{var(\hat{\theta})} = \frac{1}{\tau}(A'WA)^{-1}B(A'WA)^{-1},$$

where

$$A = \frac{\partial \bar{m}(\hat{\theta})}{\partial \hat{\theta}'}$$

and

$$B = A'W(m(\hat{\theta}) - \bar{m}(\hat{\theta}))(m(\hat{\theta}) - \bar{m}(\hat{\theta}))'WA.$$

A.3 Moments and weighting matrix

z_t contains 3 sets of exogenous variables: a full set of dummy variables for the number of days until the draw, the number of GRP's in t, $t-1$, $t-2$, and $t-3$, and variables calculating cumulative sales up to point t. This means that we attempt to pick the parameters so that the model captures well the evolution of sales over time and the reaction to advertisements.

Specifically, we stack all $\hat{u}_t(\theta)$ into a vector $\hat{u}(\theta)$ of dimension $\tau \times 1$ and define a $\tau \times M$ matrix of exogenous variables $Z = (\mathbf{1}, Z_1, Z_2, Z_3)$, where $\mathbf{1}$ is a vector of ones, Z_1 contains times until draw dummies in the columns, Z_2 contains GRP's and lags thereof in the columns, and Z_3 is a matrix with indicators such that it takes cumulative sales at the daily level, separately for each draw. Specifically, Z_3 is block-diagonal with sub-matrices $Z_{3,r}$ on the diagonal (r indexing draws). Each column of these sub-matrices is for one day and contain a set of ones on top and zeros in the bottom, such that the cumulative prediction error is calculated on a daily level when we multiply Z_3' with $\hat{u}_t(\theta)$.

After eliminating linearly dependent columns, Z has $M = 376$ columns, meaning that we have 376 exogenous variables.[2] Using this, we calculate

$$\bar{m}(\tilde{\theta}) = \frac{1}{\tau}Z'\hat{u}(\tilde{\theta}).$$

We choose the weighting matrix W to be

$$W = (Z'Z/\tau)^{-1}.$$

A.4 Smoothing

When using a simulation-based procedure to estimate a discrete choice model, one common challenge is that the simulated choice probabilities (if one use simulated maximum likelihood), or, in our case, simulated demand, is not a smooth function in the parameters. This is due to the

---

[2]Z_1 originally has 30 columns. Z_2 contains GRP's and 3 lags thereof, so it has 4 columns. Z_3 has 365 columns. Most columns in Z_1 are linear combinations of columns in Z_3. After dropping those, Z_1 has 7 columns left. Thus, we have in total $1 + 7 + 4 + 364 = 376$ columns.

fact that in discrete choice models individuals are assumed to either choose to buy or not at a given point in time. Consequently, for each simulated consumer, small changes in parameters will either have no effect on his decision (which stays at 0 or 1), or change his decision from 0 to 1. Such non-smoothness can lead to problems with the usual methods for finding an optimum of the objective function because of flat spots.

In principle, this could be addressed by increasing the number of simulated consumers. But it is not possible to fully overcome it, as the number of simulated consumers will stay finite. Therefore, as an alternative, we use a smoothed accept-reject simulator to make the demand function fully smooth in the parameters. We use this very conservatively, however, and only to avoid that the estimator gets stuck on a flat spot.

Following McFadden (1989), the simulator that we choose has the logit form. Instead of generating choices for individual i in t that are either 0 or 1, we generate smoothed choices

$$\tilde{S}_{it} = 1 - \frac{\exp\left(\frac{u_{it} - \tilde{P}_{it}(\tilde{g}_t^a)}{s}\right)}{1 + \exp\left(\frac{u_{it} - \tilde{P}_{it}(\tilde{g}_{it}^a)}{s}\right)},$$

where $\tilde{P}_{it}(\tilde{g}_t^a)$ is the simulated probability to buy given considering, u_{it} is a random draw from the standard uniform distribution and s is the smoothing parameter. The higher s the more smoothing there is. In our case, it is sufficient to use very little smoothing. We specify $s = 0.00015$.[3]

B Robustness

In this appendix, we assess how robust our results are to assuming a different market size (Appendix B.1) and to allowing for serial correlation in viewership behavior (Appendix B.2).

B.1 Assumption on market size

We first assess the robustness to making alternative assumptions about the market size. For this, we re-estimate the model assuming a market size of 500,000. We expect this to mainly affect the our parameter estimates related to the baseline probability of considering to buy a ticket and for the value to holding a ticket.

---

[3]We have experimented with different values of s and the result is not sensitive to the choice of s for values of s around 0.00015.

Table 4: Robustness checks: parameter estimates when we double the market size and allow for serially correlated viewership

parameter	baseline specification	ste.	double market size	ste.	generalized model	ste.
one minus depreciation rate goodwill stock (λ)	0.334	0.197	0.404	0.097	0.193	0.006
effect of goodwill stock on probability of considering (γ_1)	1.637	0.731	2.041	0.199	2.085	0.362
intercept of stock on probability of considering (γ_0)	-0.430	0.329	-2.270	0.118	-1.086	0.007
hourly discount factor (δ)	0.994	0.000	0.999	0.000	0.994	0.000
multiplying factor taste shock (σ)	0.318	0.010	0.109	0.024	0.339	0.006
value to having a ticket on the day of the draw						
10 January, 2014	1.477	0.056	2.513	0.119	1.496	0.046
10 February, 2014	1.669	0.039	2.575	0.106	1.638	0.053
10 March, 2014	1.493	0.053	2.507	0.121	1.505	0.054
10 April, 2014	1.448	0.067	2.490	0.125	1.489	0.058
26 April, 2014 (King's Day)	1.906	0.052	2.650	0.088	2.040	0.039
10 May, 2014	1.620	0.046	2.542	0.112	1.609	0.052
10 June, 2014	1.711	0.049	2.556	0.110	1.762	0.040
24 June, 2014 (Orange draw)	1.728	0.046	2.597	0.101	1.789	0.033
10 July, 2014	1.887	0.058	2.614	0.099	1.914	0.046
10 August, 2014	1.410	0.060	2.496	0.123	1.388	0.060
10 September, 2014	1.644	0.053	2.559	0.109	1.656	0.047
1 October, 2014 (special 1 October draw)	1.608	0.048	2.604	0.099	1.666	0.040
10 October, 2014	1.541	0.057	2.513	0.120	1.574	0.046
10 November, 2014	1.510	0.051	2.504	0.121	1.593	0.051
10 December, 2014	1.749	0.045	2.576	0.106	1.824	0.037
31 December, 2014 (New year's eve draw)	2.336	0.066	2.857	0.045	2.421	0.031

Notes: Structural estimates. See Section 5.6 and Appendix A for details on the estimation procedure. Estimates for the baseline specification are repeated in the first column (see Table 2). The second set of parameter estimates was obtained under the assumption that the market size is 500,000 instead of 250,000. The third set of estimates is for the generalized model with serially correlated viewership behavior described in Appendix B.2. The probability to consider buying is specified as $P_{it}(\text{consider}) = 1/\left(1 + \exp\left(-\left(\gamma_0 + \gamma_1 g_{it}^a\right)\right)\right)$.

Table 4 shows that the probability to consider buying a ticket is estimated to be lower, while the value to holding a ticket is estimated to be higher. In combination, this produces choice probabilities that are roughly half as big as for our baseline model with a market size that is half as big. In addition, the estimate of the hourly discount factor is higher and the estimated standard deviation of the taste shock is lower. The parameters that are least affected are the depreciation rate of the advertising goodwill stock and the coefficient on the goodwill stock that measures by how much it influences the probability of considering. In fact, the percentage point increase in that probability when we change the goodwill stock from zero to one is very similar across those two specifications.

B.2 A model with serially correlated viewership

So far, we have assumed that the probability that a consumer i is reached in t by an advertisement is given by the number of GRPs. Implicitly, this assumes that reaching a consumer in t is independent of reaching the same consumer in another period t', for instance $t - 1$. This can only be the case if viewership behavior is not serially correlated.

While this is likely violated at the minute level, it may be a reasonable approximation at the hourly level at which we estimate our model. We have no data to directly quantify how likely it is that the same consumer is reached when there are advertisements in two consecutive hours. Therefore, we assess whether this assumption substantially affects our estimates and the main conclusions we draw from them by extending our model to allow for serial correlation in viewership behavior.

In this extended version of the model, there are two states for each consumer: watching TV or listening to the radio and not watching TV or listening to the radio. When estimating the model we proceed in two steps. We first simulate, for each individual, whether they are watching TV or are listening to the radio in a given time period. Then we impose that advertising can only reach those consumers who are actually watching TV or are listening to the radio. At the same time, we assume that consumer expectations are still reasonably approximated by (2). The reason for this is that modeling consumer expectations would involve introducing an additional state variable.[4]

Formally, let state $k = 1$ be the state of not watching or listening and state $k = 2$ the one of watching TV or is listening to the radio. Specify a 2-by-2 Markov transition matrix

$$\Pi = \left[\begin{array}{cc} 0.8 & 0.2 \\ 0.4 & 0.6 \end{array} \right].$$

This means that if an individual is watching TV or is listening to the radio at time t, then there will be 40% chance that she will stop watching and 60% chance that she will continue watching

---

[4]We do not expect this to have a big effect on our parameter estimates. In Table (3), we have seen that expectations have a relatively small effect on predictions under counterfactual advertising schedules.

in period $t+1$. From this we compute the implied 2-by-1 vector P^∞ of stationary probabilities. Then, we use P^∞ to simulate individual viewership demand in the first period and Π to simulate paths in subsequent periods.

Note that here, we treat the transition probabilities as known. We could estimate them if we had data at the consumer level.

Table 4 shows the estimation result for this more general model. Now, the same pattern in the data needs to be rationalized by a model in which viewership demand is serially correlated. The results show that this can be achieved by a lower depreciation rate of the goodwill stock and a higher effect of the goodwill stock on the probability of buying a ticket. The remaining parameters are almost not affected.

C   Additional tables and figures

Table 5: Differences across draws

	(1) all draws	(2) regular draws	(3) special draws	(4) all draws
log jackpot size	0.366*	0.366***		0.509*
	(0.178)	(0.106)		(0.251)
special draw	1.509***			2.107**
	(0.492)			(0.633)
log number of days	0.182	0.153	0.805	0.408
	(0.174)	(0.106)	(1.657)	(0.558)
log jackpot size previous draw				-0.245
				(0.297)
special draw in previous draw				-0.107
				(0.969)
log number GRP previous draw				0.954
				(0.583)
Observations	16	12	4	15
R^2	0.562	0.605	0.106	0.727

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows the results of a regression of the log of total sales on the total number of days on which tickets could be bought and the jackpot size if the draw was regular. In column (1) and (4) we pool across regular and special draws and set the log of the jackpot size to zero for the latter. One observation is one draw. There are only 15 observations for the last specification because we lack data on the previous draw for the first one that is in our data.

Table 6: Effect of TV and radio advertising

	(1) TV and radio
TV GRP between 0 and 4 minutes ago	0.0133***
	(0.00118)
5 and 9 minutes	0.0415***
	(0.00140)
10 and 14 minutes	0.0453***
	(0.00123)
15 and 19 minutes	0.0308***
	(0.00120)
20 and 24 minutes	0.0206***
	(0.00108)
25 and 29 minutes	0.0144***
	(0.00113)
0.5 and 1 hour	0.0116***
	(0.000446)
1 and 1.5 hours	0.00819***
	(0.000462)
1.5 and 2 hours	0.00412***
	(0.000436)
2 and 2.5 hours	-0.000105
	(0.000408)
2.5 and 3 hours	-0.00567***
	(0.000392)
3 and 3.5 hours	-0.0112***
	(0.000374)
3.5 and 4 hours	-0.0190***
	(0.000396)
radio GRP between 0 and 4 minutes ago	0.00102
	(0.00143)
5 and 9 minutes	0.00544***

	(0.00146)
10 and 14 minutes	0.00617***
	(0.00147)
15 and 19 minutes	0.00474**
	(0.00146)
20 and 24 minutes	0.00818***
	(0.00145)
25 and 29 minutes	0.00981***
	(0.00143)
0.5 and 1 hour	0.00881***
	(0.000650)
1 and 1.5 hours	0.0116***
	(0.000652)
1.5 and 2 hours	0.0149***
	(0.000661)
2 and 2.5 hours	0.0133***
	(0.000655)
2.5 and 3 hours	0.0106***
	(0.000651)
3 and 3.5 hours	0.0110***
	(0.000632)
3.5 and 4 hours	0.00883***
	(0.000646)
draw dummies	Yes
days to draw dummies	Yes
hour dummies	Yes
Observations	515205
R^2	0.632

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows the results of a regression of the log of one plus sales on GRP's of advertising, separately for TV and radio advertising, and lags thereof. Table 1 shows effects when we pool TV and radio advertising together. Regressions were carried out at the minute level and standard errors are robust to heteroskedasticity.

Table 7: Evidence from a distributed lag model at the hourly level

	(1) baseline	(2) TV and radio	(3) before last week	(4) last week	(5) no controls
GRP current hour	0.0120*** (0.00145)		0.0110*** (0.00238)	0.00406* (0.00163)	0.0779*** (0.00536)
GRP 1 hour lagged	0.0120*** (0.00133)		0.0121*** (0.00245)	0.00529*** (0.00156)	0.0431*** (0.00533)
GRP 2 hours lagged	0.00428** (0.00134)		0.00457 (0.00235)	0.000309 (0.00147)	0.0277*** (0.00677)
GRP 3 hours lagged	0.00412* (0.00168)		0.00369 (0.00306)	-0.000451 (0.00200)	0.0545*** (0.00654)
GRP TV current hour		0.0128*** (0.00185)			
GRP TV 1 hour lagged		0.0133*** (0.00168)			
GRP TV 2 hours lagged		0.00329* (0.00163)			
GRP TV 3 hours lagged		0.00154 (0.00246)			
GRP radio current hour		0.00758*** (0.00214)			
GRP 1 hour lagged		0.00838*** (0.00218)			
GRP 2 hours lagged		0.00775** (0.00239)			
GRP 3 hours lagged		0.00802** (0.00246)			
draw dummies	Yes	Yes	Yes	Yes	No
days to draw dummies	Yes	Yes	Yes	Yes	No
hour dummies	Yes	Yes	Yes	Yes	No
Observations	7662	7662	5406	2256	7662
R^2	0.917	0.917	0.858	0.942	0.263

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: This table shows the results of regressions of the log of one plus sales on GRP's of advertising and lags thereof. In column (2) we distinguish between TV and radio advertising. Regressions were carried out at the hourly level and standard errors are robust to heteroskedasticity.

Figure 10: Cumulative sales for remaining draws
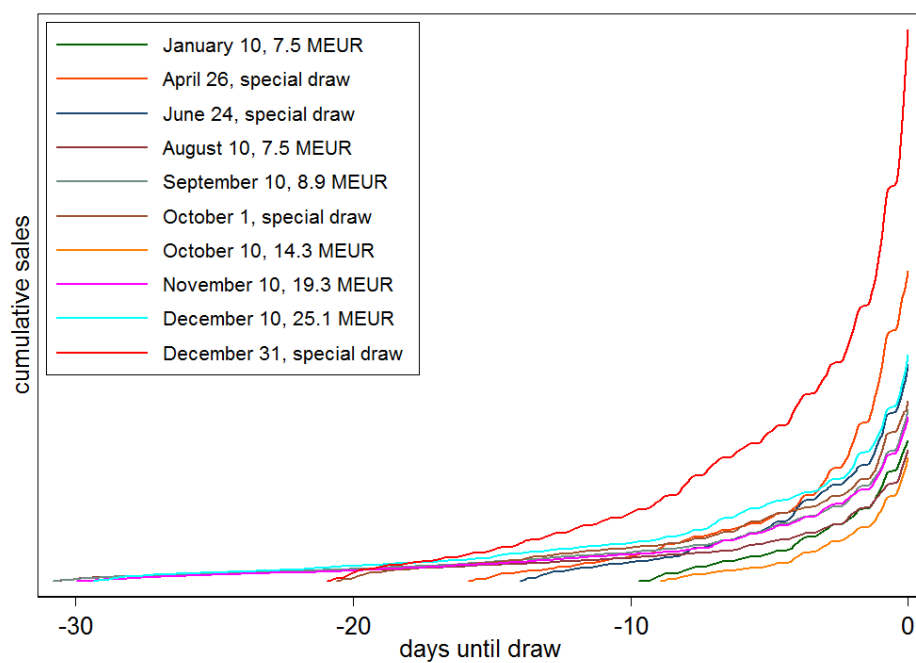
Notes: Figure 1 shows cumulative sales for 6 selected regular draws. This figure shows them for the remaining draws.

Figure 11: Advertising and sales during the day of the draw
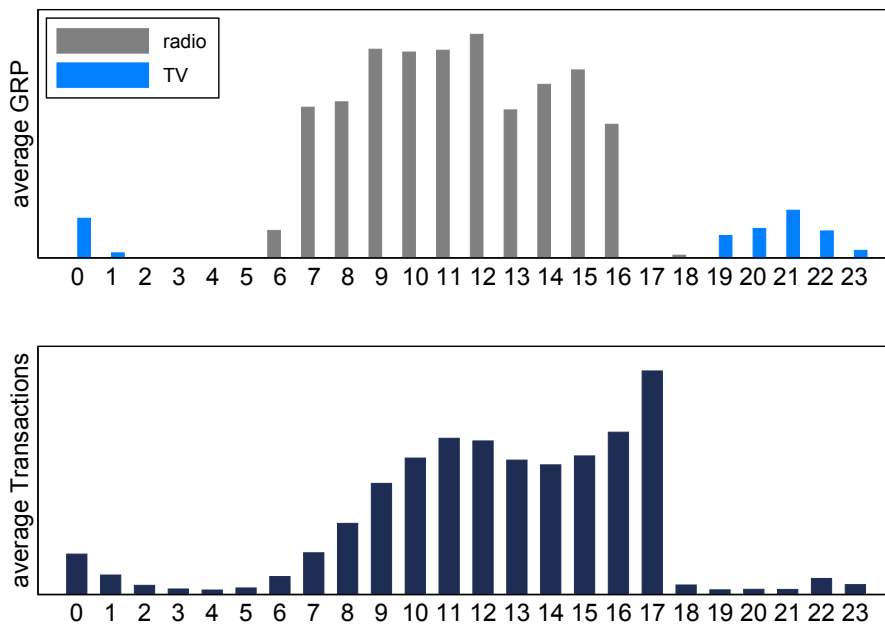

Notes: This figure shows average GRP's and sales for different times of the day. To produce this figure we first aggregate sales at the hourly level and then average over draws. On the day of the draw tickets for this draw can only be bought until 6pm. See Figure 2 for the pattern on the remaining days.

Figure 12: GRP's at the minute-level for a special draw
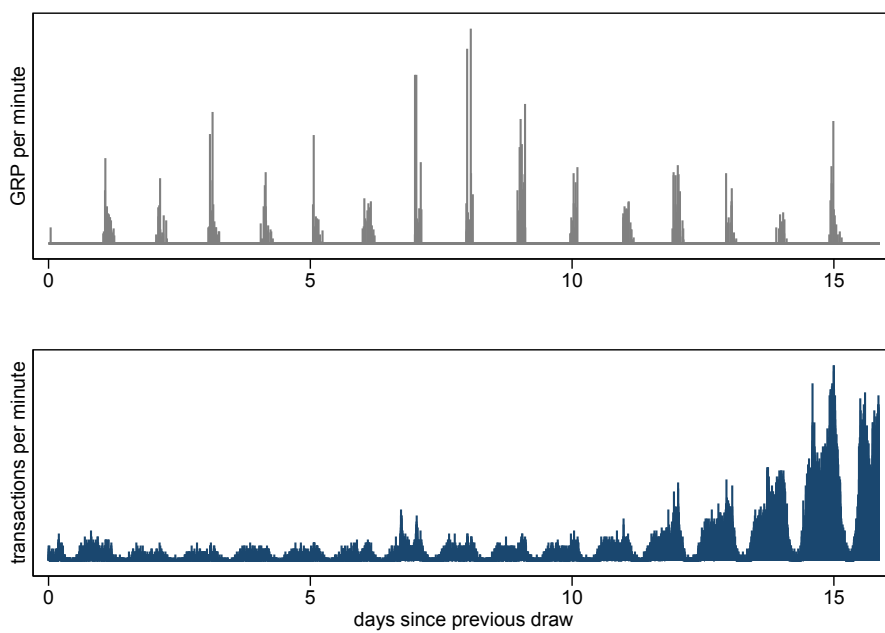
Notes: Figure 3 shows GRP's and sales for the draw on April 10, 2014. This figure shows GRP's and sales at the minute level for the special draw on April 26, 2014 (King's Day). The last regular draw took place on April 10, 2014. Tickets for the next draw can be bought from 6pm on the day of the previous draw, which is depicted as 0 days since the previous draw.

Figure 13: Advertisements that were used to construct Figure 5
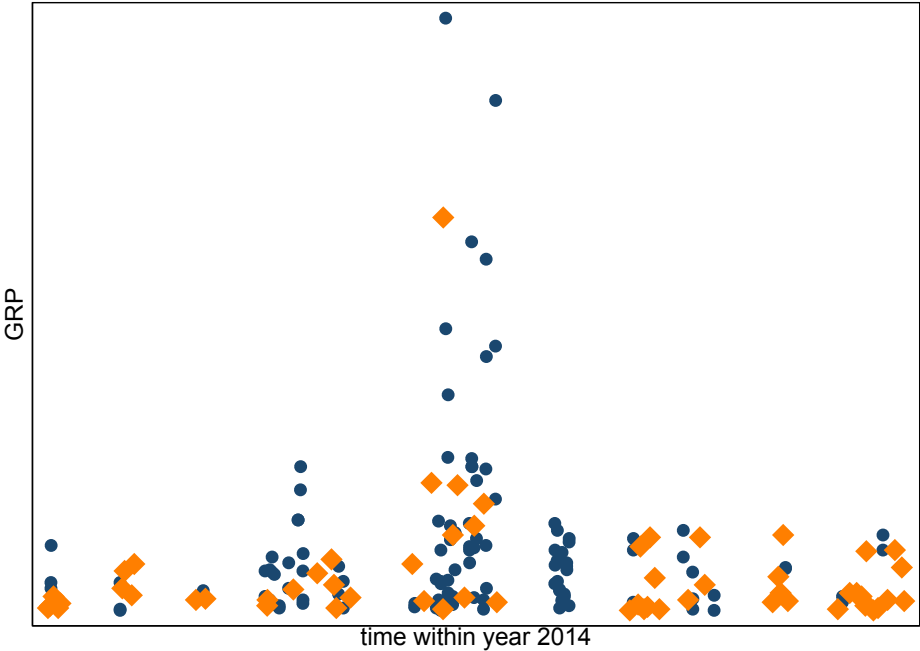
Notes: This figure shows which advertisements were used in the sample for Figure 5. It shows a dot for each advertisement with at least 9 GRP, with the number of GRP's plotted against time. The diamonds are the advertisements that were used.

Figure 14: Expectations
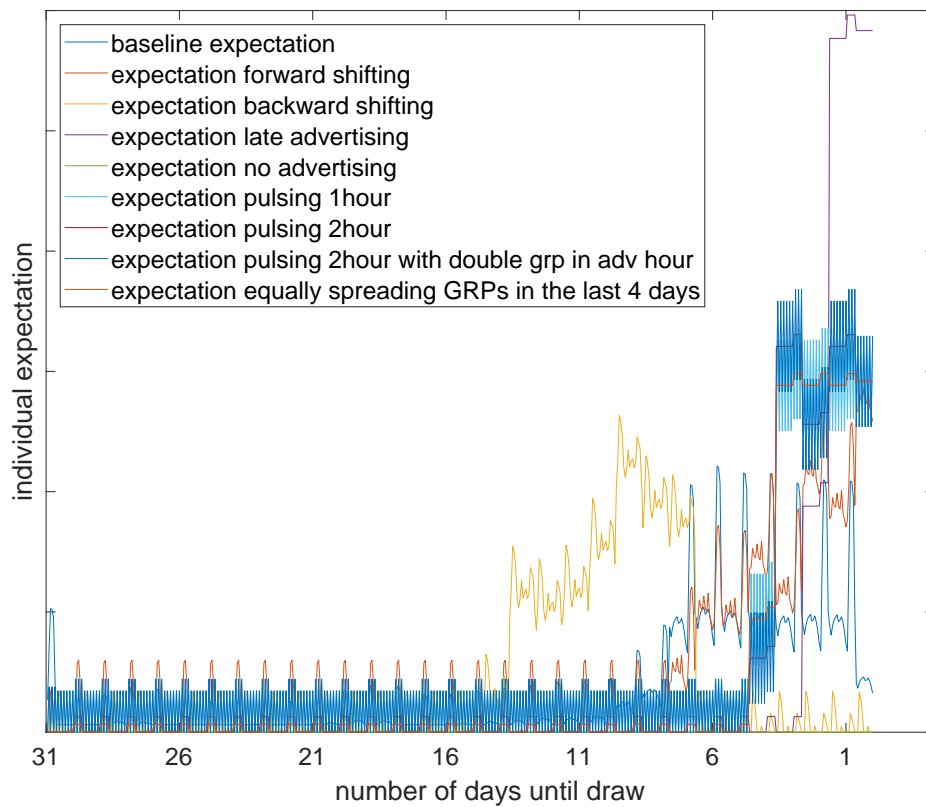
Notes: Figure shows the expected probability to see an advertisement, from the individual perspective. Obtained from regression of GRP's on hours of a day dummies, days until draw dummies for the draw in February.