创业与管理学院
School of Entrepreneurship and Management

SHANGHAITECH SEM WORKING PAPER SERIES
No. 2020-001

Advertising as a reminder: Evidence from the Dutch State Lottery

Chen He
ShanghaiTech University
Tobias J. Klein
C.E.P.R., TILEC, CentER,
Tilburg University

Janurary 10, 2020

https://ssrn.com/abstract=3517063

School of Entrepreneurship and Management
ShanghaiTech Universityhttp://sem.shanghaitech.edu.cn

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

Chen He and Tobias J. Klein†

August 2019

Abstract

Consumers who intend to buy a product may forget to do so. Therefore, they may value being reminded by an advertisement. This phenomenon could be important in many markets, but is usually difficult to document. We study it in the context of buying a product that has existed for almost 300 years: a ticket for the Dutch State Lottery. This context is particularly suitable for our analysis, because the product is simple, it is very well-known, and there are multiple fixed and known purchase cycles per year. Moreover, TV and radio advertisements are designed to explicitly remind consumers to buy a lottery ticket before the draw. This can conveniently be done online. We use high frequency advertising and online sales data to measure the effects of TV and radio advertising. We show that advertising effects are short-lived and the bigger the less time there is until the draw. This is consistent with the predictions of a simple model in which consumers suffer from limited attention and advertising affects the probability that consumers think about buying a lottery ticket and otherwise value buying it as late as possible. We provide direct evidence that advertising does not only affect the timing of purchases, but also leads to market expansion. Finally, we estimate a dynamic structural model of consumer behavior and simulate the effects of a number of counterfactual dynamic advertising strategies. We find that total sales would be 35 percent lower without advertising and that shifting advertising to the week of the draw would lead to a 16 percent increase in sales. This means that consumers react strongly to reminder advertising and wish to be reminded late, when their intention to buy is higher.

Key words: Reminder advertising, limited attention, adoption model.

JEL-classification: M37, D12, D83.

---

*We are grateful to Blue Mango Interactive and the Dutch State Lottery for providing us with the data. We thank Menno Zevenbergen and Pieter van Geel for stimulating discussions and useful comments. Without them, this project would not have been possible. The views expressed in this paper, however, do not necessarily reflect those of Bluemango or the Dutch State Lottery. We would also like to thank Jean-Pierre Dubé, Ulrich Doraszelski, Eric French, Ulrich Kaiser, Carl Mela, Peter Newberry, Martin Peitz, Pasquale Schiraldi, Stephan Seiler, Andrey Simonov, Rik Pieters, Frank Verboven, Ken Wilbur, Joachim Winter, and participants of a seminar in Oxford, at eBay in San Jose, the 2015 Media Economics Workshop in Stellenbosch, the 2016 meeting of the IO group of the Verein für Socialpolitik, the 2016 Digital Marketing Conference at Stanford, the 2017 C.E.P.R. Conference on Applied Industrial Organization, the 2018 EEA/ESEM in Cologne, and the 2019 Choice Symposium in Chesapeake Bay for their for comments and suggestions.

†He: ShanghaiTech University; e-mail: hechen@shanghaitech.edu.cn. Klein: C.E.P.R., TILEC, CentER, Tilburg University; e-mail: T.J.Klein@uvt.nl.

1 Introduction

Models of consumer behavior usually assert that not making a purchase is a deliberate choice. However, in practice, this could be driven by limited attention rather than preference. In this situation, consumers may value being reminded by an advertisement.

Although potentially important in many markets ranging from markets for consumer packaged goods to markets for health insurance, little is known about reminder advertising. One of the reasons for this is that it is generally difficult to isolate its effects and to measure how exactly it influences consumer behavior. Important questions in this context are whether and to what extent the timing of advertising matters, and whether reminder advertising mainly affects the timing of purchases or also the probability to buy at all.

In this paper, we study reminder advertising in a context of buying a product that has existed for almost 300 years: a ticket for the Dutch State Lottery. This context is particularly suitable for our analysis, because the product is simple, it is very well-known, there are known purchase cycles, and advertisements explicitly reminded consumers to buy a ticket. We develop two key empirical predictions related to advertising acting as a reminder—namely (1) that advertising effects are short-lived in our context because consumers suffer from limited attention that fades away quickly, and (2) that advertising effects are the bigger the less time there is until the draw because the likelihood that consumers buy once reminded is higher at later points in time. Then, 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.

We find the short run effects of advertising to be sizable. Reaching 1 percent of the population leads to an average increase of sales by about 2 percent within one hour. Advertising effects last up to about 2 hours and are the bigger the less time there is until the draw. This is consistent with our two empirical predictions.

Before the deadline, advertising leads in general to both purchase acceleration (individuals buying earlier rather than later) and market expansion (more people buying in total). We provide direct evidence that advertising does not only affect the timing of purchases, but also leads to market expansion, by showing that advertising has a short-run effect until the end of the period in which tickets can be bought.[1]

After presenting this reduced form evidence we develop and estimate a tractable structural model of consumer behavior to measure the effect of advertising on total sales. Consumers, at a given point in time, either buy a ticket for the following draw, or postpone 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 an attention stage. The probability to

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

devote attention to thinking about 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. Total sales do not only depend on the total amount of advertising, but on the timing of all advertisements. Our counterfactual simulations suggest that total sales would be 35 percent lower without advertising. Shifting advertising to the week of the draw would lead to a 16 percent increase in sales.

Our results are specific to the context of buying lottery tickets in the Netherlands. This context and the empirical setup are particularly well-suited to isolate and quantify the effects of reminder advertising. In other contexts, advertising may at the same time convey information, act as a reminder, and change consumer preferences. Specifically, since in our case the product has existed for a long time and consumers are well-aware of it, advertising is unlikely to change preferences or convey information about the existence or characteristics of the product. The exception to this is the size of the jackpot. Baseline sales clearly depend on it, which is why throughout we carefully control for differences across draws and use only within draw variation in the data to estimate advertising effects. In addition to that, we show in Section 6.4 that short-run advertising effects did not depend on whether the information on the size of the jackpot was good or bad. We also provide a detailed explanation for this finding based on the institutional context. By focusing on a specific context we can show that reminder advertising can have big effects. With this, we hope to stimulate more work on the topic. We discuss interesting directions in the concluding Section 10.

To the best of our knowledge, the idea that advertising may serve as a reminder is not prominently featured in the academic literature. One exception is Krugman (1972), who asked the question how often consumers should be reached by advertising and then 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.

Reminder advertising can be seen as a generalization of the concept of purchase facilitation. Originally, Rossiter and Percy (1987) described purchase facilitation as providing information to individuals who intended to buy a product, for instance about the closest retailer at which a product can be bought and how one can pay for it. Purchase facilitation resembles reminder advertising in the sense that it helps the consumer to turn the intent to buy a product into an actual purchase. However, reminder advertising overcomes a different challenge than purchase facilitation, as it reminds individuals of their purchase intent without providing new information.

The rest of this paper is structured as follows. Next, in Section 2, we relate our paper to the literature. Section 3 gives a brief overview over the market for lottery tickets in the Netherlands. Section 4 describes the data. In Section 5, we provide a precise definition of reminder advertising and develop our two empirical predictions. Section 6 tests these empirical predictions and quantifies the effects of reminder advertising. Section 7 develops our dynamic model of lottery ticket demand with advertising effects. Section 8 presents the results. Section 9 performs counterfactual experiments, and Section 10 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 Literature and contribution

The focus of our paper is on one particular role of advertising, namely to remind consumers to act on their preference. This is closely related to the literature that studies the effects of inattention and information treatments.[2]

A first set of papers in that literature studies low observed rates of switching between providers of a service (inertia) and is interested in understanding the source of this and whether there is room for policy interventions. Hortaçsu et al. (2017) do so in the context of the residential electricity market. Their hypothetical counterfactual experiments show that low-cost information interventions can increase consumer surplus. Heiss et al. (2016) and Ho et al. (2017) study inertia in the context of Medicare Part D plan choice. Heiss et al. (2016) find that removing inattention has larger effects on switching behavior than removing switching costs. Ho et al. (2017) also model the supply side and find that reducing inertia would reduce the amount consumers spend by 24 percent, due to a combination of demand and supply side changes. Kiss (2017) exploits a natural experiment and finds that advertising affected switching behavior for auto liability insurance. These papers have in common that limited attention plays a central role in explaining observed behavior and that information treatments have large effects. The focus in this literature has so far not been on how exactly this information could be provided. We therefore contribute to this literature by quantifying the effects of reminder advertising using high frequency data in a context that is particularly suitable to focus on this aspect. This allows us to better understand how consumers make decisions, as we model the dynamic process through which they end up buying a product or not, and how this is influenced by the exact timing of advertising. A prediction that comes out of our analysis and that will be interesting to test empirically in other contexts is whether information treatments are most effective at those points in time at which the inclination of consumers to buy a product when thinking about it is highest—which in our case is shortly before the draw.

A second set of papers studies the effects of reminders, typically by conducting field experiments. For instance, Calzolari and Nardotto (2016) find substantial immediate effects of simple reminders sent by email on gym attendance. They carefully develop the argument that this can be explained by limited attention. We contribute to this literature by showing that reminders can also be provided in the form of advertisements. This is particularly useful when potential clients have previously not been in touch with a provider of a service or a seller, or have not

[2]There is also a large, broader literature on inattention. See Gabaix (2018) for an excellent survey.

signed up for reminders.[3]

Next to this, we contribute to the literature on advertising effects.

More and more papers use high frequency data and estimate the effect of TV advertising on behavior online, as we do. However, they mostly document effects on search and not transactions. Zigmond and Stipp (2010) find large post-ad spikes in Google search for the advertised brands following TV ads during the opening ceremonies of the 2008 and 2010 Olympic Games. Lewis and Reiley (2013) study the effects of 46 Super Bowl commercials on online search behavior. They find large spikes in search behavior related to the advertiser or product within 15 seconds following the conclusion of the TV commercial. Kitts et al. (2014) show that effects are also measurable for smaller advertising campaigns. Joo et al. (2014) find that advertising for financial services brands increases both the number of related Google searches and searchers' tendency to use branded keywords in place of generic keywords. Joo et al. (2016) present similar findings using individual-level data at the hourly level. Chandrasekaran et al. (2018) also show that Super Bowl advertising affects brand search. Du et al. (2019) study advertising for pickup trucks. They find that the elasticity of brand search with respect to advertising is 0.09, while the price search elasticity is 0.03. Both Liaukonyte et al. (2015) and our paper complement these papers with evidence from high-frequency advertising and sales (as opposed to search) data.

More broadly, our paper relates and contributes to the large literature on the estimation of TV advertising effects.[4] Lodish et al. (1995) summarize the earlier literature on 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. More recent contributions include Stephens-Davidowitz et al. (2017) who 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; Shapiro (2018) who exploits the discontinuity of advertising exposure at the borders of television markets to show that advertising does not only increase a firm's own sales, but also the sales of rivals; and Shapiro et al. (2019) who study the distribution of advertising elasticities for a large number of products.

Another strand of the advertising literature is concerned with characterizing the mechanism through which advertising affects consumers.[5] Ackerberg (2001, 2003) develops an approach to empirically distinguish between advertisements being effective because they are informative *vis-à-vis* them being effective because they are transformative and increase the valuation for the brand. Using scanner data he finds mainly support of former. Mehta et al. (2008) also use scanner data and find evidence for both informative and transformative effects of advertis-

[3] An open question in this literature is whether consumers are forward-looking enough to sign up for a reminder service.

[4] Besides, there is a smaller, but fast growing literature on the estimation of online advertising effects. See for instance Goldfarb (2014), Lewis and Rao (2015), Blake et al. (2015), Johnson et al. (2017), Gordon et al. (2018), and the references therein.

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

ing, and that the informative effects become less important over time, consistent with learning. Hartmann and Klapper (2017) exploit exogenous variation in exposure to advertisements in the context of the Super Bowl and reject the theory that in this context advertising works by serving as a complement to brand consumption. They develop the argument that advertising instead builds a complementarity between the brand and sports viewership more broadly. We contribute to this literature by presenting evidence in favor of the view that advertising can also act as a reminder.

Yet another set of papers aims at shedding light on the effect of advertising on product consideration. Roberts and Lattin (1991, 1997) summarize the early literature on consideration. The resulting consideration set is determined by the consumer. Sovinsky Goeree (2008) and Draganska and Klapper (2011) estimate static models with a consideration stage. Van Nierop et al. (2010) and Abaluck and Adams (2017) discuss the more recent literature and demonstrate that consideration sets can be inferred from choice data. Terui et al. (2011) use scanner data and find that strong support for advertising effects on choice through an indirect route of consideration set formation that does not directly affect brand utility. Manzini and Mariotti (2014) study choice by consumers who consider an alternative with a given probability and call this limited attention, as opposed to consideration. They show that both the preference relation and the attention parameters are identified uniquely from stochastic choice data. In our model, consumers also suffer from limited attention. The probability to be attentive and actively think about buying depends on advertising and varies over time. This can be thought of in terms of consideration sets that vary over time. In that sense, we contribute to the consideration set literature by showing that advertisements that act as a reminder can have strong effects on consideration that are however short-lived. The consideration set literature abstracts in general from these short run fluctuations.

We also relate and contribute to the literature on dynamic advertising strategies. Dubé et al. (2005) build on a set of papers deriving optimal dynamic advertising strategies. They use weekly data to estimate a model that features an advertising goodwill stock, which enters the utility associated with buying a product. In their model, advertising has stronger effects when the advertising goodwill stock exceeds a certain level, with decreasing returns at higher levels. This provides a rationale for buying blocks of advertising over time, which is referred to as pulsing in this literature. Sahni (2015) provides an additional rationale for spreading advertisements over time. He shows that the effects of reaching a consumer for the third time depend on the spacing between the first and the second time (a matter of weeks rather than days in his context). He interprets this finding as related to learning in the context of product discovery, in his case about the existence of restaurants. In our case, consumers are well-aware of the product and therefore learning effects are less important and advertising mainly acts as a reminder. This can be seen as an advantage of our empirical setup that allows us to focus on how reminder advertising affects consumer behavior in the absence of learning. Focusing on reminder advertising we find that in general advertising later, when consumers are more inclined to buy once reminded, has bigger effects. Moreover, our structural estimates imply an S-shaped relationship between the probability to consider and the advertising goodwill stock. Reaching a consumer twice is associated with a probability to consider buying that is slightly more than twice that probability when the consumer is only reached once, while reaching her for the third time only leads to a smaller increase. This gives rise to a tradeoff that is different from the one studied by Dubé et al. (2005), as both advertising late and spreading it over time so that consumers are not reached three times in a row has benefits. We explore this tradeoff further in our counterfactual experiments.

Finally, our paper relates to the literature that is concerned with modeling the decision of when to buy a product. 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 an attention stage in which advertising has an effect. Our 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. De Groote and Verboven (2016) study the decision of consumers to buy subsidized (by the state) solar panels and put them on their roof. Our results can be useful to study how such a subsidy scheme could be complemented with an advertising campaign raising awareness. Our results suggest that it could be that lowering the large subsidy and at the same time accompanying the subsidy scheme by an advertising campaign could lead to higher levels of adoption at lower cost.

3 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). The Dutch State Lottery 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 De Lotto. 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.[6]

[6] In 2014, the Dutch State Lottery 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).

The lottery run by the Dutch State Lottery 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 the Dutch State Lottery, 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.

The expected payoff for an individual depends on the jackpot and the number of people who hold a ticket on the day of the draw. But importantly, there is no information revelation for a particular draw between the first day on which a ticket can be bought and the draw, as it is not communicated how many people have bought a ticket at a given point in time. This allows us to treat the expectation thereof as constant over time and capture its effect by draw fixed effects.

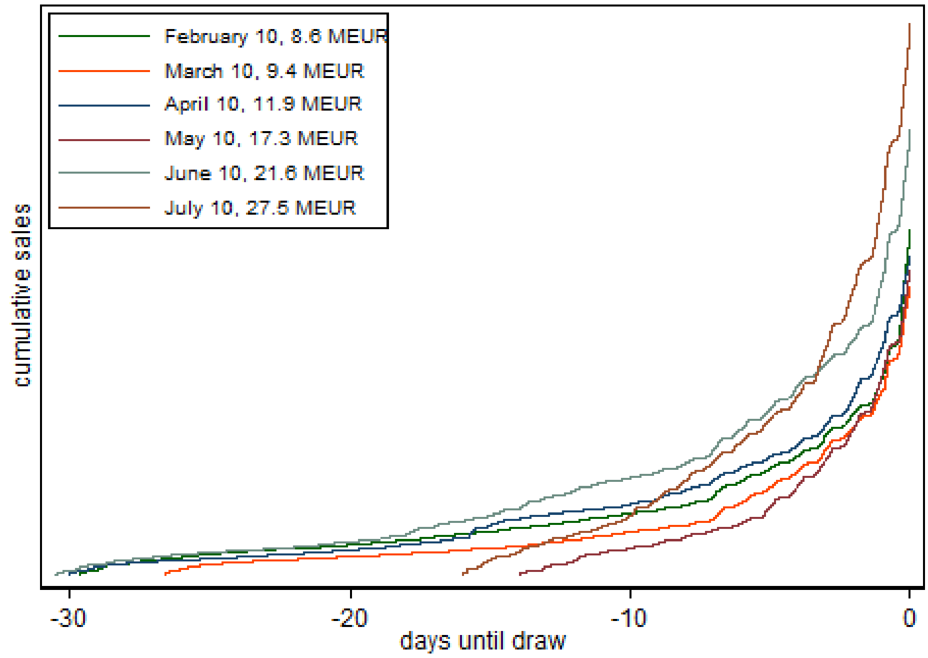
4 Data

4.1 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.[7] The advertising data consists of minute-level measurements of gross 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.

[7]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 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.

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.

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.

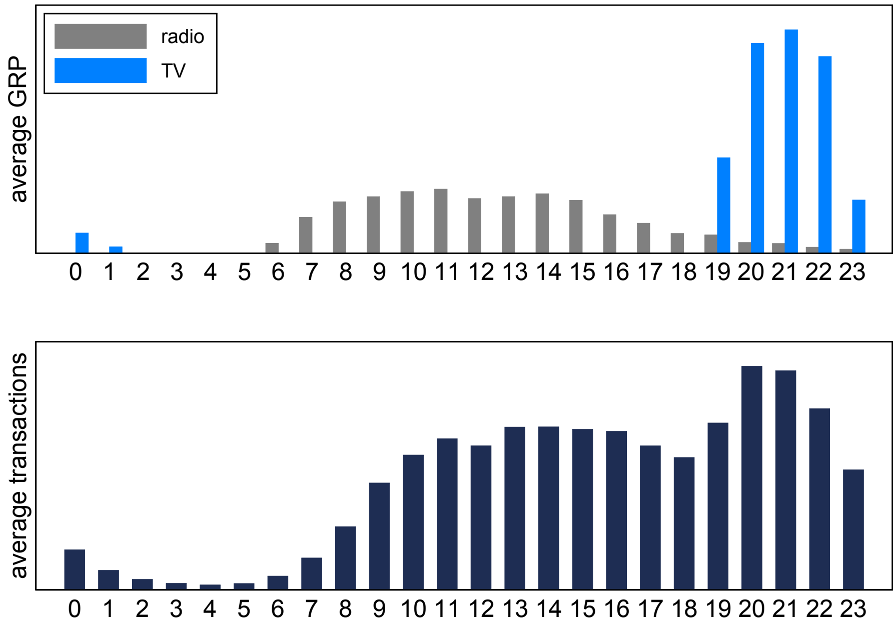
4.2 Descriptive evidence

Figure 1 shows cumulative sales for 6 selected regular draws against the time until the draw, together with the respective jackpot size.[8] 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).[9] The main take-away from this figure is that it strongly suggests that consumers value buying a ticket shortly before the draw.

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

[9]We do not expect this to have big effects, however, because most tickets are sold in the week before the draw. But we do take this into account in our structural model.

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.

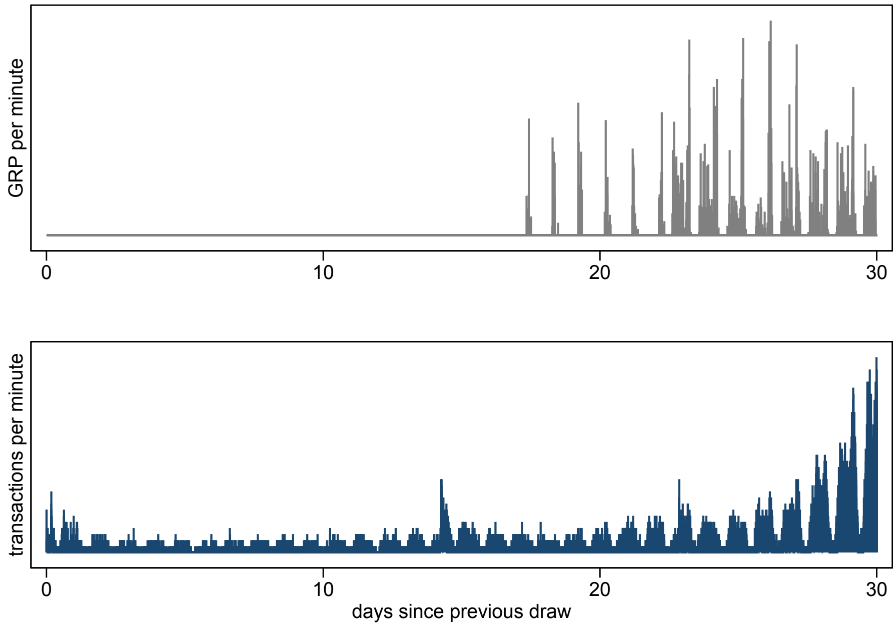
The figure shows that across draws there is a positive relationship between jackpot size and sales (that is, cumulative sales on the day of the draw). The draw on July 10 has the largest total sales of the 6 draws. It also has the largest jackpot size. The second largest sales for the draw on June 10, 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 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.[10] 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

[10]See Table 6 in Appendix C for details.

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.

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.[11]

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.[12]

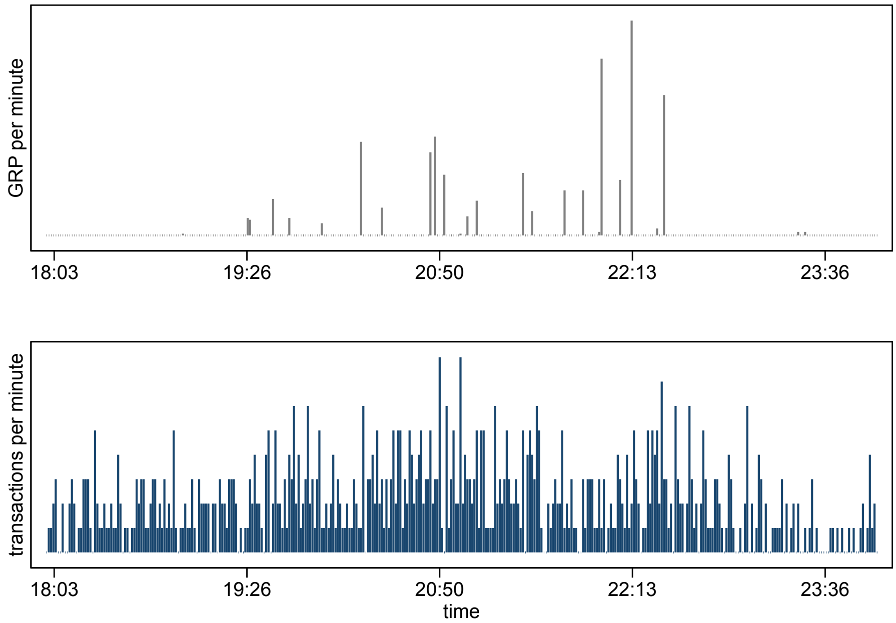
From now on we pool radio and TV GRP's together. Figure 3 shows GRP's and sales at the minute level for one regular draw.[13] We see that the firm starts advertising on the 17th day after the last regular draw, while sales only increase in the last days before the draw. This is already

[11]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 the number of GRP's increases in a continuous manner between 7pm and 9pm and did not sharply jump to a high level right after 7pm.

[12]This is a well-known challenge for the analysis of advertising effects. Our identification strategy for overcoming this challenge and measuring the effects of advertising is akin to a regression discontinuity design and described in Section 6 below.

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

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.

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.

5 Advertising as a reminder

5.1 Definition

In this section, we provide a more precise definition of reminder advertising based on consideration sets that vary over time and develop two empirical predictions related to advertising acting as a reminder.

Consider a consumer i who is well-informed about the product and its attributes and has preferences that are not influenced by advertising. The baseline probability that she thinks about buying the product at time t is $P_{it}(\text{consider})$, as in Manzini and Mariotti (2014). Once she thinks about it, she will either buy the product or not. This is also driven by taste shocks and the probability to then buy it is $P_{it}(\text{buy}|\text{consider})$. Note that both probabilities may vary over time.

An advertisement reaching consumer i will increase the likelihood that she thinks about buying the product, from $P_{it}(\text{consider})$ to $P_{it}(\text{consider}) + \Delta_{it}$. Pure reminder advertising does not contain any new information and does not influence preferences and therefore an exclusion restriction holds: it will only influence the probability to think about buying a ticket, but not the decision whether or not to buy a ticket once a consumer is thinking about it. It follows that the effect of pure reminder advertising is $\Delta_{it} \cdot P_{it}(\text{buy}|\text{consider})$.

5.2 Empirical predictions

We can now take this definition of reminder advertising as a starting point to derive two empirical predictions:

First, advertising effects are short-lived, as advertising draws additional attention, Δ_{it}, to considering to buy a ticket and we expect this attention to fade away quickly. This hypothesis was developed by Calzolari and Nardotto (2016) in the context of e-mail reminders, who also reference the literature on forgetting in this context.

Second, the absolute effect of reminder advertising is the bigger the less time there is until the draw. It follows from the definition above that the effect of reminder advertising is directly proportional to the probability to buy given consideration, ($P_{it}(\text{buy}|\text{consider})$. We expect this probability to be the bigger the less time there is until the draw. The underlying way to think about consumers is that they are exposed to limits of information processing power and attention and may therefore forget to make an intended purchase. Once they think about buying a lottery ticket, they weigh the costs of doing so at that point in time against the benefits. Costs are immediate and 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. For that reason, consumers value to be reminded to buy a ticket as late as possible. In addition, when they consider to buy a ticket early, then they will be reluctant to do so because they will likely consider buying a ticket in the future. This in turn means that the effects of increasing consideration through advertising will be the 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 motivated by the descriptive evidence in Section 4 and fully compatible with the structural model we propose in Section 7. 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).

6 The effects of advertising

For several reasons, the institutional setup that is described in Section 3 is particularly well-suited for studying the effects of reminder advertising. First, many of the advertisements explicitly remind consumers to buy a ticket.[14]

Second, the lottery ticket that is sold is a very well-known product. Its introduction predates the birth of everyone alive in the Netherlands and its characteristics have essentially not changed over time. This includes the dates on which there is a draw, which is usually on the 10th of the month and on 4 special occasions, except for the jackpot size that changes each month. For that reason, learning about the existence and time-invariant characteristics of the product on the one hand and transformative effects of advertising on the other hand will likely not play important roles.

Third, we can abstract from competing products.

Given the above, in order to provide evidence on the effects of reminder advertising and to test the two empirical predictions developed in Section 5 we must overcome two important empirical challenges. The first challenge is to measure the causal effect of advertising on sales. The approach we take in this paper is to exploit the high frequency nature of our data and the exogenous variation that is created by the institutional environment. Section 6.1 provides details. The second challenge is that advertisements do not only remind individuals to buy a ticket before the draw, but also contain information about the jackpot size. This means that we need to establish how much of the effect of advertising is related to them acting as a reminder. We turn to this question in Section 6.3.

6.1 Identifying short term effects of advertising

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 to upward-biased estimates of advertising effects even if one controls for month or week dummies. 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

[14]Our data contain no systematic information on the contents of the advertisements and therefore we are not able to say how often the reminding was explicit. An example of a TV advertisement is available at https://www.youtube.com/watch?v=RwRSqOpThzM (accessed 6 September 2018). The main message of the advertisement is that about 1 million tickets typically win a price and that there are 5 days until the draw on September 10.

time window.[15] 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 also alternatively 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 start by testing the first empirical prediction developed in Section 5.2, namely that advertising effects are short-lived, as advertising draws attention to considering buying a ticket, and this attention then vanishes again quickly. For this, 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 various ways. Then, we aggregate the data to the hourly level to verify that estimated effects are similar. Finally, we turn to the second empirical prediction and show that advertising effects are indeed the bigger the less time there is until the draw.

6.2 Direct evidence for big advertisements

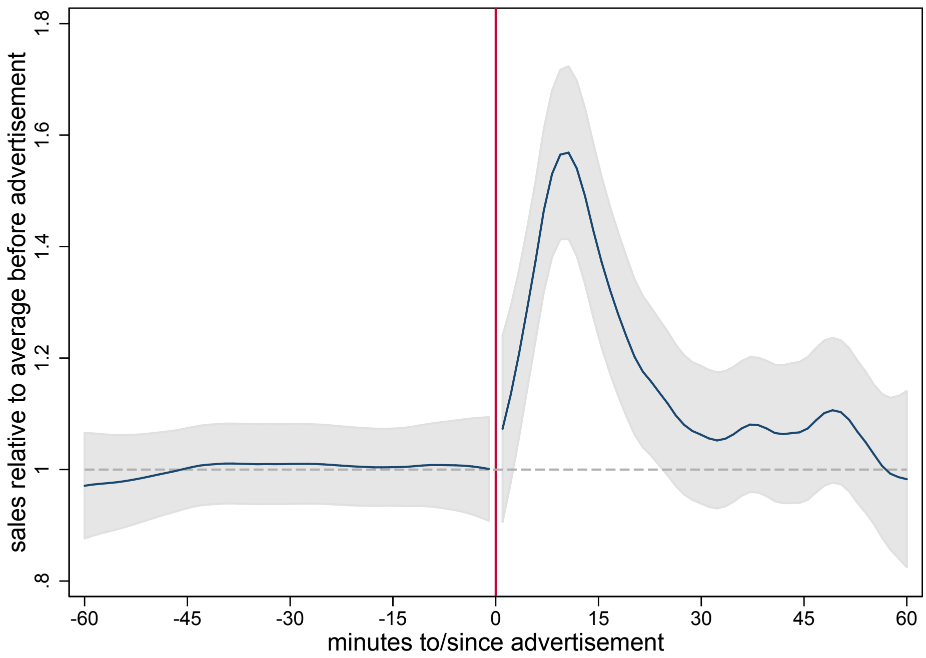
In our data, there are a number of relatively small advertisements. This means that there is often only a short amount of time between advertisements. For that reason, providing direct evidence on the effect of advertisements is challenging, as advertising effects may overlay each other. Our first approach 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.[16] 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 ad-

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

[16] 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.

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.

vertisement 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 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.[17]

In Section 5.1 we have provided a framework for thinking about advertising effects in our context. Assuming that advertising indeed only affects the probability to consider buying (we provide support for this in Section 6.4 below), Figure 5 shows how the probability to consider buying increases when consumers are reached by an advertisement. In line with our first empirical prediction developed in Section 5.2, the effect vanishes within one hour for big advertisements,

[17]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.

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

6.3 Evidence from a distributed lag model

A distributed lag model is a model in which we regress sales on lagged amounts of advertising. Following the identification argument in Section 6.1 we control for time effects using either a combination of draw, time of the day and days until the draw fixed effects, or by controlling for blocks of time. 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.[18]

Table 1 shows the results when we use the log of one plus sales as the dependent variable.[19] Column (1) is for our baseline specification with draw, days to draw, and hour of day dummies. 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.[20]

Column (2) confirms that there is indeed an upward bias when we do not control for time effects. The estimated effects are more than twice as high. Column (3) is inspired by Figure 2 showing that there is a big difference in both sales and advertising levels between day and night. Here we control for a night dummy, which takes on the value 1 between midnight and 7am, instead of the hour of the day. Results are very similar to the ones presented in column (1), suggesting that controlling for the time of the day in terms of day and night is sufficient to alleviate endogeneity concerns.[21] In column (4) we additionally control for day of the year dummies (which implies that we control for draw and days to draw dummies). Finally, in column (5) we control for hour of the year dummies. Importantly, results are again similar.

[18] Note that the empirical strategy we use here is similar, but slightly different from the one we used in Section 6.2 above. The two strategies have in common that we assume that the exact timing of advertising is random conditional on time effects. In Section 6.2 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. In one specification below we control for hour of the year fixed effects, which is very similar.

[19] 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.

[20] 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.

[21] We will make use of this when estimating the structural model. See Section 7.5 for details.

Table 1: The effect of advertising on sales

	(1) baseline	(2) no controls	(3) night dummy	(4) day dummies	(5) hour blocks	(6) small jackpot	(7) big jackpot
GRP between 0 and 4 minutes ago	0.0152***	0.0485***	0.0177***	0.0234***	0.00771***	0.0253***	0.0169***
	(0.00165)	(0.00331)	(0.00170)	(0.00246)	(0.000978)	(0.00292)	(0.00298)
5 and 9 minutes	0.0350***	0.0677***	0.0377***	0.0431***	0.0286***	0.0368***	0.0354***
	(0.00199)	(0.00333)	(0.00202)	(0.00250)	(0.00168)	(0.00340)	(0.00311)
10 and 14 minutes	0.0369***	0.0685***	0.0392***	0.0450***	0.0313***	0.0381***	0.0389***
	(0.00180)	(0.00318)	(0.00182)	(0.00228)	(0.00160)	(0.00317)	(0.00262)
15 and 19 minutes	0.0272***	0.0578***	0.0287***	0.0354***	0.0231***	0.0308***	0.0300***
	(0.00160)	(0.00317)	(0.00160)	(0.00227)	(0.00133)	(0.00317)	(0.00226)
20 and 24 minutes	0.0224***	0.0517***	0.0235***	0.0304***	0.0190***	0.0303***	0.0248***
	(0.00147)	(0.00301)	(0.00147)	(0.00226)	(0.00105)	(0.00268)	(0.00237)
25 and 29 minutes	0.0194***	0.0468***	0.0199***	0.0270***	0.0160***	0.0275***	0.0228***
	(0.00155)	(0.00295)	(0.00153)	(0.00232)	(0.00105)	(0.00275)	(0.00287)
0.5 and 1 hour	0.0152***	0.0375***	0.0141***	0.0221***	0.0127***	0.0237***	0.0178***
	(0.00142)	(0.00253)	(0.00131)	(0.00223)	(0.000975)	(0.00254)	(0.00252)
1 and 1.5 hours	0.0106***	0.0242***	0.00953***	0.0155***	0.0122***	0.0156***	0.0132***
	(0.00110)	(0.00189)	(0.00103)	(0.00166)	(0.000999)	(0.00222)	(0.00186)
1.5 and 2 hours	0.00832***	0.0192***	0.00784***	0.0124***	0.0116***	0.0153***	0.00993***
	(0.00121)	(0.00197)	(0.00119)	(0.00168)	(0.00115)	(0.00230)	(0.00179)
2 and 2.5 hours	0.00216*	0.00962***	0.00393***	0.00491***	0.00740***	0.00636***	0.00509**
	(0.000989)	(0.00176)	(0.000942)	(0.00130)	(0.00103)	(0.00140)	(0.00155)
2.5 and 3 hours	-0.00177	0.00720***	0.00144	0.00130	0.00442***	0.00355*	0.00147
	(0.000932)	(0.00155)	(0.000857)	(0.00111)	(0.000889)	(0.00151)	(0.00124)
3 and 3.5 hours	-0.00647***	0.00411*	-0.00291**	-0.00347**	0.00137	-0.00117	-0.00231
	(0.00116)	(0.00164)	(0.00102)	(0.00109)	(0.000763)	(0.00115)	(0.00169)
3.5 and 4 hours	-0.0105***	0.00476*	-0.00531***	-0.00634***	-0.000274	-0.00427*	-0.00570*
	(0.00148)	(0.00216)	(0.00138)	(0.00138)	(0.000649)	(0.00172)	(0.00221)
draw dummies	Yes	No	Yes	No	No	Yes	Yes
days to draw dummies	Yes	No	Yes	No	No	Yes	Yes
hour of day dummies	Yes	No	No	Yes	No	Yes	Yes
night dummy	No	No	Yes	No	No	No	No
day of year dummies	No	No	No	Yes	No	No	No
hour of year dummies	No	No	No	No	Yes	No	No
Observations	515205	515205	515205	515205	515205	226763	185253
R^2	0.624	0.192	0.595	0.616	0.769	0.594	0.633

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 clustered at the daily level. The first five columns differ by the set of control variables. The night dummy takes on value 1 in the time between midnight and 7am. The sixth column uses only data from regular draws with a below-median jackpot, and the seventh column uses only data from regular draws with an above-median jackpot.

Overall, this suggests that the way we control for time effects in specification (3) is sufficient.[22]

We have also estimated similar models using data aggregated to the hourly level. Table 7 in Appendix C shows the results. We report results for both a specification with hour of day dummies and a night dummy. 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 and three hours after the advertisement was aired, respectively. Comparing Table 1 to Table 7 shows that if anything aggregating the data to the hourly level will lead to slightly smaller estimated effects and in that sense estimates from hourly data will be conservative. This is important, because it would not be feasible to estimate a structural model at the minute level.

Overall, the findings reported here provide additional evidence on advertising effects being short-lived, in line with the first empirical prediction developed in Section 5.2.

6.4 Does reminder advertising also contain information?

Before testing the second empirical prediction, we now turn to the important question whether these estimates are indeed describing the effects of reminder advertising or could be explained by advertisements containing information on the jackpot size. *A priori,* it is unlikely that the short-term effect of advertising that is associated with advertising acting as a reminder depends on the information about the jackpot size, for two main reasons. First, the new jackpot becomes known right after the last draw and no new information is revealed until the next draw, also not on the number of tickets sold. This means that many consumers who would be interested in buying lottery tickets would know right after the last draw whether the next jackpot is high. Second, information on the jackpot size that is conveyed in TV and radio advertisements is usually not news to consumers. If they have not yet learned about the new jackpot size right after the last draw, then they are very likely to learn about it from display advertisements on the street. Usually, a high jackpot size creates a buzz associated with higher baseline and overall sales. Conversely, consumers can infer a low jackpot size from the absence of a buzz.

A first piece of evidence is that our estimated advertising effects are immediate and short-lived. Calzolari and Nardotto (2016) have argued that finding such a pattern for other reminders can be explained by limited attention, as reflected in our first empirical prediction. If instead advertising would be informative, then we should estimate more persistent effects.

Next to this, it is useful to return to the discussion in Section 5. Recall that reminder advertising increases the probability to consider buying, $P_{it}(\text{consider})$, by Δ_{it} and the effect of reminder advertising is therefore $\Delta_{it} \cdot P_{it}(\text{buy}|\text{consider})$. In general, advertising could also have an impact on the probability to buy given consideration and change it from $P_{it}(\text{buy}|\text{consider})$ to $P_{it}(\text{buy}|\text{consider}) + \Gamma_{it}$, in particular when advertising conveys new information. Then, the total

[22] In addition, we have estimated specification (1) with additional minute of the hour dummies to see whether seasonality within each hour could confound advertising and sales. Results were again very similar.

effect of the advertisement is

$$(P_{it}(\text{consider}) + \Delta_{it}) \cdot (P_{it}(\text{buy}|\text{consider}) + \Gamma_{it}) - P_{it}(\text{consider}) \cdot P_{it}(\text{buy}|\text{consider})$$
$$= \Delta_{it} \cdot P_{it}(\text{buy}|\text{consider}) + (P_{it}(\text{consider}) + \Delta_{it}) \cdot \Gamma_{it}.$$

The part of the effect of the advertisement that is related to it acting as a reminder can then still be defined as $\Delta_{it} \cdot P_{it}(\text{buy}|\text{consider})$.

Here, information that lets consumers positively update their priors on the jackpot size will increase the probability to buy given consideration ($\Gamma_{it} > 0$). Information that lets consumers negatively update their priors will lead to a decrease in that probability ($\Gamma_{it} < 0$). If the effect on consideration is Δ_{it} in either case, then this means that the percentage increase in sales will be bigger when the news is positive. Conversely, if advertising is mainly reminding consumers to think about buying a ticket, then the percentage effect on sales should be the same when the jackpot is big as compared to the case when it is small.

Based on this, we can test whether part of the estimated effect of advertising on sales is due to advertising conveying information about the jackpot size by comparing the estimated percentage effect when the jackpot is above the median jackpot in the sample to the effect when the jackpot is below the median jackpot. We split the sample and report estimates for both cases in column (6) and (7) of Table 1. The estimated effects are percentage changes and are strikingly similar, which suggests that indeed advertising mainly acts as a reminder.

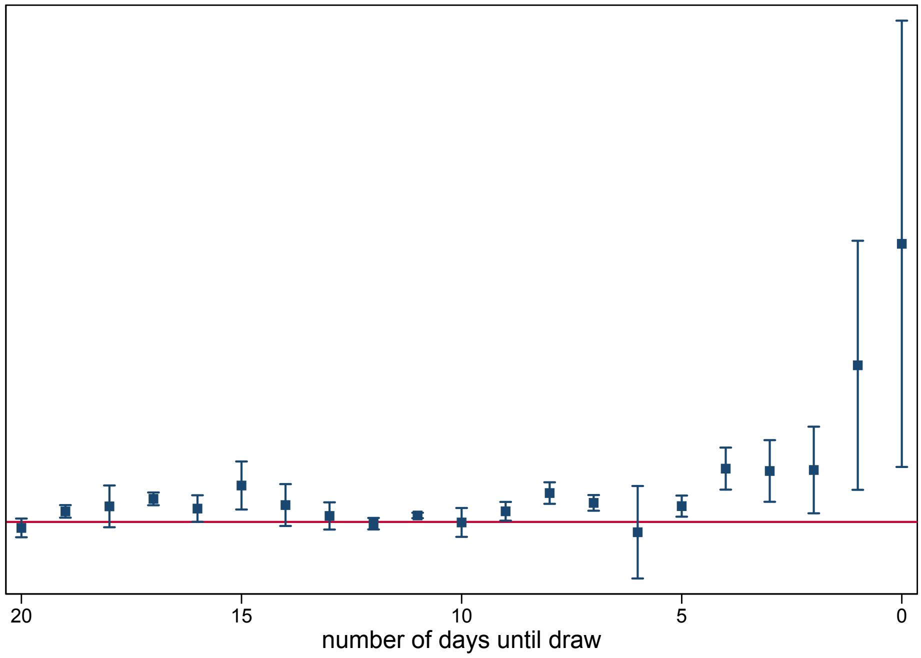
As an additional piece of evidence we can compare the estimates reported in column (1), where we only control separately for draw and days to draw dummies, to the estimates in column (4) and (5), where we allow for the interaction of the two. This means that we control for persistent effects of the information conveyed in the advertisements in specification (4) and (5). If important, then this should lead to substantially different estimates of advertising effects than those reported in column (1). This is not the case, which provides additional support for the view that the measured advertising effects can be related to advertising acting as a reminder.

6.5 The dependence of advertising effects on the time until the draw

Our second empirical prediction developed in Section 5.2 is that advertising effects are the bigger the less time there is until the draw. To test this hypothesis, 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 to the hourly level and take first differences to control for patterns in baseline sales. We specify

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

Figure 6: Effect of timing



Notes: This figure shows the immediate effect of one GRP of advertising on sales by day until the draw, for the last 21 days. See text for details.

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.[23] Moreover, guided by the finding in the previous subsection that the advertising effect lasts for about 2 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.[24]

In Figure 6, we plot the estimated effects and the corresponding 95% confidence intervals against the number of days until the draw.[25] 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

[23]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.

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

[25]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.

would have bought anyway (purchase acceleration). Or it means that consumers who would otherwise not have bought a ticket for a particular draw actually did buy a ticket (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 for a particular draw 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. In line with the two empirical predictions developed in Section 5.2, the effects are short-lived and the absolute effect of advertising on sales is higher the less time there is until the draw. Our results also show that the information about the jackpot size is not reflected in the size of advertising effects and that advertising does not only lead to purchase acceleration, but also to market expansion.

7 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. Our model can also be used to quantify these effects. In particular, 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.[26]

[26]This is why we estimate the model using the method of simulated moments. See Section 7.7 below for details.

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.

7.1 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 either considers buying a ticket or not, both with a certain probability. Advertising has a positive effect on this probability. If an individual considers buying a ticket, then she actively makes a decision. If she buys a ticket, then she receives a one-off flow of utility and cannot make any decisions anymore.[27] Otherwise, she continues in the next period and has the option of buying a ticket there.

7.2 Consideration

In our model, advertising affects the likelihood that a consumer considers buying a ticket through an advertising goodwill stock. This goodwill stock increases if the individual is exposed to an advertisement, which is uncertain from an *ex ante* perspective.

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 for instance Dubé et al. (2005) for lower frequency data, 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^a_{it}. 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

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

[27] 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.

with initial condition

$$g_{i0} = 0$$

and

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

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

$$P_{it}(\text{consider}) = \frac{1}{1+\exp\left(-(\gamma_0 + \gamma_1 g^a_{it} + \gamma_2 (g^a_{it})^2)\right)}.$$

This flexible functional form allows for both a convex and a concave relationship between the probability to consider and the advertising goodwill stock, or an S-shaped response.

7.3 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 draw-specific value of holding a ticket at the time of the draw (it will be estimated as a draw fixed effect), 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 in Figure 1 that most sales occur in the last days before the draw.

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 7.5. 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. At the same time, there is a benefit, as they won't forget to buy a ticket at a later point in time.

7.4 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_{it}^{a} summarizes all relevant information on consumer i's advertising exposure in the past. In addition, the value function in Section 7.3 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, 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

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

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.

7.5 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 and by the specification checks in Section 6.3 (related to column (3) in Table 1 and column (2) in Table 7).[28]

[28] 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 6.3 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

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 goodwill 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 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.

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 $V_{it}(\tilde{g}_{it}^a)$. 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

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

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.

Here, the expectation is taken over the advertising goodwill stock and the taste shocks. The goodwill stock $\tilde{g}^a_{iT-1}$ either changes to $\tilde{g}^{a+}_{iT}$ when the individual sees an advertisement in T, which will be the case with probability P_T, or it changes to $\tilde{g}^{a-}_{iT}$ if not, with probability $1-P_T$ (see also Section 7.2). The two values $V_{iT}(\tilde{g}^{a+}_{iT})$ and $V_{iT}(\tilde{g}^{a-}_{iT})$ are obtained using interpolation. From this, we get that the value when considering in the second to last period is

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

and the value function is

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

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

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

and the unconditional probability of buying as

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

7.6 Identification

Our data consist of aggregate sales and GRP measurements at the hourly level. They are well-suited to estimate baseline sales and advertising effects, but do not contain enough variation to estimate all parameters of the model. We therefore set the baseline probability to consider buying to what we believe is a reasonable value (in combination with a robustness check) and impose a restriction on the value to holding a ticket on the day of the draw.

Before going into detail, recall that we have normalized the coefficient on the price to be minus one and the value of the outside option of not buying a ticket to be zero. Moreover, we treat the market size and the price of a ticket as known.

Parameters capturing the evolution of baseline sales

For the moment assume that the baseline probability to pay attention is known and abstract from advertising. Recall that we have specified the probability to pay attention to be independent of the time until the draw, while the environment is clearly non-stationary, as sales depend strongly on the time until the draw (see Figure 1). For a given probability to pay attention the parameters

δ, ψ and σ are identified if only one combination of them generates the sales pattern in the data.

We have used the model to generate predictions of sales trajectories. For a given level of sales just before the draw δ changes how much sales depend on the time until the draw, i.e. the shape of the cumulative sales curve. The lower we set δ the steeper the curve becomes just before the draw and the more convex it is. This suggests that data are informative about the discount factor. The underlying reason is that buying a ticket leads to an immediate cost and a delayed benefit. The longer the delay, the more the benefit is discounted. This can be seen as an exclusion restriction. Magnac and Thesmar (2002) and Abbring and Daljord (2019) formally show that exclusion restrictions are useful for identification of the discount factor.

Our predicted sales trajectories also show that ψ directly affects the level of sales on the day of the draw and also cumulative sales for given δ. Finally, σ determines sales long before the draw. In the limit, the probability to buy given attention is 0.5 for large values of σ.

However, it turns out that it is difficult to separately identify the ψ's and σ when we generate trajectories that resemble the ones in Figure 1, as sales are very low even a few days before the draw and this is rationalized by a small enough δ. So, the level of sales long before the draw is not useful for the identification of σ. It is difficult to separately identify the ψ's and σ within a plausible range of values for those parameters, because increasing ψ and at the same time σ leads to a similarly shaped curve for cumulative sales.

One may be concerned about this, especially when the aim of estimating the structural model is to conduct welfare analyses, as the size of welfare effects is closely related to σ. However, our primary aim is to conduct counterfactual simulations for the dependence of total sales on advertising schedules. Therefore, we impose that the weighted average value to holding a ticket across all 16 draws is 2 euros, which is 2/3 of the price of a ticket. That is, we impose that $\Sigma_{i=1}^{16} w_i \cdot \psi_i = 2$, where the weight $w_i = \frac{\text{total sales of draw i}}{\text{total sales of the whole year}}$. We find this a reasonable assumption, as it roughly corresponds to the expected value to holding a ticket. The interpretation is that buying a ticket is driven by taste shocks, and not by the fundamental value to holding a ticket. These taste shocks capture the pleasure associated with participating in a lottery. Given this restriction on the average level of the ψ's, σ is identified from the average sales across draws. There are 16 draws with 16 values of total sales, and we estimate 16 parameters (15 Ψ's and σ).

Baseline level to consider buying

The baseline level of paying attention is directly linked to γ_0. There are recent advances in the literature related to identifying this baseline level, in particular Abaluck and Adams (2017) and Heiss et al. (2016). However, the necessary variation is not in our aggregate data. Therefore, we set $\gamma_0 = -4.5$ so that the baseline probability to consider buying is 1 percent. In Section B.2, we re-estimate the model with a different value for γ_0. Parameter estimates change in intuitive ways and counterfactual predictions remain similar.

Parameters capturing advertising effects

So far we have abstracted from advertising. The parameters governing the evolution of baseline sales are either estimated or assumed known. In our model, advertising affects sales only through changing the probability of paying attention. This means that there is a close link between the data and this probability, for the advertising histories in the date. We use simulated consumers, which means that for the discussion here we can treat the advertising goodwill stock as known. Then, the decay of advertising effects identifies the depreciation rate λ. The magnitude of the effects then pins down γ_1 and γ_2.

7.7 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 7.3 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 simulate because we follow simulated consumers over time, who are heterogeneous in their advertising goodwill stocks (see Section 7).

We assume the market size for Dutch online lottery tickets market is 250,000 and we simulate choices of 1,000 consumers.[29] Thus, each simulated consumer represents 250 real consumers. In the background, there is again a trade-off between computational burden and how realistic the model is. We found that simulating 1,000 consumers works well for capturing heterogeneity in the simulated advertising goodwill stock while the computational burden is still not too high. 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.[30]

[29] This market size 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 Appendix B we also present results when we assume that the market size is higher.

[30] 3 euros is the price for the smallest ticket one can buy. See Section 3 for details. See also footnote 7. 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^c_{iT}, $V_{iT}(\tilde{g}^a_{iT})$ and $\mathbb{E}\left[\max V_{iT} \,|\, \tilde{g}^a_{iT-1}\right]$ will double. Consequently, $V^c_{iT-1}(\tilde{g}^a_{iT-1})$ and $V_{iT-1}(\tilde{g}^a_{iT-1})$ 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.

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 be more likely to buy, so that those with lower advertising goodwill stocks remain. Our strategy allows and controls for that. For 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 on a grid for the advertising goodwill stock we follow the simulated consumers. For each simulated consumer we evaluate the value functions that we have already solved for at the simulated goodwill stock. From this we can compute implied choice probabilities. We can combine those with random draws u_{it} from the standard uniform distribution for each consumer at each point in time to generate simulated choices. That is, we calculate the probability of buying at the simulated goodwill stock $\tilde{g}^a_{it}$ 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 $\hat{d}_{it}$ 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.

8 Results

8.1 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 53.0 percent. Recall that we set γ_0, the intercept of goodwill stock on probability of considering at -4.5. This means that about $1/(1+exp(-(-4.5))) \approx 1$ percent of the consumers will consider buying a ticket in the absence of advertising. A one unit increase in the goodwill stock from zero to one, driven by seeing an advertisement, will increase the probability of considering to $1/(1+exp(-(\gamma_0+\gamma_1+\gamma_2))) \approx 0.194$. One hour later, the goodwill stock is $1-0.530 = 0.470$ and the probability to consider buying is $1/(1+exp(-(\gamma_0+\gamma_1 \cdot 0.470+\gamma_2 \cdot 0.470^2))) \approx 0.052$ if no advertisement reaches the consumer. Yet another hour later it is 0.022 and the probability to consider is 0.011. If the consumer is instead reached by another advertisement in the hour after she was first reached, then the augmented goodwill stock becomes

Table 2: Parameter estimates

parameter	estimate	ste.
depreciation rate goodwill stock (λ)	0.530	0.028
effect of goodwill stock on probability of considering (γ_1)	3.672	0.129
effect of squared goodwill stock on probability of considering (γ_2)	-0.599	0.116
hourly discount factor (δ)	0.996	0.000
multiplying factor taste shock (σ)	0.511	0.018
value to having a ticket on the day of the draw (ψ)		
10 February, 2014	1.512	0.104
10 March, 2014	1.516	0.084
10 April, 2014	1.483	0.066
26 April, 2014 (King's Day)	2.427	0.108
10 May, 2014	1.658	0.054
10 June, 2014	1.712	0.038
24 June, 2014 (Orange draw)	2.157	0.066
10 July, 2014	2.066	0.049
10 August, 2014	0.880	0.101
10 September, 2014	1.822	0.045
1 October, 2014 (special 1 October draw)	1.757	0.071
10 October, 2014	1.740	0.047
10 November, 2014	1.335	0.069
10 December, 2014	1.669	0.070
31 December, 2014 (New year's eve draw)	3.186	0.104

Notes: Structural estimates. Obtained using the method of simulated moments. See Section 7.7 and Appendix A for details on the estimation procedure. The probability to consider buying is specified as $P_{it}(\text{consider}) = 1/\left(1+\exp\left(-(\gamma_0+\gamma_1 g^a_{it}+\gamma_1 (g^a_{it})^2)\right)\right)$. We set $\gamma_0 = -4.5$ and impose that the weighted average value to having a ticket on the day of the draw is 2 euros, using weights proportional to actual sales. The implied value for the draw on 10 January is 1.686 euros.

1.470 in the second period and the probability to consider buying is 0.402. And when she is reached again one hour later, it is 0.499. Hence, we're actually facing an S-curve in the sense that reaching a consumer twice in a row leads to a probability to consider buying that is slightly bigger than twice the probability of reaching a consumer once. In principle this creates a weak incentive to pulse, as the convexity is not very strong. We do not expect pulsing to be optimal, however, as reaching the consumer three times in a row leads to a smaller increase in the probability to consider than reaching him for the first or second time, while reaching a consumer twice is only slightly more effective in generating attention than reaching two consumers once, respectively. Therefore, we expect advertisements to be more effective when they are spread over time. We explore this in more detail in our counterfactual experiments in Section 9.

The hourly discount factor is estimated to be 0.996. This means that one month before a draw, the value consumers attach to a ticket is only 11.5% 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 $0.511 \cdot \sqrt{\pi^2/6} \approx 0.66$ euros ($\sqrt{\pi^2/6}$ is the standard deviation of a type 1 extreme value random variable). Finally, the 15 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.

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

8.2 The dependence of advertising effects on time

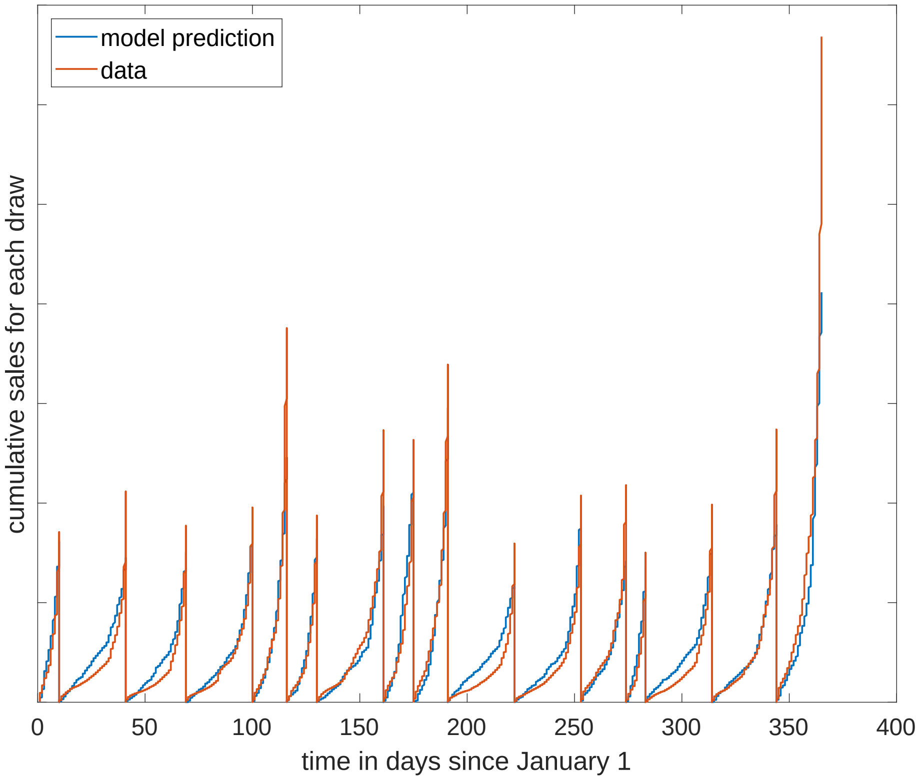
A key quantity the model predicts is the immediate effect of advertising on sales and how this effect depends on the time until the draw. Figure 8 shows, for our structural estimates, how the probability of buying a ticket, if the individual has not done so yet, changes when she is exposed to an advertisement. 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 6. The figure shows that our model can generate this effect.

8.3 The elasticity of sales with respect to advertising

To get a first idea about the size of the implied advertising effects, we have increased all GRP's by 10% and have simulated the effect of this on sales. From this we calculate an elasticity of sales with respect to advertising of 0.34.

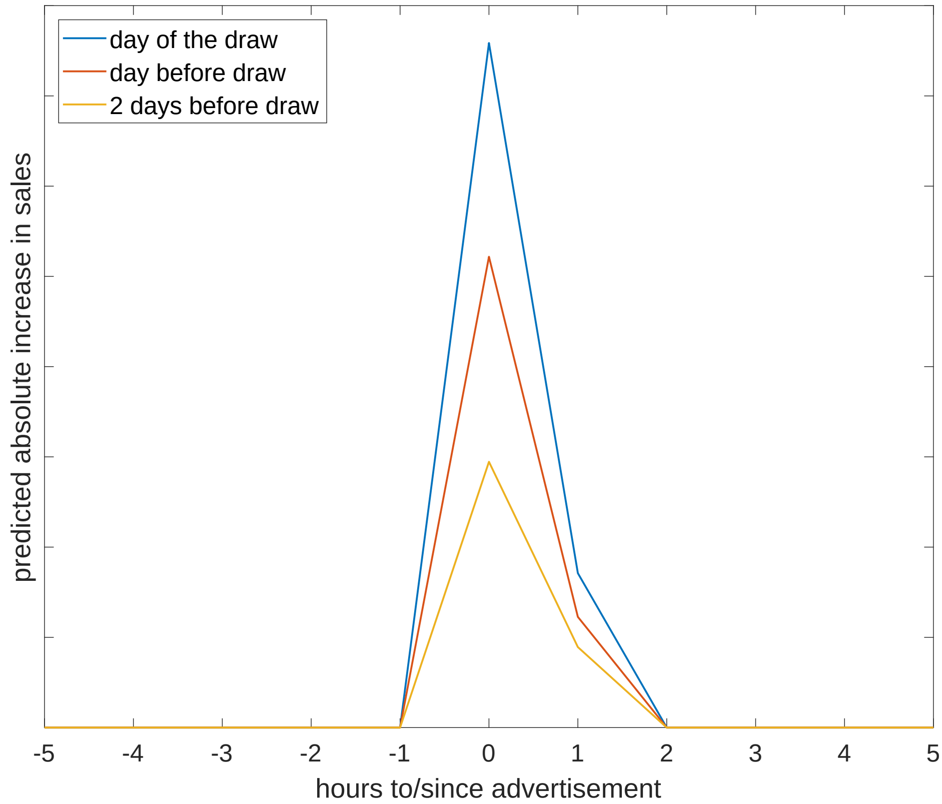
To put this into perspective, the average elasticity Hu et al. (2007) report for their meta study of offline effects of classical TV advertising is 0.113. Shapiro et al. (2019) estimate elasticities

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.

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.

for 288 consumer packaged goods (CPG) and find an average elasticity of 0.014, with about 65 to 75 percent of the elasticity estimates not being statistically different from zero.

Our estimate of 0.34 is considerably higher. This could be the case for various reasons. First, the product is different. The two aforementioned studies are for classical CPG's and not lottery tickets. Second, in general advertising serves many purposes. In our case advertising is designed to remind consumers to buy a ticket. It could be that reminder advertising is particularly effective. Third, in our case consumers have the possibility to immediately buy the product online, whereas they might forget doing so when they need to first visit the store at some future point in time, as it is the case for the classical CPG's.

9 Counterfactual experiments

Having estimated the model, we return to the question what the effect of alternative dynamic advertising schedules would have been, including the case of no advertising at all. 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 in our context to assume as an approximation that the price of one GRP does not vary over time, especially because the firm decided to not buy advertising at pre-specified times and regular draws are on the 10th day of the month, independent of the weekday. In that sense it is meaningful to ask the question whether it is possible to sell more tickets when one allocates the same number of GRP's in a different way.

We consider 8 alternative strategies and compare the total 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. Comparing sales under this strategy to baseline sales will allow us to conclude what the overall effect of advertising was. 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.[31] This allows us to assess how strong the effect of advertising at times at which consumers are more likely to buy once considering is. In the third, we spread all advertisements equally in the last 4 days before the draw to quantify this further. Thereafter, we look into more refined dynamic strategies. Our estimates imply that there is a weak S-shaped relationship between the advertising goodwill stock and the probability to consider buying. In principle, this generates an incentive to pulse, but as mentioned in Section 8.1 above we do not expect this to be optimal. It is nevertheless interesting to quantify the effect. We simulate the effects of three pulsing strategies. 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. These counterfactual strategies ignore scheduling constraints at the level of the media companies. To take this into account, the last two counterfactual strategies take, respectively, the schedule as it is in the third week and move it to the fourth week (i.e. adding it to the one in 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 in advertising strategy. Making this distinction is interesting because it allows us to quantify the importance of changes in expectations. The underlying intuition is that consumers should be less inclined to buy earlier if they believe that it is more likely to be reminded later. One can think of this as a second order effect, with the first order effect being the change in the actual advertising strategy. We make the distinction to assess how important this effect on expectations is.

Table 3 shows the result. We first focus on the last column, for rational expectations. Not advertising at all leads to 65 percent of the original sales. Generally, allocating advertising

[31] Recall that we count the night as one hour.

Table 3: Effect of various advertising strategies

strategy	expectations unchanged	expectations rational
data (reference point)	100%	100%
no advertising at all	64%	65%
all advertising in the last 2 days before the draw	147%	142%
spreading advertisements equally in the last 4 days before the draw	124%	121%
pulsing strategy in the last 4 days before draw (1 hour blocks)	122%	120%
pulsing strategy in the last 4 days before draw (2 hour blocks)	119%	117%
pulsing strategy in the last 4 days before draw (1 hour double, 3 hour none)	117%	115%
shift advertising from third week to fourth week	117%	116%
shift advertising from fourth week to third week	81%	81%

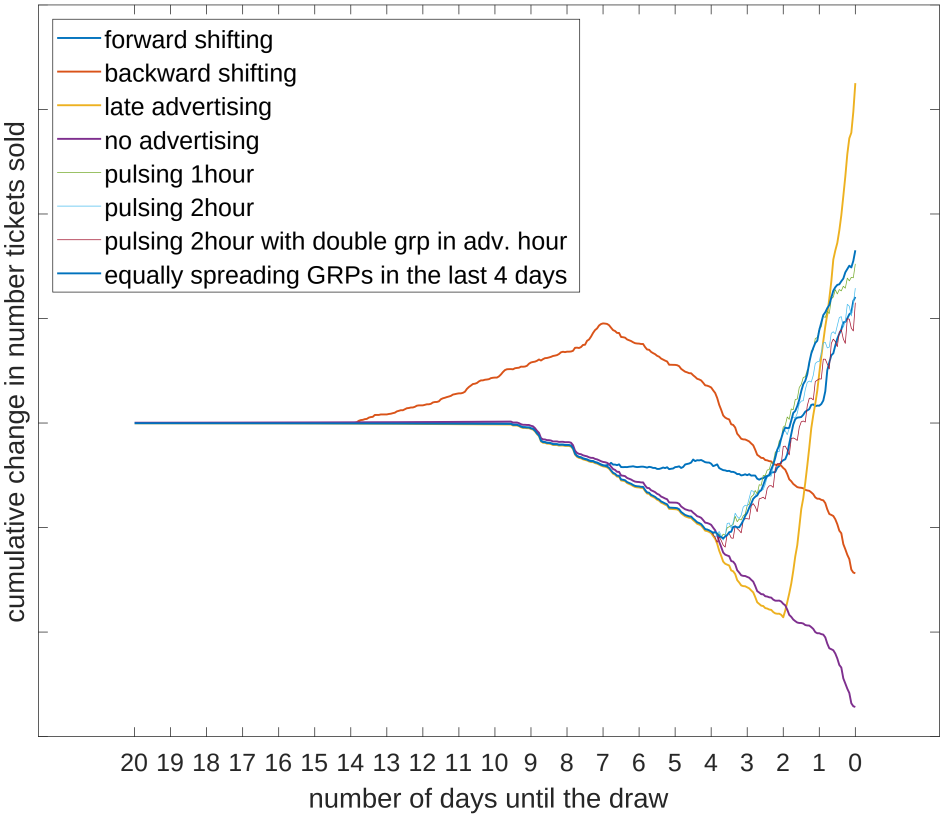
Notes: This table shows the effect of using alternative dynamic advertising strategies for the February draw. See text for a description of these strategies. In the column labeled "unchanged" consumer expectations are consistent with the advertising data we used to estimate the model and not with the 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.

to later points in time increases sales. The most drastic measure we considered was to move it to the last two days before the draw. This lead to an increase in sales of 42 percent. The more realistic experiment of moving advertising from the third to the fourth week still leads to an increase in sales of 16 percent. Comparing the outcome for the pulsing strategies to the one for the strategy where all advertisements are spread equally over the last 4 days before the draw reveals that pulsing is indeed not optimal, but does also not lead to substantially lower sales. This is because the convexity that we estimate is not strong enough to make it a superior strategy.

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 faster than sales for the baseline strategy (hence the difference in cumulative sales is positive) in the third week and less fast in the fourth week (hence the difference is negative), as advertising leads to purchase acceleration. Overall, fewer tickets are sold, which is reflected in the lower end point.

The results presented in Table 3 also show that expectations of consumers matter, but are quantitatively not of first-order importance. Qualitatively, when expectations are rational, then effects become 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.

Figure 9: Effect of different advertising strategies



Notes: This figure 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.

10 Summary and concluding remarks

In 2018, global advertising spending on all kinds of advertising amounted to 552 billion US dollars. 208 billion of this were spent in the US, which is about 1 percent of GDP.[32] There is still no consensus among academic scholars on how one should think about this in general. Possible reasons include that the effect of advertising is highly-context specific and depends on the type of advertisement.

Advertisements are typically classified as either being transformative or informative (Bagwell, 2007). Transformative advertising either changes preferences or enters utility as a complement to a product. Informative advertising provides consumers with information on the availability and characteristics of products, including prices. The aim of this paper is to provide clean empirical support for the view that advertising can also act as a reminder when consumers

[32]Taken from the winter 2018 update to the MAGNA advertising forecast, available at https://magnaglobal.com/magna-advertising-forecasts-winter-2018-update/ (accessed May 2019). More than half of global spending, 301 billion US dollars, were on non-digital advertising that includes TV, radio and print advertising.

suffer from limited attention and to quantify the effects. For this, we use high frequency data to credibly identify advertising effects and we focus on one particular context—the sale of lottery tickets in the Netherlands—that is particularly suitable for the purpose of our study. The reason for this is that the other two pathways—transformative and informative advertising—are *a priori* unlikely to be at play and we can show that there is also no empirical support for advertising conveying information.

We develop two empirical predictions that we test in our reduced-form analysis. Our first prediction is that advertising effects are short-lived, in line with the view that individuals suffer from limited attention. Our second prediction is that advertising effects are stronger the less time there is until the draw. The thought experiment we undertake to measure advertising effects and test these predictions 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 2 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. This provides exogenous variation in the attractiveness to buy a ticket—our data strongly suggest that it is more attractive to do so soon before the draw. The presence of a deadline allows us to show that advertising does not only lead to purchase acceleration (shifting the timing of purchases), but also to market expansion (that more tickets are sold overall). After finding support for our empirical predictions in our reduced-form analysis we spell out a structural adoption model that can generate short-lived advertising effects and advertising effects that are stronger the less time there is until the draw. The model features an attention stage where the probability to consider buying a ticket does not depend on the time, but is affected by advertising. 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 starting from the actual advertising schedule in the data and shifting advertising to later points in time has positive effects on sales. Overall, we find large short-lived effects of advertising on sales and a large effect on total sales.

The context of our study is the market for lottery tickets. This context is particularly helpful for obtaining model-free evidence on the effects of advertising and to attribute the effects to advertising acting as a reminder, rather than it conveying information about the existence and characteristics of products, or advertising changing consumer preferences. Focusing on this context however means that our results do not directly generalize, which opens up opportunities for future research.

A nice follow-up to our study could be a comprehensive content analysis of TV and radio advertisements that measures how much of the impressions are related to advertising conveying information, changing consumer preferences, acting as a complement to consumption, or reminding them to turn an intended purchase into an actual one. For this a survey could be used to train the algorithm classifying advertisements. With this, we also hope that more light can be shed on the question to what extent the slogans "Drink Coca Cola" and "Just do it." are actually reminders rather than persuasion or a complement to consumption.

Reminder advertising stimulates consumers to act on their preferences. By providing evidence for this pathway we hope to stimulate more work on the topic. This seems important to us, as many fundamental results in economics hinge on the presumption that consumers actually act on their preference.

References

Abaluck, J. and A. Adams (2017). What do consumers consider before they choose? identification from asymmetric demand responses. NBER Working Paper 23566, National Bureau of Economic Research.

Abbring, J. H. and Ø. Daljord (2019). Identifying the discount factor in dynamic discrete choice models. *arXiv preprint arXiv:1808.10651v2*.

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.

Blake, T., C. Nosko, and S. Tadelis (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica 83*(1), 155–174.

Calzolari, G. and M. Nardotto (2016). Effective reminders. *Management Science 63*(9), 2915–2932.

Chandrasekaran, D., R. Srinivasan, and D. Sihi (2018). Effects of offline ad content on online brand search: Insights from super bowl advertising. *Journal of the Academy of Marketing Science 46*(3), 403–430.

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 (2019). Immediate responses of online brand search and price search to tv ads. *Journal of Marketing 83*(4), 81–100.

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

Gabaix, X. (2018). Behavioral inattention. CEPR Discussion Paper DP13268. Centre for Economic Policy Research, London, UK.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization 44*(2), 115–129.

Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2018). A comparison of approaches to advertising measurement: Evidence from big field experiments at facebook. Working Paper, Kellogg School of Management, Northwestern University.

Hartmann, W. R. and D. Klapper (2017). Super bowl ads. *Marketing Science 37*(1), 78–96.

Heiss, F., D. McFadden, J. Winter, A. Wuppermann, and B. Zhou (2016). Inattention and switching costs as sources of inertia in Medicare Part D. NBER Working Paper No. 22765, National Bureau of Economic Research, Cambridge, MA, USA.

Ho, K., J. Hogan, and F. Scott Morton (2017). The impact of consumer inattention on insurer pricing in the medicare part d program. *RAND Journal of Economics 48*(4), 877–905.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017). Power to choose? an analysis of choice frictions in the residential electricity market. *American Economic Journal: Economic Policy 9*(4), 192–226.

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.

Johnson, G. A., R. A. Lewis, and E. I. Nubbemeyer (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research 54*(6), 867–884.

Joo, M., K. C. Wilbur, B. Cowgill, and Y. Zhu (2014). Television advertising and online search. *Management Science 60*(1), 56–73.

Joo, M., K. C. Wilbur, and Y. Zhu (2016). Effects of TV advertising on keyword search. *International Journal of Research in Marketing 33*(3), 508–523.

Kiss, A. (2017). Salience and switching. Working Paper, University of Amsterdam, Amsterdam, The Netherlands.

Kitts, B., M. Bardaro, D. Au, A. Lee, S. Lee, J. Borchardt, C. Schwartz, J. Sobieski, and J. Wadsworth-Drake (2014). Can television advertising impact be measured on the web? web spike response as a possible conversion tracking system for television. In *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, pp. 1–9. ACM.

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

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.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete decision processes. *Econometrica 70*(2), 801–816.

Manzini, P. and M. Mariotti (2014). Stochastic choice and consideration sets. *Econometrica 82*(3), 1153–1176.

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

Mehta, N., X. Chen, and O. Narasimhan (2008). Informing, transforming, and persuading: Disentangling the multiple effects of advertising on brand choice decisions. *Marketing science 27*(3), 334–355.

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.

Roberts, J. H. and J. M. Lattin (1991). Development and testing of a model of consideration set composition. *Journal of Marketing Research*, 429–440.

Roberts, J. H. and J. M. Lattin (1997). Consideration: Review of research and prospects for future insights. *Journal of Marketing Research*, 406–410.

Rossiter, J. R. and L. Percy (1987). *Advertising and promotion management.* McGraw-Hill.

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

Sahni, N. S. (2015). Effect of temporal spacing between advertising exposures: Evidence from

online field experiments. *Quantitative Marketing and Economics 13*(3), 203–247.
Shapiro, B., G. J. Hitsch, and A. E. Tuchman (2019). Generalizable and robust "tv" advertising effects. Working Paper, University of Chicago, Chicago, IL.
Shapiro, B. T. (2018). Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of Political Economy 126*(1), 381–437.
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.
Terui, N., M. Ban, and G. M. Allenby (2011). The effect of media advertising on brand consideration and choice. *Marketing Science 30*(1), 74–91.
Van Nierop, E., B. Bronnenberg, R. Paap, M. Wedel, and P. H. Franses (2010). Retrieving unobserved consideration sets from household panel data. *Journal of Marketing Research 47*(1), 63–74.
Zigmond, D. and H. Stipp (2010). Assessing a new advertising effect: Measurement of the impact of television commercials on internet search queries. *Journal of Advertising Research 50*(2), 162–168.

Online Appendix

A Details on the econometric implementation

In Section 7.7, 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 the draws differ with respect to the total number of hours in which a ticket can be bought (the draw-specific T in the model) and the draw-specific 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 $\{\lambda, \sigma, \delta, \gamma_0, \gamma_1, \gamma_2\}$. In addition, we estimate 15 values $\psi_2, \ldots, \psi_{16}$ of holding a ticket at the time of the respective draw (recall that we have imposed that the weighted average value of all actually sold tickets in our data is 2 euros). Thus the full set of structural parameters to be estimated is $\theta \equiv \{\lambda, \sigma, \delta, \gamma_0, \gamma_1, \gamma_2, \psi_2, \ldots, \psi_{16}\}$.

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 7.7. Starting from this we 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. Section A.3 below provides details. 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.

[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 7, we have described the model only for one draw. Within each draw, t runs from 1 to the draw-specific T.

Two constraints are imposed when the objective function is minimized. First, we assume that the weighted average of ψ-s across all 16 draws is 2. That is, $\sum_{i=1}^{16} w_i \psi_i = 2$ with $w_i = \frac{s_i}{\sum_{i=1}^{16} s_i}$ and s_i is the total sales of draw i. For a given set of $\{\psi_2, \ldots, \psi_{16}\}$ one can compute the implied ψ_1 as $\psi_1 = \frac{2 \cdot \sum_{i=1}^{16} s_i - \sum_{i=2}^{16} s_i \psi_i}{s_1}$. We impose that $\frac{2 \cdot \sum_{i=1}^{16} s_i - \sum_{i=2}^{16} s_i \psi_i}{s_1} > 0$ to ensure that the implied value of holding the first lottery ticket is also positive. Second, since we specify a quadratic functional form of probability of consider: $P_{it}(\text{consider}) = \frac{1}{1+\exp\left(-(\gamma_0 + \gamma_1 g_{it}^a + \gamma_2 (g_{it}^a)^2)\right)}$. We impose that $-\gamma_1 - 6\gamma_2 < 0$ so that $P_{it}(\text{consider})$ is an increasing function of goodwill stock up to at least 3.

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

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

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 in the moments $\mathbb{E}[m(z_t, \hat{u}_t(\theta))] = 0$ 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.

To implement this 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. 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

[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.

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

We follow the usual two-step GMM procedure. We first choose the weighting matrix W to be

$$W = ((\frac{1}{\tau}\sum_t m(z_t, \hat{u}_t(\theta_s))m(z_t, \hat{u}_t(\theta_s))/\tau)^{-1}$$

to get consistent estimates $\hat{\theta}_s$ where θ_s is a vector of starting values that we chose. Then we calculate the optimal weight matrix at $\hat{\theta}_s$: $W^o(\hat{\theta}_s) = (\frac{1}{\tau}\sum_t m(z_t, \hat{u}_t(\hat{\theta}_s))m(z_t, \hat{u}_t(\hat{\theta}_s))')^{-1}$ and then get the efficient estimates.

A.4 Smoothing

When using a simulation-based procedure to estimate a discrete choice model, one common challenge is that the simulated choice probabilities or, in our case, simulated demand, is not a smooth function in the parameters. This is due to the 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 her decision (which stays at 0 or 1), or changes it discretely. 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]

[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.

B Robustness

In this appendix, we assess how robust our parameter estimates are to assuming a different market size (Appendix B.1), to assuming a different baseline probability of paying attention (Appendix B.2), and to allowing for serial correlation in viewership behavior (Appendix B.3). We also assess how these alternative specifications affect our results of counterfactual experiments (Appendix B.4).

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 300,000. We continue to impose that $\gamma_0 = -4.5$ and that the weighted average value to holding a ticket is 2 euros and, focusing on baseline sales for the moment, expect that the higher market size will lead to estimates that imply a lower probability to buy given consideration. As discussed in Section 7.6, the evolution of sales over time is closely related to the discount factor, which we do not expect to change much for that reason. But when the variance of the taste shocks, that is directly related to σ, decreases, then the model will produce lower sales given consideration (as the average value to holding a ticket is 2 euros and the price is 3 euros, so in the model consumers always buy due to taste shocks; in fact more so the less time there is until the draw).

parameter estimates. We find that indeed, σ is estimated to be lower. This produces choice probabilities that are roughly 80% as big as for our baseline model with a market size that is 20% bigger than the baseline case. As expected, the estimate of the depreciation rate of the advertising goodwill stock and of the discount factor are almost unchanged. γ_1 and γ_2 are slightly different, but in fact the implied percentage point increase in the probability to consider buying when we change the goodwill stock is very similar across those two specifications.

B.2 Assumption on probability to consider buying

Next, we assess the robustness to making alternative assumptions about the baseline probability to consider buying in the absence of advertising. For this, we re-estimate the model imposing that $\gamma_0 = -2.944$ so that the probability to consider buying without advertising increases from 1% (baseline specification) to 5%. We continue to impose that the weighted average value to holding a ticket is 2 euros and, focusing on baseline sales for the moment, expect that the higher baseline probability to consider buying will lead to estimates that imply a lower probability to buy given consideration.

Table 4 shows resulting parameter estimates. Similarly to the specification of 120% market share, the model produces a lower estimate of σ, which gives rise to lower sales given consideration and thereby offsets the higher probability to consider buying. The estimated depreciation rate of the goodwill stock is higher. The estimate of the discount factor and of γ_1 and γ_2 are

Table 4: Robustness checks

parameter	(1) baseline specification	ste.	(2) 120 percent market size	ste.	(3) increased γ_0	ste.	(4) generalized model	ste.
depreciation rate goodwill stock (λ)	0.530	0.028	0.557	0.029	0.765	0.045	0.450	0.022
effect of goodwill stock on probability of considering (γ_1)	3.672	0.129	4.055	0.254	3.487	0.855	3.545	0.063
square effect of stock on probability of considering (γ_2)	-0.599	0.116	-0.676	0.213	-0.560	0.834	-0.576	0.063
hourly discount factor (δ)	0.996	0.000	0.995	0.000	0.994	0.000	0.995	0.000
multiplying factor taste shock (σ)	0.511	0.018	0.482	0.015	0.418	0.006	0.519	0.015
value to having a ticket on the day of the draw								
10 February, 2014	1.512	0.104	1.609	0.132	1.691	0.022	1.444	0.066
10 March, 2014	1.516	0.084	1.598	0.075	1.665	0.041	1.587	0.085
10 April, 2014	1.483	0.066	1.501	0.079	1.566	0.060	1.526	0.074
26 April, 2014 (King's Day)	2.427	0.108	2.467	0.093	2.401	0.034	2.374	0.086
10 May, 2014	1.658	0.054	1.649	0.045	1.649	0.055	1.549	0.084
10 June, 2014	1.712	0.038	1.739	0.038	1.787	0.047	1.784	0.043
24 June, 2014 (Orange draw)	2.157	0.066	2.114	0.097	2.224	0.045	2.096	0.043
10 July, 2014	2.066	0.049	2.106	0.059	1.995	0.039	2.058	0.084
10 August, 2014	0.880	0.101	1.026	0.083	1.174	0.085	1.007	0.082
10 September, 2014	1.822	0.045	1.800	0.041	1.743	0.050	1.781	0.077
1 October, 2014 (special 1 October draw)	1.757	0.071	1.792	0.083	1.841	0.051	1.816	0.081
10 October, 2014	1.740	0.047	1.824	0.045	1.708	0.059	1.7565	0.056
10 November, 2014	1.335	0.069	1.335	0.071	1.452	0.050	1.424	0.054
10 December, 2014	1.669	0.070	1.770	0.081	1.942	0.062	1.704	0.063
31 December, 2014 (New year's eve draw)	3.186	0.104	2.979	0.069	2.859	0.057	3.130	0.094

Notes: Structural estimates. See Section 7.7 and Appendix A for details on the estimation procedure. Estimates for the baseline specification in Table 2 are repeated in column (1). The second set of parameter estimates was obtained under the assumption that the market size is 300,000 instead of 250,000. The fourth set of estimates is for the generalized model with serially correlated viewership behavior described in Appendix B.3. The probability to consider buying is specified as $P_{it}(\text{consider}) = 1/\left(1+\exp\left(-(\gamma_0+\gamma_1 g^a_{it}+\gamma_2(g^a_{it})^2)\right)\right)$. For specifications (1), (2), and (4), we set $\gamma_0 = -4.5$. The fourth set of estimates was obtained by setting $\gamma_0 = -2.944$ so that the baseline probability of paying attention is 5%. Throughout, we impose that the weighted average value to having a ticket on the day of the draw is 2 euros, using weights proportional to actual sales. The implied values for the value to holding a ticket on 10 January are 1.686, 1.742, 1.649 and 1.779, respectively.

almost unchanged.

B.3 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 and by re-estimating it for a given degree of serial correlation.

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

$$\Pi = \begin{bmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{bmatrix}.$$

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 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 is rationalized by a model in which viewership demand is serially correlated. The results show that this can be achieved by a slightly higher depreciation rate of the goodwill

[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.

stock. Moreover, the multiplying factor of the taste shock increases, which tentatively leads to higher predicted sales given consideration. The remaining parameters are almost not affected.

B.4 Counterfactual experiments with different specifications

In this sub-section, we assess how these alternative specifications will affect our results of counterfactual experiments. For this, we re-produce the result of counterfactual experiments in Section 9 using three different specifications. In all three specifications, we assume the 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 results thus correspond to the ones in the first column of Table 3.

Table 5 shows the result. The result for the baseline specification is repeated in the first column. Column (2) displays the result using the set of estimates that we obtain when we specify the market size to be 120 percent of what we used before. Results for a baseline probability to consider buying of 5% are reported in column (3). Finally, column (4) shows the results using the estimates of the generalized model. Overall, we find similar magnitudes across different specifications.

Table 5: Robustness checks: counterfactual experiments with alternative specifications

strategy	(1) baseline specification	(2) 120 percent market size	(3) increased probability of paying attention	(4) generalized model
data (reference point)	100%	100%	100%	100%
no advertising at all	64%	58%	74%	60%
all advertising in the last 2 days before the draw	147%	153%	140%	144%
spreading advertisements equally in the last 4 days before the draw	124%	123%	120%	117%
pulsing strategy in the last 4 days before draw (1 hour blocks)	122%	121%	120%	116%
pulsing strategy in the last 4 days before draw (2 hour blocks)	119%	118%	118%	112%
pulsing strategy in the last 4 days before draw (1 hour double, 3 hour none)	117%	117%	121%	110%
shift advertising from third week to fourth week	117%	114%	114%	110%
shift advertising from fourth week to third week	81%	71%	80%	74%

Notes: This table shows the results of the counterfactual experiments in Section 9 using three alternative specifications. In all three specifications, we assume the 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.

C Additional tables and figures

Table 6: 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 7: Evidence from a distributed lag model at the hourly level

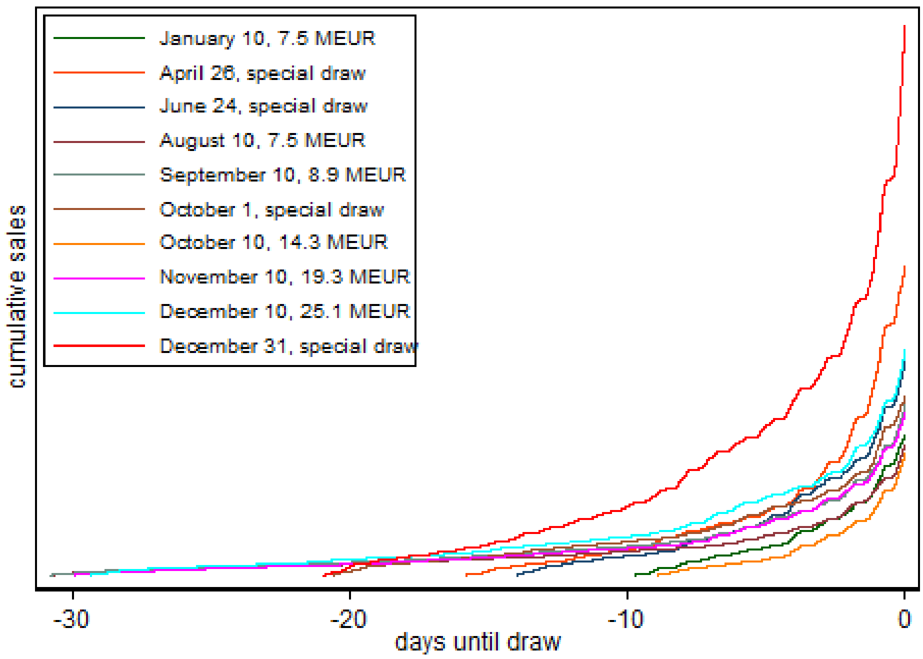
	(1) baseline	(2) night dummy
GRP current hour	0.0120***	0.0198***
	(0.00145)	(0.00160)
GRP 1 hour lagged	0.0120***	0.0128***
	(0.00133)	(0.00115)
GRP 2 hours lagged	0.00428**	0.00414**
	(0.00134)	(0.00127)
GRP 3 hours lagged	0.00412*	0.00879***
	(0.00168)	(0.00191)
draw dummies	Yes	Yes
days to draw dummies	Yes	Yes
hour of day dummies	Yes	No
night dummy	No	Yes
Observations	7662	7662
R^2	0.917	0.863

Standard errors in parentheses

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

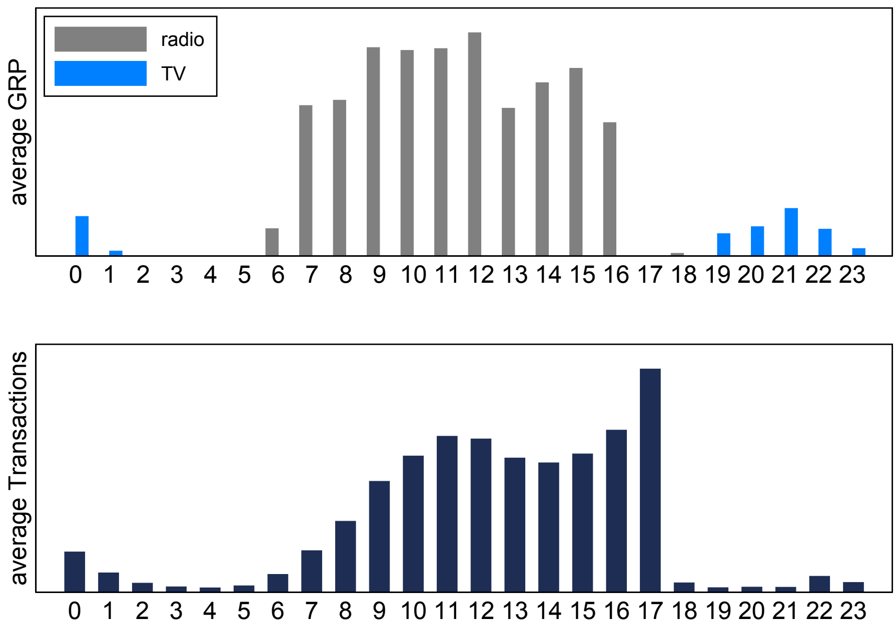
Notes: This table shows the results of regressions of the log of one plus sales on GRP's of advertising and lags thereof. Regressions were carried out at the hourly level and standard errors are clustered at the daily level.

Figure 10: Cumulative sales for remaining draws



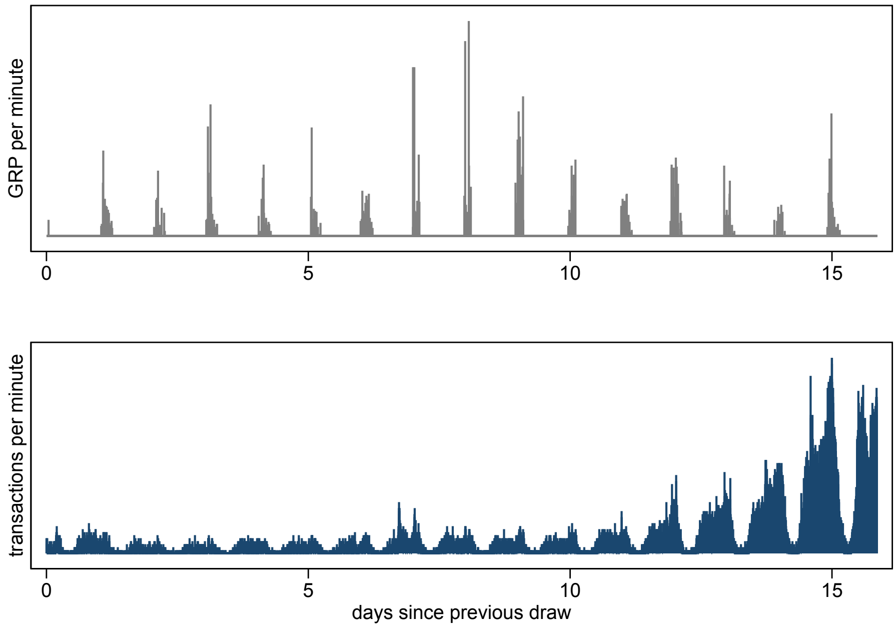
Notes: Figure 1 shows cumulative sales for 6 selected regular draws. This figure shows them for the remaining draws.

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



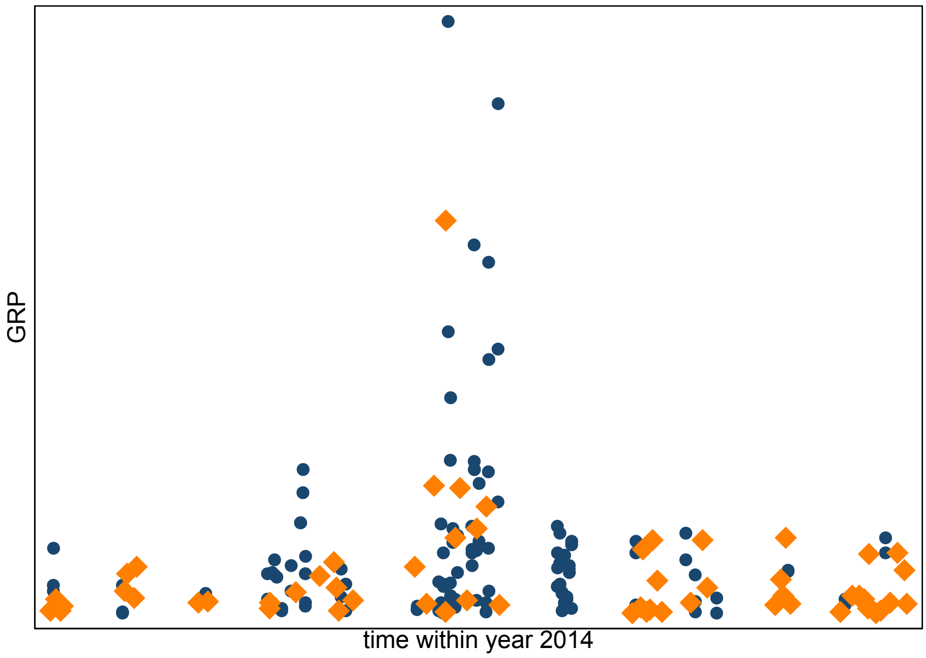
Notes: This figure shows average GRP's and sales for different times of the day. To produce this figure we first aggregate sales at the hourly level and then average over draws. On the day of the draw tickets for this draw can only be bought until 6pm. See Figure 2 for the pattern on the remaining days.

Figure 12: GRP's at the minute-level for a special draw



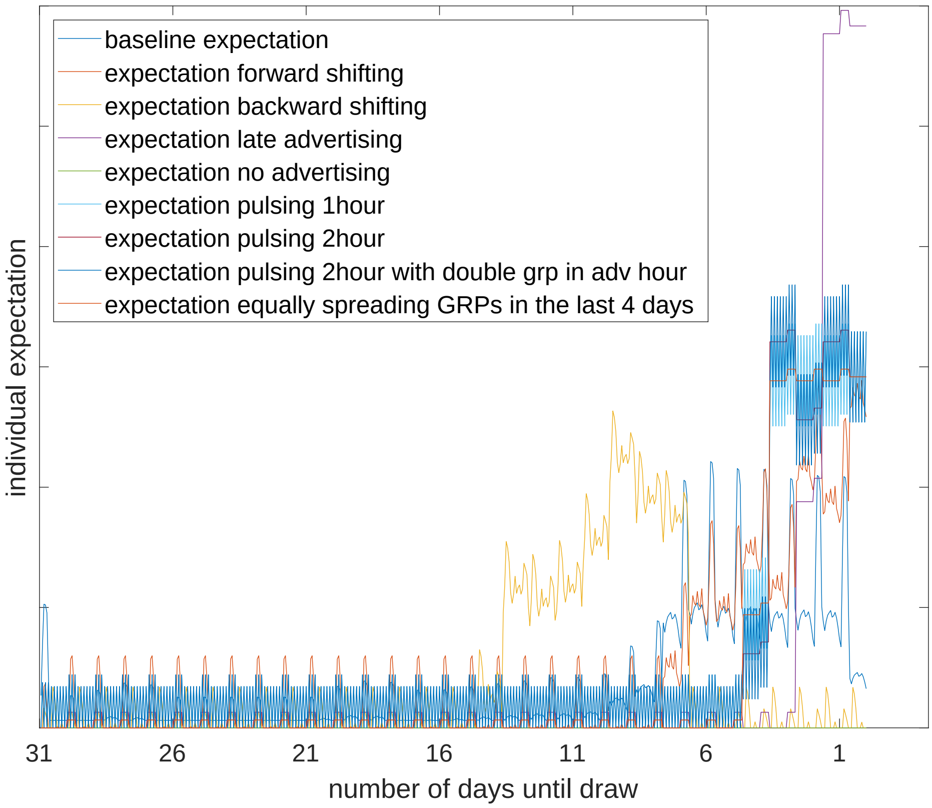
Notes: Figure 3 shows GRP's and sales for the draw on April 10, 2014. This figure shows GRP's and sales at the minute level for the special draw on April 26, 2014 (King's Day). The last regular draw took place on April 10, 2014. Tickets for the next draw can be bought from 6pm on the day of the previous 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 probability to see an advertisement. Obtained from regression of GRP's on hours of a day dummies, days until draw dummies for the draw in February.