Optimizing Online Sales using Targeted Advertising*

Chen He †

This Version: June 2017

Abstract

Advertising can affect consumer behavior at the consideration and the purchasing stage. This paper uses high frequency data on TV and radio advertising from different channels together with website traffic and online sales data to measure the effects of advertising. The high frequency nature of the data allows me to cleanly identify these effects and to show that they depend on the channel on which the firm advertises. I find positive effects of advertising on consideration and conversion that last for up to 4 hours. I then point out that the observed increase in the conversion rate could be due to the fact that those who are motivated to visit the website through advertisements are different from those who usually visit. The former ones have a higher probability to buy given that they visit. Ignoring this and studying consideration and conversion separately could result in an underestimated conversion rate and thus a suboptimal advertising strategy, in particular when advertising on different channels reaches different audiences. Motivated by this, I propose and estimate a new integrated model of consideration and conversion. In the model, consumers first decide whether or not to visit the website. This decision is driven by an option value. Importantly, this option value is allowed to depend on unobserved consumer characteristics. Therefore, unlike standard discrete choice models, the model allows consumers who visited the website to have a higher probability of buying than those who did not visit the website and consequently it can generate the observed pattern in the data even if advertising had no direct effect on consumers once they visit the website. My estimates show that one would overestimate both the effects of advertising and the cost of visiting the website if one would ignore this selection. Finally, I show that shifting advertising across channels could lead to increased sales.

Key words: purchase funnel, targeted advertising, online sales, consumer heterogeneity.

JEL-classification: M37.

*I would like to thank my advisors Tobias Klein and Bart Bronnenberg for their invaluable help and support. I am also grateful to Blue Mango Interactive and the Dutch State Lottery for providing me with the data. The views expressed in this paper, however, do not necessarily reflect those of Blue Mango or the Dutch State Lottery. All remaining errors are my own.

†He: CentER, Tilburg University; e-mail: C.He@uvt.nl.

1 Introduction

Advertising has for a long time been one of the most important instruments for firms to market their products and as a result, they spend a large amount of money on advertising. The advertising industry has even been expanding in recent years. Globally, media owners advertising revenues grew by 5.7% in 2016, to $493 billion (Letang and Stillman, 2016). TV advertising revenues grew by nearly 4% in 2016, reaching a market share of 38% (Letang and Stillman, 2016). Moreover, the trend shows that the industry is gradually and surely shifting to more Internet-based targeted ads and the "pay-per-click" model for online advertising, from the traditional "pay-per-impression" based ads. According to the latest industry forecast, digital-based ad sales will become the number one media category in 2017, reaching a market share of 40% (Letang and Stillman, 2016).

Though important, understanding how a firm should allocate its advertising budget remains a difficult question. Broadly speaking, at least two aspects that the firm needs to consider when it comes to allocating its advertising budget are when to advertise and where to advertise. Regarding the first aspect, He and Klein (2016) document that the timing of advertising is important and by advertising at the "right time", the firm could improve its profit for given advertising budget.

This paper addresses the second aspect: on which channel should the firm air its advertisement? This is important because the audiences are different across channels and therefore respond differently to advertising. I study this question in the context of the online sales, where TV and radio advertising can affect individual behavior in two stages, i.e., it can affect website traffic and conversion (the probability of buying conditional on visiting). I use high frequency data on TV and radio advertising from different channels together with online sales and website visits data to measure the effects of advertising. The high frequency nature of the data allows me to cleanly identify these effects and to show that they depend on the channel on which the firm advertises.

The contributions of this paper are threefold: first, I show that advertising causally increases online traffic as well as online conversion rate and that the effects depend on the channel on which the advertising is aired. This implies that targeting indeed matters. Second, I point out that the observed increase in the conversion rate could be due to the fact that those who are motivated to visit the website through advertisements are different from those who usually visit. The former ones have a higher probability to buy given that they visit. Ignoring this and studying consideration and conversion separately could result in an underestimated conversion rate and thus a suboptimal advertising strategy, in particular when advertising on different channels reaches different audiences. Motivated by this, third, I specify and estimate a new integrated model of consideration and conversion, in which the effects of advertising are channel-specific.

Crucially, simultaneously modeling the individual decisions on visiting and purchasing in an integrated two-stage model allows me to distinguish the conversion rate between two groups:

those who are motivated to visit by advertisements and those who usually visit. When the manager lays out a one-stage model only for purchasing decision and estimates the conversion rate using only data of those who visit, then what she gets is the average conversion rate: average between those who go on the website just like that and those who go on the website because they see an ad. But the one that matters for the manager is the one for those who actually go on the website because they see an ad since the other group's visiting decision is not affected by advertisements. *Only simultaneously modeling two-stage decisions* then allows me to study the effects of ads on the conversion rate for the treatment group.

In the model, consumers first decide whether or not to visit the website. This decision is driven by an option value. Importantly, this option value is allowed to depend on unobserved consumer characteristics. Therefore, unlike standard discrete choice models, the model allows consumers who visited the website to have a higher probability of buying than those who did not visit the website and consequently it could generate the observed pattern in the data even if advertising had no direct effect on consumers once they visit the website. My estimates show that one would overestimate both, the effects of advertising and the cost of visiting the website if one would ignore this selection. Using the model structure and the obtained estimates, I demonstrate that shifting advertising across channels could lead to increased sales.

The empirical context of this article is the market of the Dutch State Lottery. Studying the effects of advertising in the context of lotteries is promising. A lottery ticket is a simple product whose attributes are generally known or easy to describe. In addition, the market for lottery tickets in the Netherlands is very concentrated and therefore it is not unreasonable to assume that the firm acts as a monopoly in this market and therefore I could abstract away from supply side complication. Moreover, the website is very simple. It essentially offers only one product. This means that consumers can obtain little value from window shopping and thus I can reasonably assume that if a consumer visits the website because of watching the ad, she seriously considers buying a ticket. This is in contrast to those more general websites (e.g., Amazon) that offer many products and thus they are still worth visiting because of the value obtained from window shopping, even without purchasing.

This study is most closely related to He and Klein (2016). There are two major differences: first, He and Klein (2016) answer the question when the firm should advertise. For that purpose, they estimate a dynamic model that abstracts from targeting and makes no distinction between the two stages, consideration and purchasing. Here, I instead develop a static model of consumer choice and study on which channel should the firm air its advertisements. I simultaneously study the effects of advertising on both consideration and conversion. The decomposition of online sales into the two sub-stages (visit and purchase given visit) is useful for marketers. The first stage is informative about how many consumers can the advertisements attract. The second stage answers the question whether advertising can attract the right consumers.

Besides He and Klein (2016), this study is also related to several strands of the literature. The marketing literature has described consideration and conversion as different stages of the purchase funnel. Hoban and Bucklin (2015) show that display advertising positively affects visitation to the firm's website for users in most stages of the purchase funnel, but not for those who previously visited the site without creating an account. Lodish et al. (1995) show that TV advertising could increase sales, but not always. Sherman and Deighton (2001) show that banner advertising leads to more site visits. The authors also show that targeting can improve CTR. Manchanda et al. (2006) show that banner advertising increases online purchase. Haans et al. (2013) find that click-through rates are higher for advertisements involving expert evidence and statistical evidence than for those involving causal evidence, but the latter leads to a higher conversion rate. In the world of TV advertising, Kitts et al. (2014) show that TV advertising can increase the number of new visitors to a brand's website. More closely related, Liaukonyte et al. (2015) use a quasi-experimental design to show that TV advertising triggers website visitation and online shopping, and the effect crucially depends on the content and media placement of the advertisement. In terms of comparing the relative effectiveness of advertising channels, Danaher and Dagger (2013) find that catalogs, television, and direct mail most strongly influence sales and profit, followed by radio and newspaper. The findings in my study are consistent with those in Tellis et al. (2000) and Chandy et al. (2001). The authors find that TV advertisements lead to more consumer telephone calls, but their effects dissipate very rapidly. My study contributes to this strand of literature by investigating the effects of ads on consideration and conversion simultaneously in an integrated dataset. This is possible since, in my data, website visits and online sales come from the same group of consumers.

From a modeling perspective, this study is related to the literature on the effect of advertising on choice sets. The way in which advertising affecting consumer choice is related to Dubé et al. (2005) who model advertising to enter the flow utility of consumers non-linearly with diminishing marginal return of additional unit of advertising. I generalize their model by allowing the effect of advertising to be different across channels. My study is also related to the search and choice set literature because visiting the website can be viewed as a proxy of including the product into the choice set. The literature has found different ways to address the challenge that choice sets of consumers are usually unobserved (to the researcher) which can be seen as a missing data problem. Bronnenberg and Vanhonacker (1996), Ackerberg (2001) and Albuquerque and Bronnenberg (2009), among others, use auxiliary information such as past purchases. Kim et al. (2010, 2016) treat the choice set as the result of a process of sequential search. In their model, the consumer will search an additional option if the marginal benefit of searching is larger than the marginal cost of searching. My study shares the same spirit as theirs. In my model, consumers will visit the website if and only if doing so is better than not visiting. Roberts and Lattin (1991) develop a two-stage model of consideration and choice. However, their model does not feature advertising. Goeree (2008) directly augment the product choice model with a model of choice set formation. In her model, the probability that a consumer considers a given brand is a function of her demographics and advertising. Since she cannot observe the choice set, she estimates the model using simulated choice sets. Draganska and Klapper (2011) combine micro-level survey data on brand awareness with demand and advertising data to estimate an aggregate discrete choice model. They use consumer survey data of brand awareness to construct the choice set. They find evidence that advertising has a direct effect on the probability of inclusion in the choice set in addition to its effects on consumers preferences. Their paper is one of the very few cases where choice sets are observed by researchers. Somewhat differently, Clark et al. (2009) find that advertising has a significant positive effect on brand awareness but no significant effect on perceived quality.

One of the main topics of this article is targeted advertising. The previous literature has investigated this topic using descriptive approach. Goldfarb and Tucker (2011a) use a large-scale field experiment to show that targeted online advertisements increase purchase intent. In a related study, Lambrecht and Tucker (2013) find that online targeted advertisements have a positive effect on consumers with narrowly construed preferences. From the supply side, Goldfarb and Tucker (2011b) show that advertisers are willing to pay more for targeted search advertisements. In addition to developing a model, my study contributes to this strand of the literature by showing how targeted offline advertising can help improve online sales.

The rest of this paper is structured as follows. Section 2 gives a brief overview of the background information for lottery tickets in the Netherlands. Section 3 describes the data and shows descriptive evidence. Section 4 shows additional descriptive evidence on the effect of advertising on visits and sales. Section 5 develops the model of lottery ticket demand with advertising effects. Section 6 presents the results. Section 7 performs counterfactual experiments for the supply side. Finally, Section 8 concludes.

2 The market for lottery tickets in the Netherlands[1]

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

The lottery run by Staatsloterij is classical. A ticket has a combination of numbers and Arabic letters and a consumer can choose some of them. The size of the prize depends then

[1] The content in this section overlaps with He and Klein (2016). I include it here so that the readers do not need to refer to the other paper.

on how many numbers and letters of a ticket match with the ones of the winning combination. On top of that, there is a jackpot whose size varies over time. For all draws but the very last one in a year, consumers can choose between a full ticket that costs 15 euros and multiples of one fifth of a ticket. For the last draw, the price of a ticket is 15 euros and consumers can buy multiples of one half of a ticket. Winning amounts are then scaled accordingly. The tickets can be purchased in two ways: they can either be purchased online via the official website of Staatsloterij, or offline, for example, in a supermarket or a gas station. About 80 percent of the sales are offline.

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

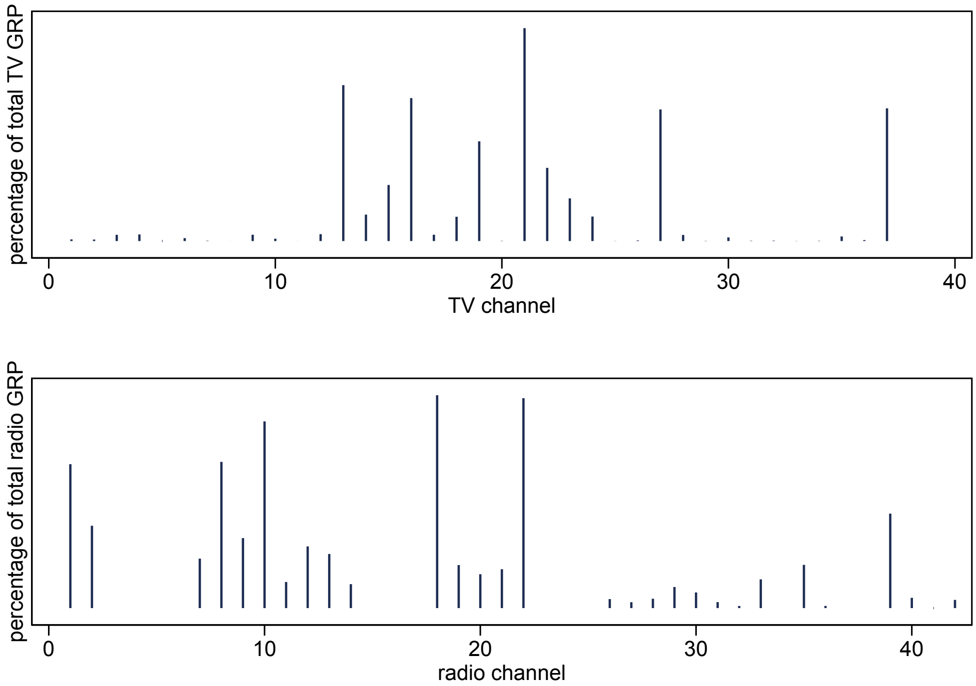
3 Data and descriptive statistics

3.1 Overview

The data consists of 3 parts: the number of visits to the seller's website, the number of online sales and TV and radio advertisement. All of them are measured at the minute level. The advertising data is measured in terms of Gross Rating Points (GRP's), separately for each TV and radio channel. GRP's measure impressions of the target population. More specifically, GRP's is defined as the percentage of people that have been reached times$\times$ the average frequency of the reach. For example, 5 GRP's mean that 5 percent of the target population (in this case the general population) are reached by the ads. It can also mean that 2.5 percent of the target population is reached twice. This is a standard measure in the advertising industry. Measuring ads using GRP's has the advantage over the traditional measure of advertising in terms of monetary expenditures since it gives a precise measure of the percentage of the target population that has been reached. Related to my identification strategy, the GRP's in the data are the actually delivered GRP's, rather than the contracted GRP's. I will return to this in section 4.

Besides, I observe the jackpot size for the 12 regular draws in 2014. There is no information on jackpot size for the 4 special draws, as more involved rules apply to them. This makes it difficult to calculate an equivalent one-shot jackpot size. For example, on the drawing day, every 15 minutes consumers can win an additional 100,000 euros. In the empirical analysis, I will capture differences across draws in a flexible way. Throughout the paper, I am not allowed to report levels of visits and sales and advertising. Therefore, I will only present relative numbers and (semi-) elasticities in the tables and figures below and some vertical axis will have no units of measurements.

Figure 1: distribution of GRP's across channels



Notes: This figure shows the distribution of GRP's across different channels. TV (radio) GRP's are computed as percentage of total TV (radio) GRP's.

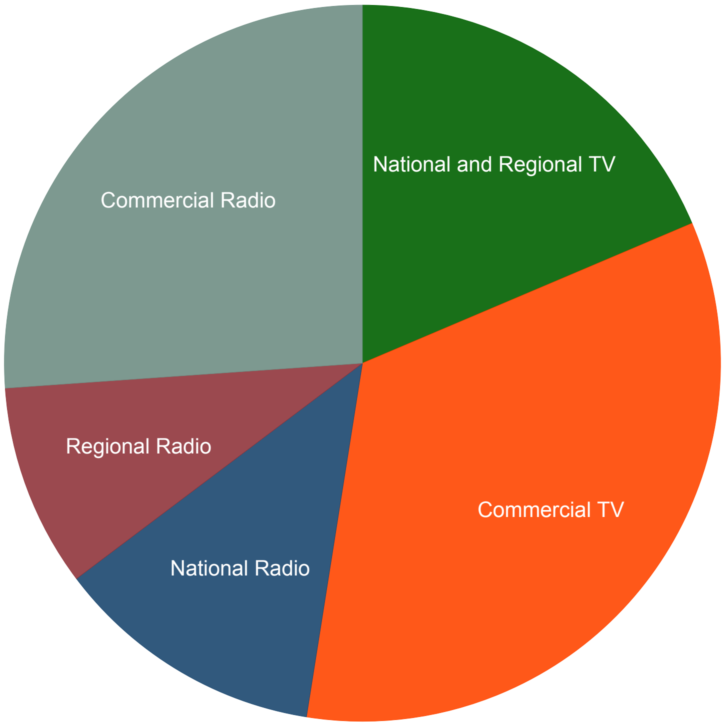
3.2 Descriptive evidence

Figure 1 shows the distribution of GRP's across different channels. One can see that the firm diversifies its budget over a large number of channels and there are many channels where the firm only spend few GRP's. This motivates me to classify all channels into different groups based on channel characteristics. I assign all channels into 5 groups: Group 1 consists of public and regional TV stations. Group 2 is commercial TV stations. Group 3 is the public radio stations. Group 4 consists of regional radio stations. Finally, group 5 is commercial radio stations.[2] Two criterions are used in this classification. First, after classification, the audiences across each channel should be different in terms of their interest in lottery tickets and the audiences within one channel should be similar. Second, after classification, the total number of spending for GRP's should not differ too much across different groups. I do not have data on channel specific viewership and thus could not verify the first criterion. I verify the second criterion using Figure 2, which shows the share of GRP's after the combination of channels. One can see that the firm spends most of the budget on commercial TV and radio channels.

Table 1 shows the mean and various rescaled percentiles for GRP's by each channel group.

[2]The reason that I do not separate public and regional TV stations into 2 groups is that the GRP's spent on regional TV stations are very few compared with other groups. I thus put public and regional TV stations into one group.

Figure 2: Share of GRP's across channels



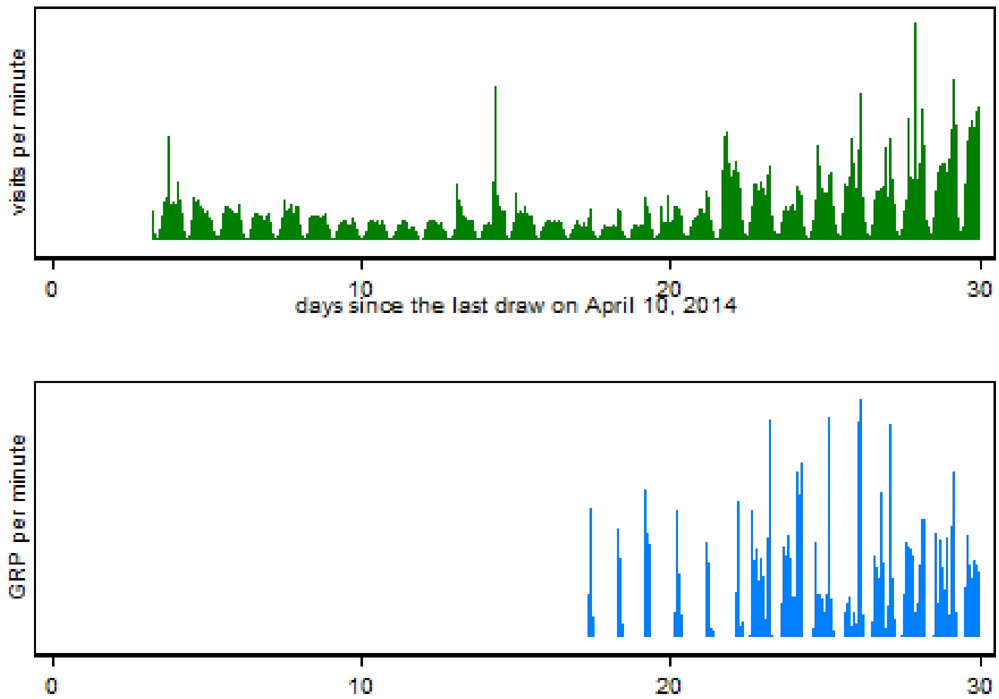
These numbers are rescaled by the overall average GRP if there is an ad. One can see that the mean size of GRP for an ad is larger on TV and national radio channels than on other two channels. This means advertising on TV and national radio channels are more effective in reaching people.

Figure 3 shows GRP's and visits at the minute level for one representative regular draw. First, notice that I disregard data for the first 3 days since the last draw. This is because the number of visits is extremely large in the first 3 days after the last draw. These site visits are generated by those who check online whether they have won in the last lottery. Clearly, those visits have nothing to do with advertising and thus are disregarded. In the reduced-form regressions and structural estimation, I also disregard the first 3 days of data for each draw.

Table 1: Rescaled percentiles for GRP's at the minute level

channel	5th	25th	50th	75th	95th	max	mean
national & local TV channel	0.07	0.28	0.63	1.74	6.74	41.74	1.67
commercial TVchannel	0.07	0.35	0.92	1.81	4.86	29.79	1.46
national radio channel	0.21	0.63	1.25	2.71	4.58	14.72	1.74
local radio channel	0.07	0.07	0.07	0.21	1.67	5.76	0.35
commercial radio channel	0.07	0.21	0.49	1.04	2.85	6.25	0.83

Figure 3: GRP's and visits at the minute-level for a regular draw



Second, we see that the firm only starts advertising on the 17th day after the last regular draw.

Figure 4 shows GRP's and sales at the minute level for the same draw. Unlike site visits, most of the sales occur during the last few days.

Next, figure 5 and 6 zoom in further and shows the pattern for one of the days in Figure 3 and 4. In both figures, the lower part depicts the GRP's. The higher the GRP spikes, the more people are reached. In particular, each color represents a different group of channels. It is interesting to notice that the raw data presented in Figure 5 and 6 has already shown some evidence of short-run site visits and sales responding to advertising. For example, there are some spikes of GRP's just before 20:50, followed by spikes of visits and sales several minutes later.

Finally, an interesting question is that, given consumers have visited the website, whether advertising could increase their probability of buying a ticket. That is, whether advertising could convert website traffic into sales. The variable of interest here is the probability of buying a ticket conditional on visiting. At first glance, it is tempting to measure this conditional probability by the number of sales over the number of visits at each minute. However, taking a closer look at Figure 3 and 4, one notices that the number of visits always responds to advertising faster compared to the number of sales. This is especially the case for the spikes around 22:13. In other words, there is a delay on sales response after advertising compared to visits. Moreover, such a delay is "random", which makes the ratio of sales over visits within the same minute meaningless. In the following section, I investigate the effect of advertising on

Figure 4: GRP's and sales at the minute-level for a regular draw

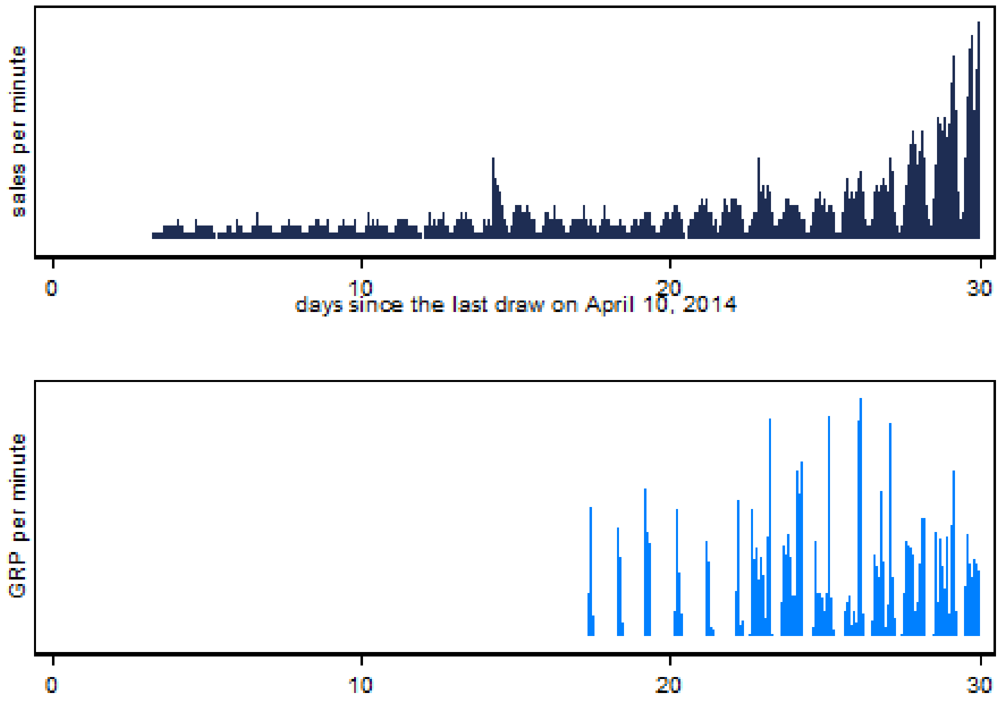


Figure 5: GRP's and visits at the minute-level for a short time window

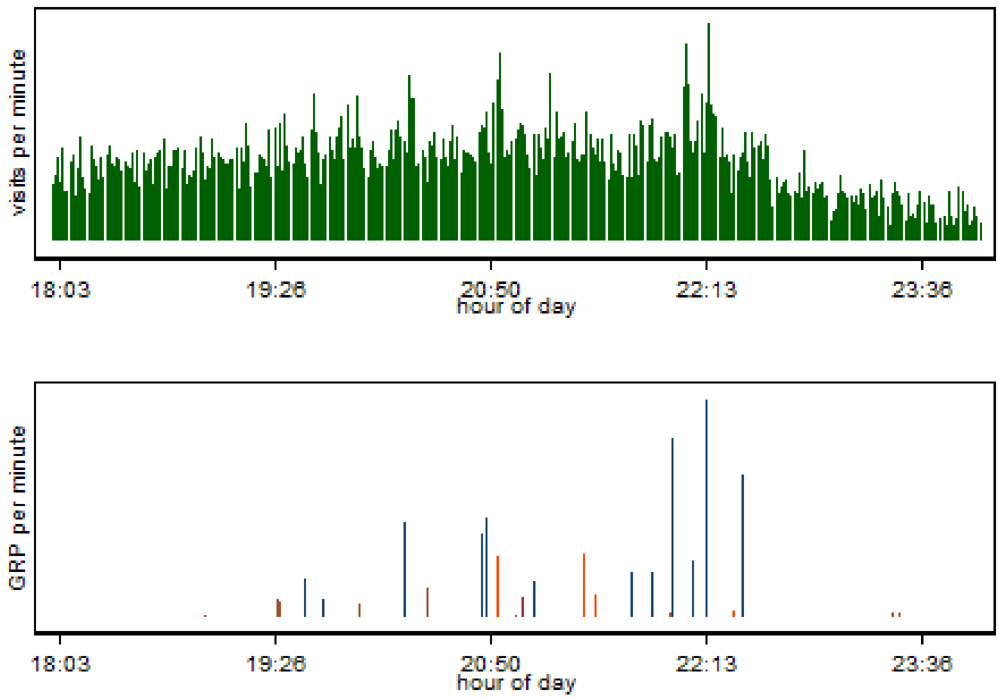
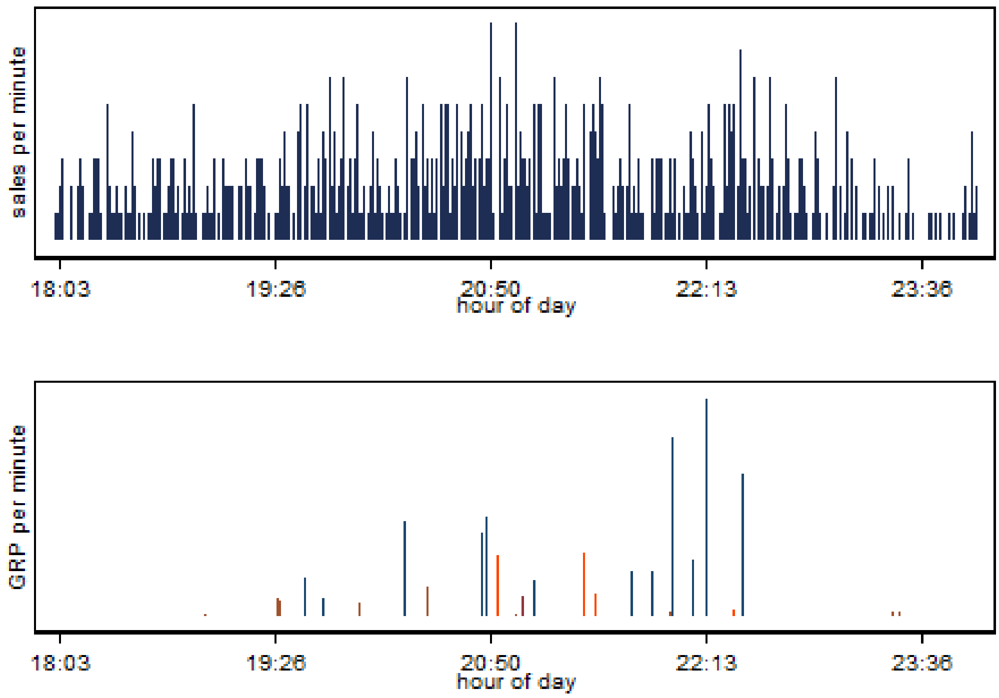


Figure 6: GRP's and sales at the minute-level for a short time window



conversion rate, taking into account the delay between the number of visits and sales.

4 Descriptive evidence

Motivated by descriptive statistics, I now characterize the short-term effects of advertising more systematically using regressions. I control for draw, time of the day and days until the draw fixed effects. The effect of advertising in this section has a causal interpretation. The establishment of causality relies on the industry practice called "make-goods", as explained in Dubé et al. (2005). The idea is that there is a difference between the contracted GRP's and the actually delivered GRP's. Although it is possible that the firm strategically chooses its desired GRP levels, which makes contracted GRP's endogenous, the actually delivered GRP's is random after controlling for draw, time of the day and days until the draw fixed effects. The intuition is that at a given minute in time, the instantaneous viewing rate for a particular show on a particular channel is random.

There is a related approach by Liaukonyte et al. (2015), who reconstruct baseline on the level of ads and then attribute the systematic differences between the pre- and post-ads windows to the ad insertion. The difference between their approach and mine is that I control for baseline effects using regression while Liaukonyte et al. (2015) reconstruct baseline on the level of ads (like matching estimator).

4.1 Effect of advertising on visits and sales

Throughout, I use a common regression framework: distributed lag model. A distributed lag model is a model in which I regress variables of interest on lagged amounts of advertising. A nice feature of distributed lag model is that it imposes little structure. I control for the draw, time of the day and days until the draw fixed effects. More precisely, I specify

$$y_t = \beta_0 + \underbrace{\Sigma_i^N \beta_{1i} \cdot grp_{t-i+1}}_{\text{lags of GRP's}} + \underbrace{x_t' \cdot \beta_2}_{\text{time and draw dummies}} + \varepsilon_t, \tag{1}$$

where x_t is a vector of dummy variables including draw, the hour of the day and days until the draw dummies. The dependent variable y_t is log of one plus the number of visits (sales). Notice that I do not distinguish GRP's from different channels in this specification. That is, I treat every unit of GRP's the same. The main interest here is to measure the effect of advertising, no matter where they come from, on the number of visits and the number of sales. Table 2 summarizes the result.

Column (1) shows the effect of advertising on visits. The main effect is observed in the first hour, but there are effects thereafter. The maximal effect is an increase in the number of visits of about 2.9 percent for each additional GRP of advertising, between 5 and 9 minutes after the advertisement was aired.

Next, Column (2) reports the effect of advertising on sales. Compared with column (1), I find that the maximal effect is an increase in sales of about 3.8 percent for each additional GRP of advertising, between 10 and 14 minutes after the advertisement was aired. Notice that the regressions show evidence on the delay of sales response: the maximal effect on sales is between 10 and 14 minutes after the advertisement was aired whereas it is between 5 and 9 minutes after the advertisement was aired for the number of visits. This means, on average, there is a delay of 5 minutes between visits and sales.

To summarize, I find advertising has a significant positive effect on both site visits and online sales. The effect of advertising on sales has an average delay of 5 minutes compared to visits.

4.2 Effect of advertising on online conversion rate

As documented in the previous subsection, there is a "random" delay between the time when consumers visit the website and the time when they actually make a purchase. This makes regressing sale-visit ratio on the distributed lags of GRP's meaningless. To measure the effect of advertising on online conversion rate, ideally one needs the stand-alone advertisement with no advertising before and after itself. This is crucial since the effect of advertising would otherwise overlap with each other in the presence of multiple advertising. Unfortunately, I do not

Table 2: The effect of advertising

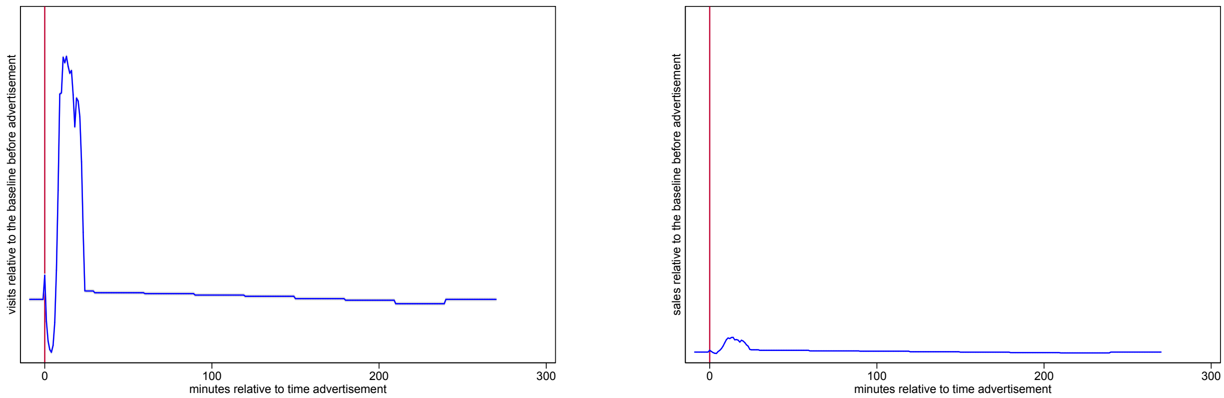
	(1) log(1+visits)	(2) log(1+sales)
GRP between 0 and 4 minutes ago	0.0241***	0.0167***
	(0.000918)	(0.00106)
5 and 9 minutes	0.0286***	0.0352***
	(0.000832)	(0.00106)
10 and 14 minutes	0.0106***	0.0382***
	(0.000650)	(0.000923)
15 and 19 minutes	0.00931***	0.0286***
	(0.000661)	(0.000966)
20 and 24 minutes	0.00918***	0.0239***
	(0.000682)	(0.000969)
25 and 29 minutes	0.00908***	0.0209***
	(0.000708)	(0.00105)
0.5 and 1 hour	0.00727***	0.0164***
	(0.000295)	(0.000420)
1 and 1.5 hours	0.00635***	0.0111***
	(0.000297)	(0.000413)
1.5 and 2 hours	0.00479***	0.00871***
	(0.000292)	(0.000423)
2 and 2.5 hours	0.00345***	0.00310***
	(0.000301)	(0.000366)
2.5 and 3 hours	0.000884**	-0.000795*
	(0.000297)	(0.000349)
3 and 3.5 hours	-0.000898**	-0.00565***
	(0.000287)	(0.000322)
3.5 and 4 hours	-0.00499***	-0.00940***
	(0.000286)	(0.000322)
draw dummies	Yes	Yes
days to draw dummies	Yes	Yes
hour dummies	Yes	Yes
Observations	441223	441223
R^2	0.841	0.655

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/visits on GRP's of advertising and lags thereof. Regressions were carried out at the minute level and standard errors are robust to heteroskedasticity.

Figure 7: Effect of a 10-GRP advertisement on site visits and online sales



have many stand-alone advertisements in the data.[3] Most of the advertisements stand close to each other, as can be seen from Figure 5. Thus, I measure the effect of advertising on online conversion rate in a counterfactual setting using the estimates in Table 2. More specifically, I first compute the effect of a 10 GRP's stand-alone advertisement on the number of visits. The interpretation of this number is the total number of extra visits due to the GRP's compared to baseline visits (without GRP's). On the left side of the Figure 7, this is the total area under the impulse response curse. Then, I compute a similar number for sales, which is depicted again on the right side of the Figure 7. The effect of advertising on conditional sales is then measured by the ratio of two.

Comparing this conversion rate with the average conversion rate over all periods, I find that the conversion rate is higher than that without advertising.[4]

As an alternative approach, I characterize the effect of ads on conversion rate using the same regression as in (1) , with y_t be the conversion rate. To overcome the aforementioned delay problem, I aggregate the data to the hourly level. The underlying idea is that the delay problem is eliminated after aggregation. As can be seen from Table 3, advertising increases conversion rate in the present and one hour in the future. The maximal effect is an increase of about 0.06 percent point for each additional GRP. Again, this difference in conversion rate is between those who are motivated to visit the website through advertisements and those who usually visit.

4.3 Effect of advertising across channels

To study the heterogeneous effect of advertising across channels, I extend the distributed lag model in (1) with interaction terms of channel dummies with GRP's and their lags, while controlling for the same set of dummy variables. More precisely, I estimate

[3]Regression with only stand-alone advertising would be subject to serious sample selection issue.

[4]I did not report the overall conversion rate because that would reveal information on the real data.

Table 3: The effect of advertising on conversion rate

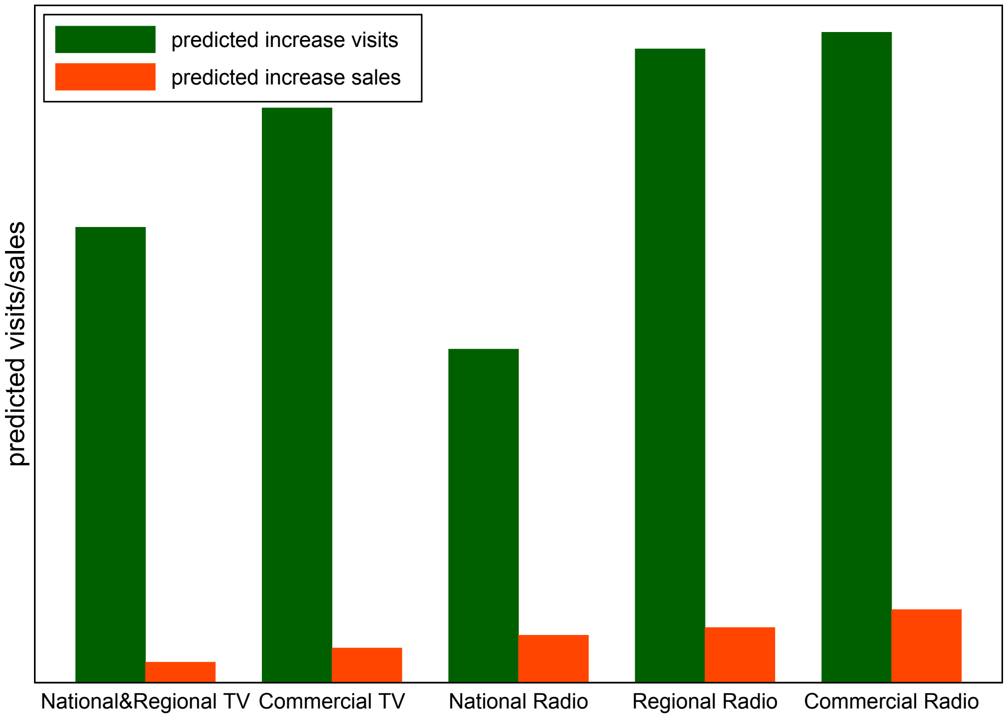
	(1) conversion rate
GRP the current hour	0.0237**
	(0.00898)
GRP one hour ago	0.0624***
	(0.0101)
GRP two hours ago	0.0108
	(0.00909)
GRP three hours ago	0.0157
	(0.00836)
GRP four hours ago	-0.00415
	(0.00738)
draw dummies	Yes
days to draw dummies	Yes
hour dummies	Yes
Observations	7531
R^2	0.773

Standard errors in parentheses

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

Notes: This table shows the results of regressions of the conversion rate on GRP's of advertising and lags thereof. Regressions were carried out at the hourly level and standard errors are robust to heteroskedasticity.

Figure 8: Effect of a 5-GRP advertisement on site visits and online sales for different channels



Notes: Group 1=National & Local TV; 2=Commercial TV; 3=National Radio; 4=Local Radio; 5=Commercial Radio.

$$y_t = \beta_0 + \underbrace{\Sigma_i^N \Sigma_j^5 \beta_{1\,ji} \cdot grp_{jt-i+1}}_{\text{interaction between GRP's and channels}} + \underbrace{x_t' \cdot \beta_2}_{\text{time and draw dummies}} + \varepsilon_t, \tag{2}$$

where grp_{jt} is GRP's from channel group j at time t. Figure 8 summarizes the reult. First, channel group 5 (commerical radio channels) is the most effective in attracting both online traffic and online sales. However, channel 3 has the highest conversion rate. One can find the the full table in the appendix A.

5 A model of lottery ticket demand

5.1 Motivation of estimating a model

There are three reasons that motivate me to estimate a model. First, as the descriptive evidence suggests, the online conversion rate increases after ads. This could be due to the fact that those who are motivated to visit the website through advertisements are different from those who usually visit. The former ones have a higher probability to buy given that they visit. Spelling out a two-stage model with correlated structure allows me to capture this selection. Therefore, the effects of ads on the conversion rate that the model captures are the difference between two group of visitors: Second, I can use the model to rigorously describe the idea that the firm faces a tradeoff: advertising on the more effective channels has a higher return, but at the same time that the marginal return of additional unit of advertising is diminishing. Last but not the least, once the structural parameters are estimated, I can evaluate various counterfactual targeting advertising strategies.

5.2 General structure

There are 5 channel groups. Within each $j = 1,2,...,5$ channel group, consumers are homogeneous in observed characteristics but are heterogeneous in their unobserved (to the econometrician) taste shocks. Moreover, consumers are assumed to watch one channel group. Consumers differ across each channel group in two ways. First, they differ in how much ads they have been reached. This can be seen directly from data since each channel group has different GRP's. Second, for given amount of advertising, individuals from different groups react differently to advertisements.

There are N expected discounted utility-maximizing consumers. Each consumer i comes from one of the 5 channel groups. Time $t = 1,2,\ldots,T$ is discrete and finite and measured at the hourly level. T is the last hour of the draw. In every hour, each individual has to make two sequential decisions. She first decides whether or not to visit the website. If she does, then she pays the cost of visiting the website (e.g., time cost of opening the website on a computer or smartphone). Otherwise, she receives the utility of outside option and continues in the next period and has the option of visiting the website there. After the individual has visited the website, she then decides whether or not to purchase a lottery ticket. If she does, then she receives a one-off flow of utility. Otherwise, she receives the utility of outside option and continues in the next period.

5.3 Advertising

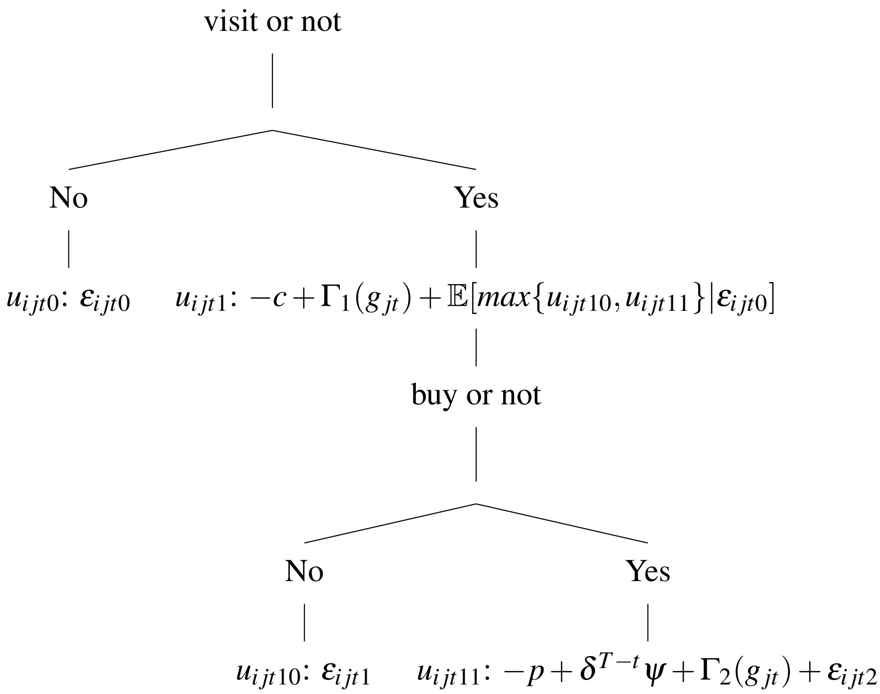
The flow utility from visiting the website and buying a ticket (described below) is modeled to depend on the advertising goodwill stock. Loosely speaking, the advertising goodwill stock summarizes how many ads that the consumer has been exposed and it will increases if the individual is exposed to an advertisement. The size of the increase depends on GRP's. Moreover, the goodwill stock depreciates over time.

More specifically, let the goodwill stock on channel j at the beginning of period t denote by g_{jt}. The firm has the opportunity to increase the goodwill stock by purchasing GRP's. The goodwill stock will depreciate at the beginning of the next period. Let λ denote the depreciation rate and assume that the initial goodwill stock is 0. The law of motion for the goodwill stock is

$$g_{jt} = \lambda g_{jt-1} + GRP_{jt}$$

and

Figure 9: Model summary



$$g_{j0} = 0 \ \forall j.$$

The specification is similar to that in Dubé et al. (2005). There are two differences. First, unlike Dubé et al. (2005) who model the GRP's to enter the goodwill stock non-linearly, I specify a linear goodwill stock production function. Moreover, I allow the goodwill stock to be different for each channel group j.

5.4 Consumers

Consumer i from channel group j at time t decides whether or not to visit the website of lottery tickets. Visiting the website yields flow utility

$$u_{ijt1} = \underbrace{-c}_{\text{cost of visiting the website}} + \underbrace{\Gamma_1(g_{jt})}_{\text{effect of ads on visiting}} + \underbrace{I_{ijt}}_{\text{option value}}, \tag{3}$$

where c is the cost of visiting the website, g_{jt} is the advertising goodwill stock and I_{ijt} is the option value of purchasing the lottery ticket, which will be explained later. As in Dubé et al. (2005), g_{jt} enters consumer's flow utility non-linearly. In particular, I specify

$$\Gamma_1(g_{jt}) = \gamma_{j1} \log(1+\gamma_3 \cdot g_{jt}). \tag{4}$$

4 links the model to the data. In particular, g_{jt} measures to what extent the consumers from channel j have been exposed to ads. The coefficient γ_{j1} measures that consumers from different

groups react differently to ads. Notice that the log functional form of $\Gamma_1(\cdot)$ implies diminishing marginal returns of advertising at a given point in time. This means the firm needs a tradeoff between a larger value of γ_{j1} and the diminishing marginal return of goodwill stock implied by the log functional form of $\Gamma_1(\cdot)$. The parameter in front of the goodwill stock, γ_3 , affects the curvature of the log function. The larger this parameter, the stronger diminishing marginal return of an additional unit of goodwill stock. If a consumer chooses not to visit the website, she receives the utility of outside option, which is normalized to 0, plus an unobserved taste shock ε_{ijt0}:

$$u_{ijt0} = 0 + \varepsilon_{ijt0}.$$

Visiting the website gives the consumer an option to buy a ticket. First, suppose that she chooses not to buy a ticket after visiting the website, she again receives a flow utility of outside option

$$u_{ijt10} = \varepsilon_{ijt1}.$$

Now suppose that she instead purchase a ticket, she receives flow utility

$$u_{ijt11} = -p + \delta^{T-t}\psi + \Gamma_2(g_{jt}) + \varepsilon_{ijt2},$$

where p is the price of the ticket, δ is the hourly discount factor and ψ is the value of holding a ticket at the time of the draw. The same as the case of flow utility of visiting the website, g_{jt} enters consumer's flow utility non-linearly:

$$\Gamma_2(g_{jt}) = \gamma_{j2}\log(1 + \gamma_3 \cdot g_{jt}).$$

Notice that I allow different effects of advertising on flow utility of visiting and buying given visiting. The taste shocks in flow utilities: ε_{ijt0}, ε_{ijt1}, ε_{ijt2} are assumed to be jointly multivariate normally distributed:

$$\{\varepsilon_{ijt0}, \varepsilon_{ijt1}, \varepsilon_{ijt2}\} \sim N(0, \Sigma)$$

with

$$\Sigma = \begin{bmatrix} 1 & \sigma_{01} & 0 \\ . & 1 & 0 \\ . & . & 1 \end{bmatrix}.$$

The covariance between ε_{ijt0} and ε_{ijt1}, σ_{01}, is what makes this model non-standard: it allows the taste shock in the visiting stage to be correlated with that in the purchasing stage.[5] From

[5]Notice that since I normalize the variance of taste shocks to 1. The covariance σ_{01} becomes the correlation coefficient and thus it can never be larger than 1 or smaller than -1.

the consumer's perspective, this means that the taste shock of the outside option at the visiting stage is informative about that at the purchasing stage. From the modeling perspective, those who enter the purchasing stage decision are selected by their taste shock. It will become clear later that only those with a small enough ε_{ijt0} choose to visit the website. Consequently, those who visited the website have a large probability of purchasing given visiting, which, in turn, generates the spikes observed in sales data. To see this more formally, consider the option value of visiting the website: I_{ijt}. By definition, it is the expected value of the maximum of flow utility between buying a ticket and not buying one:

$$I_{ijt} = \mathbb{E}_{\varepsilon_{ijt1},\varepsilon_{ijt2}}\left[\max\{u_{ijt10},u_{ijt11}\}|\varepsilon_{ijt0},g_{jt},T-t\right]. \tag{5}$$

Unlike standard models where the expectation operator is taken over the taste shocks unconditionally, here because of the non-zero σ_{01}, the expectation is taken conditional on the visiting stage shock ε_{ijt0}. In other words, the option value, I_{ijt}, is a function of ε_{ijt0}. The main motivation for the distributional assumption is computational: it can be shown that if the taste shocks are type-I extreme value distributed, then the conditional expectation has no closed form solution.[6] Another motivation is that the normal distribution, unlike the type-I extreme value distribution, is symmetric about its mean.

(5) has an economic interpretation: the consumer has taken into account the likelihood of buying a ticket when she decides whether or not to visit the website. If she knows for sure that she would not purchase a lottery ticket, then there is also little reason for her to visit the website. Put differently, those who are motivated to visit the website through advertisements are different from those who usually visit: they have a higher probability to buy given that they visit. In the model, this fact is captured by the correlation between ε_{ijt0} and ε_{ijt1} through the option value I_{ijt}. Only those consumers that belong to the set $\{\varepsilon_{ijt0}|u_{ijt1} > u_{ijt0}\}$will visit the website. The key elements of the full model are summarized in Figure 9 .

5.5 Discussion

Having spelled out the model, I give a short discussion on what the model meant to capture and how it captures. What the model meant to capture is the difference in conversion rate between two group of visitors: those who are motivated to visit by ads and those who visit anyway. The former ones are the treatment group and the latter ones are the control group. The model captures the difference between the two groups of visitors in the following way: the effect of ads on visiting decision, $\Gamma_1(\cdot)$, is the treatment. After the treatment, the increase in predicted visits due to $\Gamma_1(\cdot)$ comes from the treatment group. The effect of ads on conversion, $\Gamma_2(\cdot)$, then captures the increase in conversion rate for the treatment group relative to the control group. In addition, the effectiveness of advertising is channel specific. This is the source of variation in

[6]If ε_{ijt1} and ε_{ijt2} are drawn from type-I extreme value distribution and I_{ijt} is taken unconditionally over ε_{ijt0}, then it has the familiar log sum closed form solution.

the effectiveness across channels.

5.6 Solving the model

I now describe how to solve the model for given values of the parameters, which I then vary in the outer loop of the estimation procedure. In the following, I discuss the solution of the model in backward order. That is, I first discuss the solution of the model in the purchase stage, given that the consumer has visited the website. I then discuss the decision that the consumer face in the visiting stage.

The key insight for the purchasing stage decision is that conditional on having visited the website, the decision between buying and not buying a ticket is a binary probit choice. To see this, notice that it follows from result of multivariate normal distribution that conditional on ε_{ijt0}, the joint distribution of $\{\varepsilon_{ijt1}, \varepsilon_{ijt2}\}$ is also (bivariate) normally distributed:

$$\{\varepsilon_{ijt1}, \varepsilon_{ijt2}\}|\varepsilon_{ijt0} \sim \left(\begin{bmatrix} \sigma_{01}\varepsilon_{ijt0} \\ 0 \end{bmatrix}, \Sigma_{12} \right),$$

with

$$\Sigma_{12} = \begin{bmatrix} 1 - \sigma_{01}^2 & 0 \\ 0 & 1 \end{bmatrix}.$$

Also note that conditional on ε_{ijt0}, Σ_{12} implies $\varepsilon_{ijt1}$is independent from ε_{ijt2}. Joint normality of error terms means the purchasing stage decision is a standard binary probit model with uncorrelated error terms. This result dramatically reduces the computational burden of the model since it implies that, conditional on ε_{ijt0}, the choice stage decision can be solved without simulating integral. To see this, define $\tilde{u}_{ijt} \equiv u_{ijt11} - u_{ijt10}$, then the probability of buying a ticket, given having visited the website and ε_{ijt0} is given by

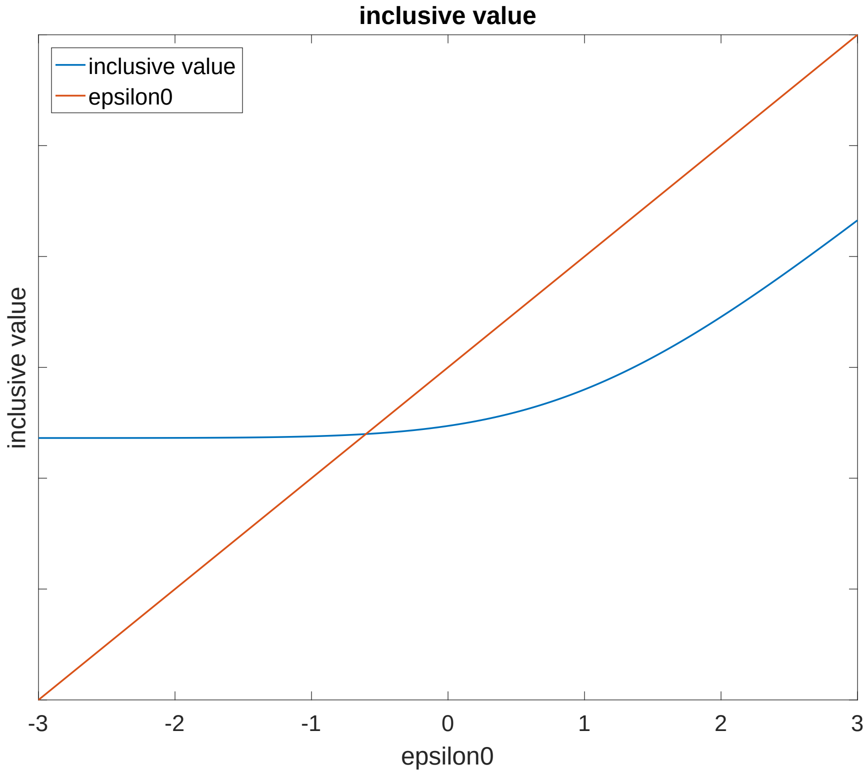
$$P(buy|visit, \varepsilon_{ijt0}) = \tilde{\Phi}(\tilde{u}_{ijt}) \tag{6}$$

where $\tilde{\Phi}(\cdot)$ is the cdf of normal distribution with mean $\sigma_{01}\varepsilon_{ijt0}$ and variance $2 - \sigma_{01}^2$.

One can see how the selection procedure is embedded in the model more clearly from (6). The smaller the outside option value is (the more negative ε_{ijt0} is), the lower the mean of $\tilde{\Phi}(\cdot)$ and hence the higher $P(buy|visit, \varepsilon_{ijt0})$ is. Moreover, the larger the σ_{01} is, the smaller the variance is. It is this correlated structure that generates the spikes observed in sales. The correlated structure has an economic interpretation of the role of advertising in consumer's decision process: consumers visited the website because of the advertisements. Moreover, since they take into account the possibility of buying, they are more likely to buy given they visit.

Having discussed how to solve the purchase stage decision, I now turn to the upper layer of the model, visiting stage decision. The key challenge in this stage is to evaluate the option value

Figure 10: Option value



term, I_{ijt}, in (5). For notational purpose, define $V_0 \equiv 0$ and $V_1 \equiv -p + \delta^{T-t} a + \Gamma_2(\cdot)$ so that $I_{ijt} = \mathbb{E}[\max\{V_0 + \varepsilon_{ijt1},\ V_1 + \varepsilon_{ijt2}\} | \varepsilon_{ijt0}, g_{jt}, T-t]$. It follows immediately from independence of ε_{ijt1} and ε_{ijt2} that $\varepsilon_{ijt1} | \varepsilon_{ijt0} + V_0 \sim N\left(V_0 + \sigma_{01}\varepsilon_{ijt0}, 1 - \sigma_{01}^2\right)$ and $\varepsilon_{ijt2} + V_1 \sim N(V_1, 1)$. Using a result from Nadarajah and Kotz (2008), it follows that

$$
\begin{aligned}
I_{ijt} = {} & \left(V_0 + \sigma_{01}\varepsilon_{ijt0}\right) \Phi\left(\frac{V_0 - V_1 + \sigma_{01}\varepsilon_{ijt0}}{\sqrt{2-\sigma_{01}^2}}\right) \\
& + V_1 \Phi\left(\frac{V_1 - V_0 - \sigma_{01}\varepsilon_{ijt0}}{\sqrt{2-\sigma_{01}^2}}\right) \\
& + \sqrt{2-\sigma_{01}^2}\,\phi\left(\frac{V_0 - V_1 + \sigma_{01}\varepsilon_{ijt0}}{\sqrt{2-\sigma_{01}^2}}\right),
\end{aligned} \tag{7}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the probability density function (pdf) and the cumulative density function (cdf) of the standard normal distribution. (7) has a structure that is easy to interpret. Indeed, if one ignores the last term for the moment, the option value is the weighted average of the two flow utilities in the purchasing stage, with the cdf being the weight.

Figure 10 gives a graphical illustration of the option value. The red line is the 45-degree line and the blue line represents u_{ijt1}, which is the option value plus a scalar utility shifter, $-c+\Gamma_1(\cdot)$. The individual will pay a visit to the website if $u_{ijt1} > \varepsilon_{ijt0}$. Notice that the flat part of the blue curve is the utility of buying, u_{ijt1}. The positive slope part of the blue curve is the utility of visit but not buy, which is a function of ε_{ijt0}. The intuition goes as follows: if the individual has a small ε_{ijt0} draw (the taste shock when not visit), then since she expects she may get another small draw of ε_{ijt1} if visit but not buy as well (since the positive correlation). Thus, she expects that $u_{ijt10} = \varepsilon_{ijt1} < u_{ijt11} = -p+\delta^{T-t}\psi+\Gamma_2(g_{jt})+\varepsilon_{ijt2}$ and thus the option value equals to the flow utility of visit and buy, which is the flat part of the blue curve. Conversely, if the individual has a large ε_{ijt0} draw, then since she expects she may get another large draw of ε_{ijt1} if visit but not buy as well. Consequently, she expects that $u_{ijt10} = \varepsilon_{ijt1} > u_{ijt11} = -p+\delta^{T-t}\psi+\Gamma_2(g_{jt})+\varepsilon_{ijt2}$ and thus the option value equals to the flow utility of visit but not buy, which is the positive slope part of the blue curve.

Figure 10 implies that individual will only visit the website when she has a small draw of ε_{ijt0}. Denote the upper and lower threshold points respectively by ε_{t0min} , the probability of visiting the website is given by

$$P_{jt}(visit) = \Phi_{\varepsilon_{ijt0}}(\varepsilon_{t0min}), \tag{8}$$

where $\Phi_{\varepsilon_{ijt0}}(\cdot)$ denotes the cdf of normal distribution with mean 0 and variance 1.

Figure 10 has a meaningful economic interpretation. It says that those with a small value of ε_0 will visit the website. Moreover, given that they visit the website, they also have a large probability to purchase. As such, the model is able to produce the observed pattern in the data that the conversion rate increases after the ads.

Notice that it is not always the case that the flat part of the option value crosses ε_{ijt0}. In general, depend on the slope of ε_{ijt0} and variance of ε_{ijt1} and ε_{ijt2}, it could have 0,1 or 2 intersections across ε_{ijt0}. In appendix C, I provide a more general case with un-normalized variance of ε_{ijt1} and ε_{ijt2}.

Having discussed the option value and the probability of visiting the website, it is ready to derive the other two choice probabilities. Consider the probability of buying a ticket conditional on visiting the website and ε_{ijt0}, given by (6). To get the conditional probability of buying given visiting without conditional on ε_{ijt0}, one simply needs to integrate out ε_{ijt0}. Notice however that one needs to integrate out ε_{ijt0} not on the entire real line, but only in those regions where consumers would visit the website. In figure 10, this would be the two regions below the minimum threshold. Mathematically, it follows that

$$P_{jt}(buy|visit) = \int \tilde{\Phi}(\tilde{u})d\bar{F}(\varepsilon_{ijt0}), \tag{9}$$

where $\bar{F}(\varepsilon_{ijt0})$ is the cdf of normal distribution "truncated" in certain regions. Details on computing the conditional probability can be found in section B. Finally, the probability of buying a ticket is given by[7]

$$P_{jt}(buy) = P_{jt}(visit)P_{jt}(buy|visit). \tag{10}$$

5.7 Empirical implementation

There is an inner and an outer loop. In the inner loop, I solve the consumer's choice probability for given values of the parameters and compute the value of a method of simulated moment (SMM) objective function. In the outer loop, I then estimate the parameters. The moments I use are related to visits and sales at a given point in time given the advertising activity before that, and the evolution of cumulative visits and sales.

I assume the market size for Dutch online lottery tickets market is 250,000 and I set the market size in the model, denoted by M, to be 1000.[8] Thus each consumer in the model represents 250 real consumers. To implement this, I take aggregate sales and site visitations and divide them by 250. In addition, since I do not observe the market share for each channel group, I assume each channel group has the same market share: $0.2M$.[9] Finally, in the estimation, one time unit is equal to one hour and I count the time between midnight and 7 am as 1 hour. This is a compromise between computational burden and how realistic the model is.

In the data, I only observe that a consumer has bought a ticket, but not which one. I assume that the price of the tickets bought is 3 euros. The key simplifying assumption is that everybody buys the same ticket (and not that some consumers buy multiple ones, for instance).

To estimate the parameters, I first compute the option value at every period numerically on a grid. This gives me, at each point in time, the threshold point that defines the truncation region in $\bar{F}(\cdot)$ and thus $P_{jt}(visit)$. Next, I compute the $P_{jt}(buy|visit)$ by integrating out $\varepsilon_{t0} - s$ in the truncation region obtained from the previous step. The simulated aggregate demand is then given by $0.2M \cdot \Sigma_j P_j(visit)$, and $0.2M \cdot \Sigma_j P_j(buy)$. I then match those two demands to actual aggregated visits and sales. Further details are provided in Appendix D.

6 Results

In this section, I present the estimated results. After I show the parameter estimates and fit in subsection 6.1, I decompose choice probabilities in subsection 6.2. Next, in subsection 6.3, I calculate the elasticities of advertising implied by the parameter estimates. This section is concluded with a comparison between the estimates from the proposed model vs. the model with no correlation between ε-s.

[7]Formally, it should be $P_{jt}(buy\&visit) = P_{jt}(visit)P_{jt}(buy|visit)$. But since I assume that the decisions are sequencitally, it follows that $P(buy\&visit) = P(buy)$.

[8]I experimented with different market sizes and found that results of the counterfactual simulations are not very sensitive to it.

[9]I plan to use data on the average market share of the channels in terms of the audience to refine the equal market share assumption in the future.

Table 4: Estimation results: key parameters

parameter	estimate	std.err.
depreciation rate goodwill stock (λ)	0.450	0.104
hourly discount factor (δ)	0.986	0.004
covariance between taste shock (σ_{01})	0.240	0.008
curvature parameter (γ_3)	3.000	0.174
channel specific effects on flow utility of buy		
national & local TV channel (γ_{12})	0.055	0.084
commercial TVchannel (γ_{22})	0.101	0.041
national radio channel (γ_{32})	0.200	0.140
local radio channel (γ_{42})	0.245	0.131
commercial radio channel (γ_{52})	0.255	0.062
channel specific effects on flow utility of visit		
national & local TV channel (γ_{11})	0.035	0.026
commercial TVchannel (γ_{21})	0.035	0.019
national radio channel (γ_{31})	0.030	0.046
local radio channel (γ_{41})	0.037	0.052
commercial radio channel(γ_{51})	0.037	0.026

Notes: Structural estimates. Obtained using the method of simulated moments. See Sections 5.7 and Appendix D for details.

6.1 Parameter estimates and fit

In this section, I present my estimation results and assess the fit of the model. Table 4 shows the estimated parameters for the key parameters of the model. The effect of advertising depreciates quickly, at an hourly rate of about 55(=1-0.450) percent. The hourly discount factor is estimated to be 0.986. This means that one month before a draw, the value that consumers attach to a ticket is less than 1% of the value on the day of the draw. Next, the estimated covariance between taste shock (σ_{01}) is estimated to be 0.240. Recall that I have normalized the variance of the taste shock to 1. This implies that σ_{01} is also the correlation coefficient. The significant positive correlation shows that the visiting and purchasing stage are indeed positively correlated. That is, those who visit the website are also more likely to purchase.

Now I come to the effect of advertising on visits and sales. The effect of the goodwill stock on flow utility of buying a ticket (γ_{j2}) differs from channel to channel. Channel 5, the commercial radio channel group, is the most effective. Its effect is estimated to be 0.255. In contrast, channel 1, the national and regional tv group, is the least effective channel. The result is in line with the descriptive evidence in described in Figure 8. Next, the effect of the goodwill stock on flow utility of visiting the website (γ_{j1}) is much smaller. The reason is due to the specification of flow utility: the spikes in the visits data is measured by the sum of two parts: the common part, $\Gamma_1(\cdot)$, and the individual specific option value I_{ijt}. The estimated small

Table 5: Estimation results: draw fixed effect

parameter	estimate	std.err.
cost of visiting the website		
10 January, 2014	1.750	0.023
10 February, 2014	2.030	0.025
10 March, 2014	2.020	0.021
10 April, 2014	2.020	0.016
26 April, 2014 (King's Day)	1.690	0.032
10 May, 2014	1.900	0.026
10 June, 2014	2.040	0.017
24 June, 2014 (Orange draw)	1.620	0.018
10 July, 2014	1.800	0.040
10 August, 2014	1.990	0.020
10 September, 2014	2.020	0.020
1 October, 2014 (special 1 October draw)	1.880	0.023
10 October, 2014	1.800	0.029
10 November, 2014	2.020	0.032
10 December, 2014	1.920	0.030
31 December, 2014 (New year's eve draw)	1.620	0.060
value to having a ticket on the day of the draw		
10 January, 2014	0.020	0.205
10 February, 2014	0.250	0.510
10 March, 2014	0.190	0.505
10 April, 2014	0.260	0.256
26 April, 2014 (King's Day)	0.930	0.337
10 May, 2014	0.460	0.233
10 June, 2014	0.410	0.327
24 June, 2014 (Orange draw)	0.170	0.191
10 July, 2014	0.400	0.274
10 August, 2014	0.210	0.381
10 September, 2014	0.460	0.413
1 October, 2014 (special 1 October draw)	0.220	0.462
10 October, 2014	0.390	0.217
10 November, 2014	0.420	0.464
10 December, 2014	0.270	0.483
31 December, 2014 (New year's eve draw)	1.690	0.154

Notes: Structural estimates. Obtained using the method of simulated moments. See Sections 5.7 and Appendix D for details.

value of γ_{j1} implies that the option value accounts for most of the spikes in the visits data. The economic interpretation behind this is that advertising generates online visits mainly through informing high-value consumers. Once taking out the effect of advertising on "high-value" consumers, the remaining effect on an average consumer is small.

Apart from the key parameters, I have also estimated the fixed effect for each draw. These are the cost to visit the website (c) and the value of holding a ticket at the time of the draw (ψ). Table 5 presents the result. In general, a larger number of visits in the month implies a lower cost of visiting and similarly, larger sales implies higher estimates of draw fixed effects for sales. Moreover, one can see that draws with a short time period are estimated to have a lower cost of visiting the website.

Figure 11 shows the model fit. With only a few parameters, the model arguably fits the patterns in the data relatively well.

6.2 Decompose probabilities

One of the advantages of estimating a structural model is that, once estimated, one can decompose probability of sales into the probability of visits and probability of buy given visits. As a result, one can study the effect of advertising on these three probabilities. Figure 12 shows the plot of these probabilities for one typical draw.

Two things are worth to notice: first, the conditional probability of buying a ticket, given that the consumer has visited the website, increases over time. This means that it is a trend that the closer to the deadline, the more those with a high probability of buying are entering the pool.[10] Second, it is clear from the figure that advertisement does have a positive effect on the conditional probability of buying, that is, the conversion rate.

6.3 Elasticities of advertising

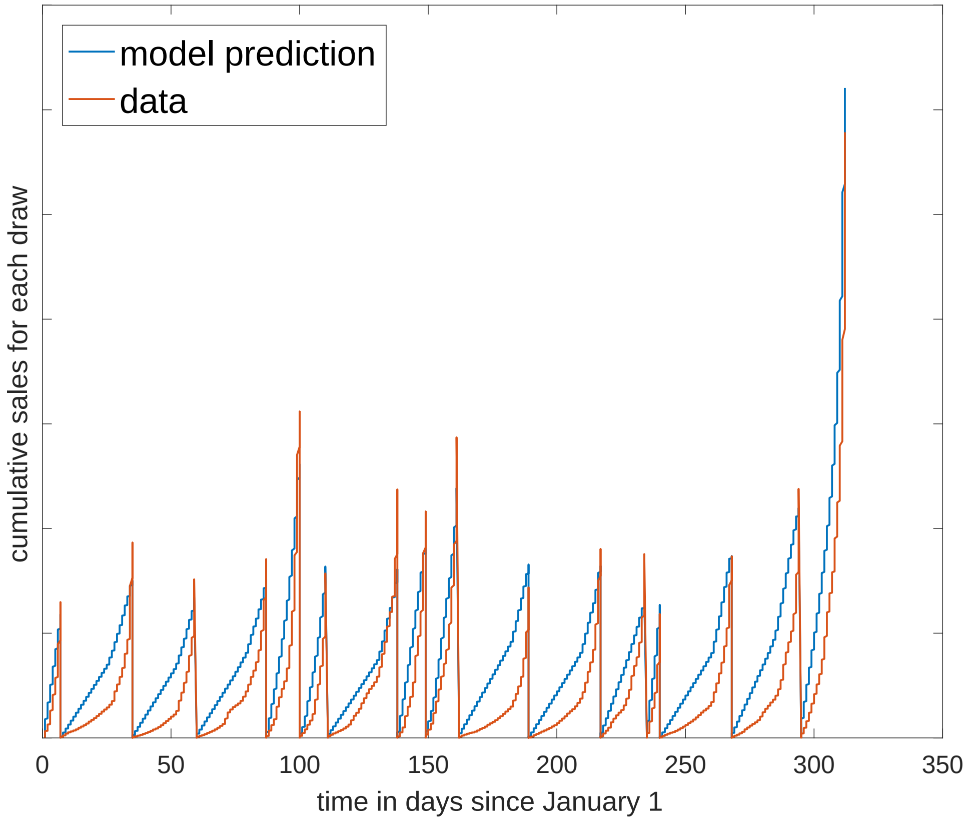
In this subsection, I calculate the elasticities of advertising implied by the model parameter estimates. The elasticity is defined as Elasticitity $= [\frac{P(1.01\overline{GRP})}{P(\overline{GRP})} - 1] \cdot 100$, where $\overline{GRP}$ is the average GRP's over the last two weeks before the draw and $P(.)$ is the choice probability.

I calculate the elasticity in the following way: I increase GRP only for one channel at a time. I do this for each hour during the last 5 days before the draw using the average value for the cost of visiting the website and the value of holding a ticket and then take the average elasticities across these hours. Table 6 shows the result. For example, 0.174 in row 2, column 3 means that 1% increase in GRP on commercial TVchannel will increase the prob(buy) by 0.174%. Interpretations of other numbers are similar. These elasticities are more or less in line with those found in other literature. [11]

[10] In my model, those consumers with a high probability of buying, given visiting the website, are those with low ε_0.

[11] Sethuraman et al. (2011) summarize the elasticities of advertising for different studies.

Figure 11: Model fit



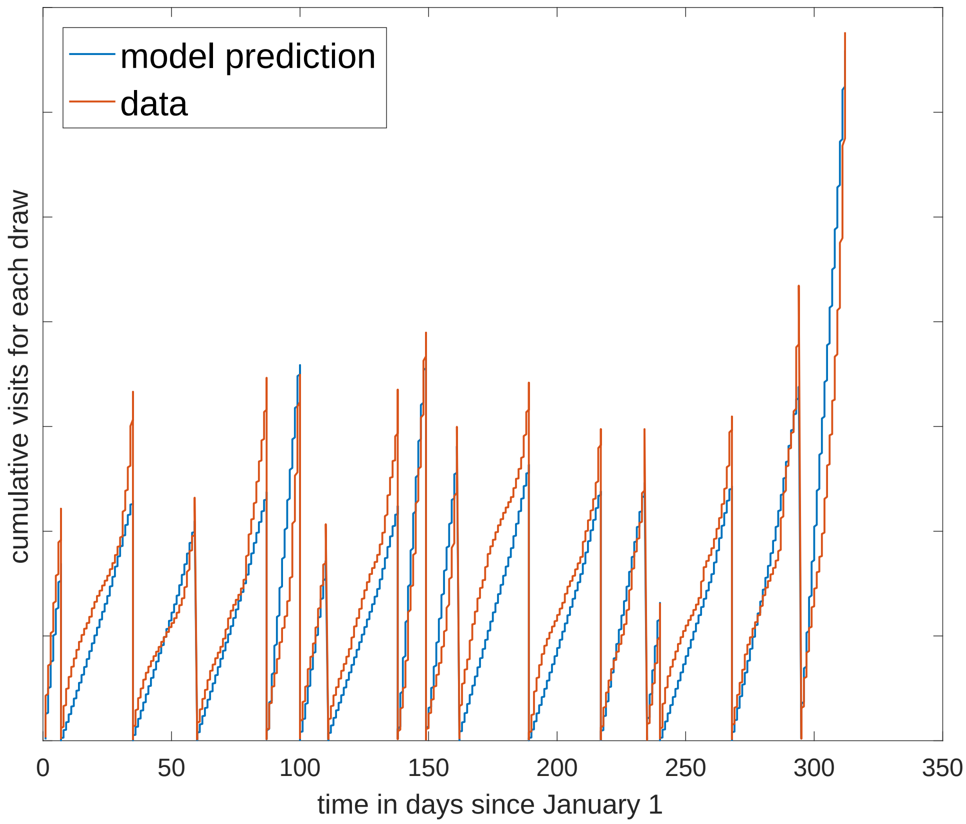
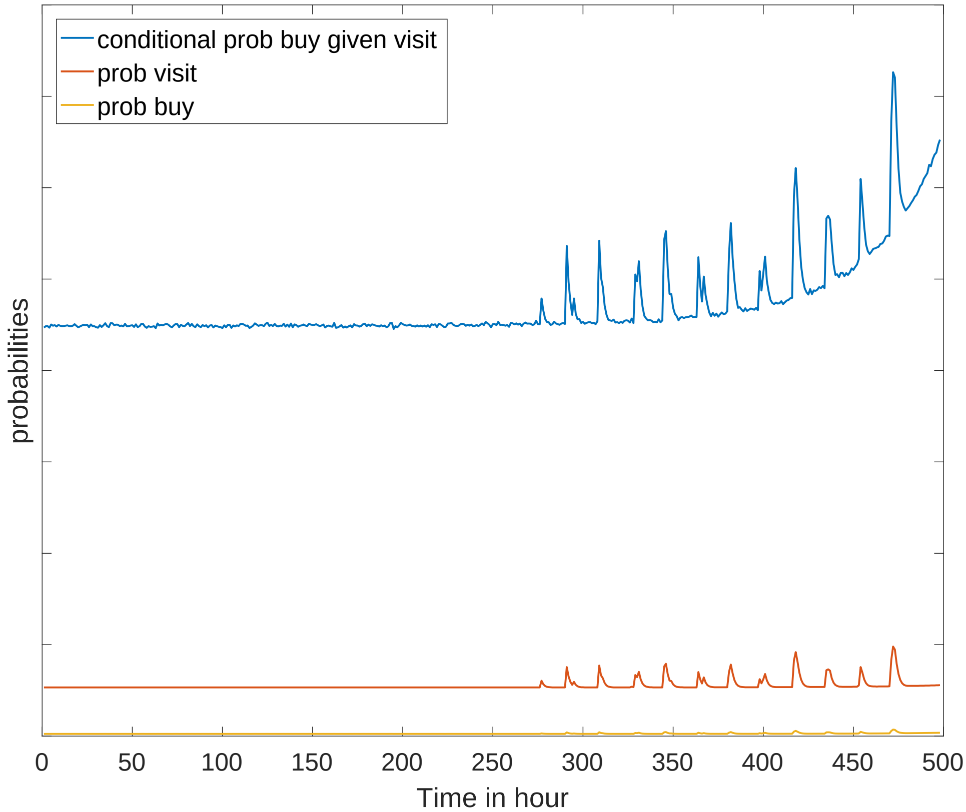


Figure 12: Decompose probabilities



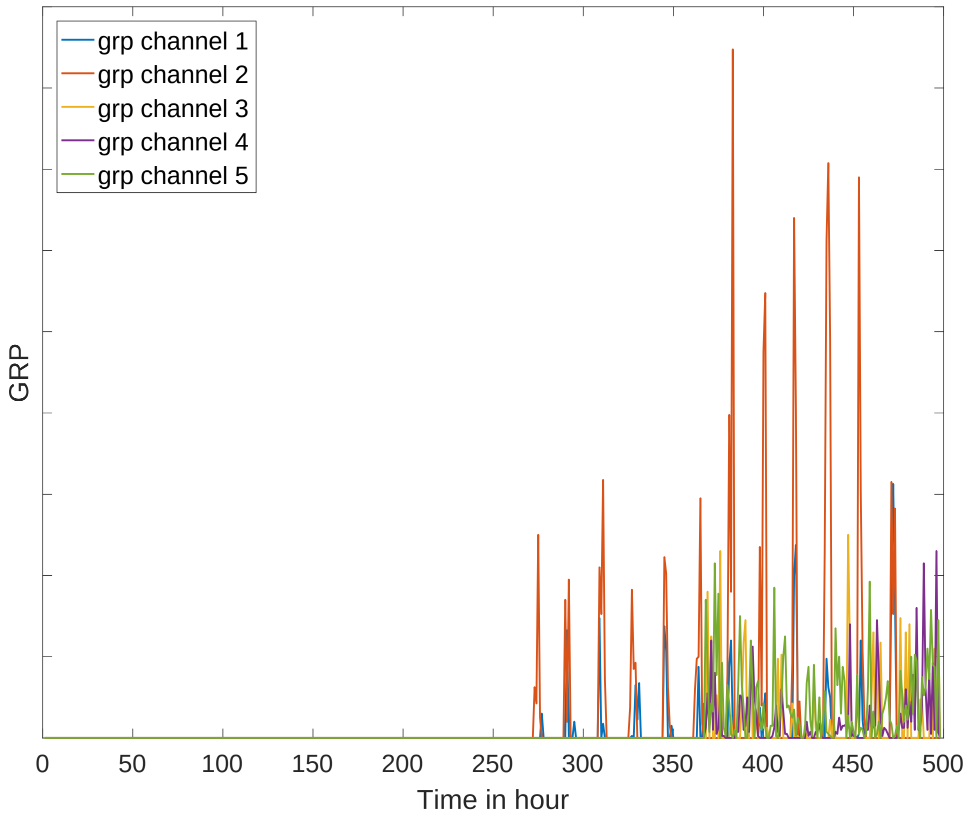


Table 6: Elasticities of advertising

channel	prob(visit)	prob(buy|visit)	prob(buy)
national & local TV channel	0.091	0.042	0.132
commercial TVchannel	0.090	0.084	0.174
national radio channel	0.108	0.167	0.276
local radio channel	0.126	0.203	0.329
commercial radio channel	0.138	0.209	0.347

6.4 The proposed model vs. the model with no correlation between ε-s

In this subsection, I compare the estimates from two models. The first one is the proposed model with the correlation between taste shocks. The second model is the one without the correlation. Apart from the correlation, these two models are identical otherwise. Table 7 and 8 show the results.

First, Table 7 shows that if one ignored the correlation between taste shocks, one would get much larger estimates for the effect of advertising on purchasing. Second, Table 8 shows that the cost of visiting the website would be much larger. The intuition goes as follows: without selection, both the serious buyers and the “average” visitors (those with low probability to buy) will visit the website. Consequently, for given parameters, the model would predict a much lower conversion rate. As a result, in order to get a higher conversion rate that fits the data, the cost of visiting the website must be larger (so fewer people will visit) and the effect of ads on purchasing a ticket would be larger (so more people will buy).

7 Counterfactual experiments

Having estimated the model, I turn to the supply side. I do not have access to data on the profitability of an additional sold ticket, and also not on the cost of one GRP in a given TV/radio station. It is, however, not unreasonable to assume as an approximation that the price of one GRP does not vary over different stations. Therefore, it is meaningful to study whether a given (monetary) budget could be allocated better over different channels, by asking the question whether it is possible to sell more tickets when one allocates the same number of GRP’s in a different way.

Importantly, there are two dimensions where I could vary GRP’s. The first dimension is time: when to advertise? For example, one can allocate more GRP’s to a day that is closer to the deadline. The second dimension is location: where to advertise? In this paper, I focus on the second one. To do this, one needs to keep both the timing and the size of GRP’s as they are and then only varies GRP’s across channels.

Table 7: Comparison estimates: key parameters

Comparison	with correlation	without correlation
depreciation rate goodwill stock (λ)	0.450	0.539
hourly discount factor (δ)	0.986	0.964
covariance between taste shock (σ_{01})	0.240	0
curvature parameter (γ_3)	3.000	2.934
channel specific effects on flow utility of buy		
national & local TV channel (γ_{12})	0.055	0.187
commercial TV channel (γ_{22})	0.101	0.165
national radio channel (γ_{32})	0.200	0.358
local radio channel (γ_{42})	0.245	0.445
commercial radio channel (γ_{52})	0.255	0.527
channel specific effects on flow utility of visit		
national & local TV channel (γ_{11})	0.035	0.035
commercial TVchannel (γ_{21})	0.035	0.035
national radio channel (γ_{31})	0.030	0.030
local radio channel (γ_{41})	0.037	0.037
commercial radio channel (γ_{51})	0.037	0.037

7.1 Setup

Suppose there is a typical draw: the cost of visiting the website and the value of holding a ticket are the average values over all draws. Moreover, there are no ads during the entire period of this draw. Now suppose the manager, Sophia, buys 10 GRP's between 7 pm-7:59 pm and 10 GRP's between 8 pm-8:59 pm on the day 7 days before the draw. Then she faces the following question: how to allocate 10 GRP's between 7 pm-7:59 pm and 10 GRP's between 8 pm-8:59 pm over different channels so that the online sales are maximized. Sophia's question can be summarized in the following mathematical form:

$$\text{Max } q(w, \hat{\theta})$$

$$\text{s.t. } (w_{17} + w_{27} + w_{37} + w_{47} + w_{57}) \cdot 10 \leq 10$$

$$(w_{18} + w_{28} + w_{28} + w_{28} + w_{28}) \cdot 10 \leq 10$$

$$\forall w_{jt} \geq 0,$$

where q is total sales, $\hat{\theta}$ is the estimated parameters which the manager takes as given and w_{jt} is the weight for channel j at hour t. For example, w_{27} is the weight for channel 2 at the hour

Table 8: Comparison estimates: draw fixed effect

Comparison	with correlation	without correlation
cost of visiting the website		
10 January, 2014	1.750	2.250
10 February, 2014	2.030	2.530
10 March, 2014	2.020	2.520
10 April, 2014	2.020	2.520
26 April, 2014 (King's Day)	1.690	2.190
10 May, 2014	1.900	2.400
10 June, 2014	2.040	2.540
24 June, 2014 (Orange draw)	1.620	2.120
10 July, 2014	1.800	2.300
10 August, 2014	1.990	2.490
10 September, 2014	2.020	2.410
1 October, 2014 (special 1 October draw)	1.880	2.380
10 October, 2014	1.800	2.300
10 November, 2014	2.020	2.419
10 December, 2014	1.920	2.420
31 December, 2014 (New year's eve draw)	1.620	2.120
value to having a ticket on the day of the draw		
10 January, 2014	0.020	0.047
10 February, 2014	0.250	0.404
10 March, 2014	0.190	0.268
10 April, 2014	0.260	0.341
26 April, 2014 (King's Day)	0.930	1.315
10 May, 2014	0.460	0.450
10 June, 2014	0.410	0.589
24 June, 2014 (Orange draw)	0.170	0.274
10 July, 2014	0.400	0.619
10 August, 2014	0.210	0.266
10 September, 2014	0.460	0.573
1 October, 2014 (special 1 October draw)	0.220	0.276
10 October, 2014	0.390	0.376
10 November, 2014	0.420	0.348
10 December, 2014	0.270	0.309
31 December, 2014 (New year's eve draw)	1.690	2.961

Table 9: Effect of various advertising strategies

strategy	visits	sales	conversion rate
data (reference point)	100%	100%	100%
allocate all GRP's on channel			
national & local TV channel	99.52%	98.18%	98.65%
commercial TVchannel	99.55%	98.42%	98.86%
national radio channel	99.58%	99.03%	99.45%
local radio channel	99.69%	99.60%	99.91%
commercial radio channel	99.70%	**99.70%**	100.01%
optimal allocation	99.97%	**100.39%**	100.42%

Notes: This table shows the effect of using alternative advertising strategies. See text for a description of these strategies. The conversion rate is calculated by the predicted sales over the predicted visitations.

between 7 pm-7:59 pm.

7.2 Results

First suppose that Sophia allocates the budget based on "past experience", i.e., she uses the average weight over the 7 days before the draw from previous draws. Next, she uses the targeting strategy, that is, allocating all GRP's on one channel. Finally, she solves the maximization problem to find the optimal allocation.

Table 9 shows the result of various advertising strategies. The first row is the reference level: she uses the average weight in the data. Row 2 to Row 6 show the model prediction under the targeted strategy. As implied by the parameter estimates, the commercial radio channel which has the largest estimated effect performs best. The national & local TV channel, on the other hand, is least effective.

Finally, the optimal allocation shows that Sophia could do better than using pure targeted strategy. In particular, the optimal strategy is to spend about 43% of the budget on the commercial radio channels, about 37% on the local radio channels, about 18% on the national radio channel and about 2% on the commercial TV channels. That is, although the commercial radio channel has the largest estimated effect, due to the functional form of diminishing marginal return of advertising, she could improve sales by allocating some GRP's to other channels so that the marginal returns at the optimal allocation are more or less the same across channels.

8 Summary and concluding remarks

Advertising can affect consumer behavior at the consideration and the purchasing stage. This paper uses high frequency data on TV and radio advertising from different channels together with online sales and website visits data to measure the effects of advertising. I find positive effects of advertising on consideration and conversion and the effects depend on the channel on which the firm advertises.

Besides, I point out that the observed increase in the conversion rate could be due to the fact that those who are motivated to visit the website through advertisements are different from those who usually visit. The former ones have a higher probability to buy given that they visit. Ignoring this and studying consideration and conversion separately could result in an underestimated conversion rate and thus a suboptimal advertising strategy, in particular when advertising on different channels reaches different audiences. I provide an explanation for this by spelling out a new integrated model of consideration and conversion that can generate the observed pattern in the data. I estimate the structural parameters of this model and show that one would overestimate both, the effects of advertising and the cost of visiting the website if one would ignore this selection. I simulate the effects of counterfactual targeted advertising strategies and conclude that shifting advertising across channels can lead to increased sales.

The proposed model can be used by firms to optimize online sales by choosing TV and radio advertising schedules, but of course, the model could be used in other situations as well. A closely related one is the real-time bidding in advertising auctions for online advertisements on, say, Facebook or Google. In this context, firms need to bid for the advertisement position on the user's web browser. The value of the bid then depends crucially on the user's expected conversion rate, as the price is only paid when users actually click on the advertisement. Using a two-stage model that takes into account selection is particularly important: when the manager estimates the conversion rate using only data of those individuals who visit then she gets an average conversion rate— average between those who go on the website just like that and those who go on the website because they see an ad—but the one that matters for ad auctions is the one for those who actually go on the website because they see an ad.

In a broader sense, the model proposed in this paper can also be used in situations where either the agent does not need to experience the good or doing so is not possible. An example could be buying an airline ticket. Airline tickets are similar to the lottery tickets since they also have simple characteristics: price and the time of the flight. Moreover, there is also a deadline for buying an airline ticket determined by the consumer: the date of her trip. Finally, consumers cannot experience the product before the purchase. Therefore, those who visit the website of airlines because of the ads are more likely to purchase as well: they have no reason to check the ticket price from, say, London to New York City if they have no plan to travel between these two cities. In other words, those who visit the website are also selected. In this paper, I have shown that taking such selection issues into account will help firms to design better advertising strategies.

References

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

Albuquerque, P. and B. J. Bronnenberg (2009). Estimating demand heterogeneity using aggregated data: An application to the frozen pizza category. *Marketing Science 28*(2), 356–372.

Bronnenberg, B. J. and W. R. Vanhonacker (1996). Limited choice sets, local price response and implied measures of price competition. *Journal of Marketing Research*, 163–173.

Chandy, R. K., G. J. Tellis, D. J. MacInnis, and P. Thaivanich (2001). What to say when: Advertising appeals in evolving markets. *Journal of Marketing Research 38*(4), 399–414.

Clark, C. R., U. Doraszelski, and M. Draganska (2009). The effect of advertising on brand awareness and perceived quality: An empirical investigation using panel data. *Quantitative Marketing and Economics 7*(2), 207–236.

Danaher, P. J. and T. S. Dagger (2013). Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. *Journal of Marketing Research 50*(4), 517–534.

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.

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

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

Goldfarb, A. and C. Tucker (2011a). Online display advertising: Targeting and obtrusiveness. *Marketing Science 30*(3), 389–404.

Goldfarb, A. and C. Tucker (2011b). Search engine advertising: Channel substitution when pricing ads to context. *Management Science 57*(3), 458–470.

Haans, H., N. Raassens, and R. van Hout (2013). Search engine advertisements: The impact of advertising statements on click-through and conversion rates. *Marketing Letters 24*(2), 151–163.

He, C. and T. J. Klein (2016). Advertising as a reminder: evidence from the dutch state lottery.

Hoban, P. R. and R. E. Bucklin (2015). Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research 52*(3), 375–393.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2010). Online demand under limited consumer search. *Marketing science 29*(6), 1001–1023.

Kim, J. B., P. Albuquerque, and B. J. Bronnenberg (2016). The probit choice model under sequential search with an application to online retailing. *Management Science*.

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.

Lambrecht, A. and C. Tucker (2013). When does retargeting work? information specificity in online advertising. *Journal of Marketing Research 50*(5), 561–576.

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

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.

Manchanda, P., J.-P. Dubé, K. Y. Goh, and P. K. Chintagunta (2006). The effect of banner advertising on internet purchasing. *Journal of Marketing Research 43*(1), 98–108.

Nadarajah, S. and S. Kotz (2008). Exact distribution of the max/min of two gaussian random variables. *IEEE Transactions on very large scale integration (VLSI) systems 16*(2), 210–212.

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.

Sethuraman, R., G. J. Tellis, and R. A. Briesch (2011). How well does advertising work? generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research 48*(3), 457–471.

Sherman, L. and J. Deighton (2001). Banner advertising: Measuring effectiveness and optimizing placement. *Journal of Interactive Marketing 15*(2), 60–64.

Tellis, G. J., R. K. Chandy, and P. Thaivanich (2000). Which ad works, when, where, and how often? modeling the effects of direct television advertising. *Journal of Marketing Research 37*(1), 32–46.

A Additional tables and figures

Table 12: Effect of TV and radio advertising from different channels

	(1) visits	(2) sales
GRP1 between 0 and 4 minutes ago	0.0236***	0.00761***
	(0.00191)	(0.00165)
5 and 9 minutes	0.0301***	0.0338***
	(0.00176)	(0.00188)
10 and 14 minutes	0.0100***	0.0384***
	(0.00114)	(0.00163)
15 and 19 minutes	0.00402***	0.0226***
	(0.000966)	(0.00166)
20 and 24 minutes	0.00315***	0.0130***
	(0.000938)	(0.00134)
25 and 29 minutes	0.00204*	0.00976***
	(0.000941)	(0.00164)
0.5 and 1 hour	-0.00128**	0.00445***
	(0.000400)	(0.000598)
1 and 1.5 hours	0.00214***	0.00326***
	(0.000479)	(0.000630)
1.5 and 2 hours	-0.000858*	0.000693
	(0.000415)	(0.000616)
2 and 2.5 hours	0.00274***	0.000863
	(0.000520)	(0.000570)
2.5 and 3 hours	-0.00108*	-0.00385***
	(0.000512)	(0.000564)
3 and 3.5 hours	-0.00284***	-0.00711***
	(0.000483)	(0.000506)
3.5 and 4 hours	-0.00700***	-0.0127***
	(0.000491)	(0.000545)
GRP2 between 0 and 4 minutes ago	0.0346***	0.0184***

	(0.00141)	(0.00165)
5 and 9 minutes	0.0478***	0.0453***
	(0.00127)	(0.00195)
10 and 14 minutes	0.0144***	0.0503***
	(0.00103)	(0.00173)
15 and 19 minutes	0.0101***	0.0372***
	(0.00103)	(0.00161)
20 and 24 minutes	0.00817***	0.0280***
	(0.00105)	(0.00167)
25 and 29 minutes	0.00861***	0.0183***
	(0.00109)	(0.00160)
0.5 and 1 hour	0.00948***	0.0161***
	(0.000437)	(0.000688)
1 and 1.5 hours	0.00890***	0.0117***
	(0.000445)	(0.000662)
1.5 and 2 hours	0.00759***	0.00746***
	(0.000454)	(0.000640)
2 and 2.5 hours	0.00517***	0.00158*
	(0.000470)	(0.000616)
2.5 and 3 hours	0.00186***	-0.00419***
	(0.000471)	(0.000587)
3 and 3.5 hours	-0.00199***	-0.0124***
	(0.000498)	(0.000588)
3.5 and 4 hours	-0.00794***	-0.0221***
	(0.000496)	(0.000575)
GRP3 between 0 and 4 minutes ago	0.00354	0.00413
	(0.00186)	(0.00252)
5 and 9 minutes	-0.00582**	0.0114***
	(0.00206)	(0.00263)
10 and 14 minutes	-0.00617***	0.0105***
	(0.00180)	(0.00268)
15 and 19 minutes	-0.00563**	0.00525*
	(0.00185)	(0.00260)

20 and 24 minutes	-0.00405*	0.00693**
	(0.00180)	(0.00250)
25 and 29 minutes	-0.00424*	0.00691**
	(0.00188)	(0.00255)
0.5 and 1 hour	0.00134	0.00878***
	(0.000831)	(0.00115)
1 and 1.5 hours	-0.00500***	0.00926***
	(0.000835)	(0.00112)
1.5 and 2 hours	-0.00313***	0.0122***
	(0.000845)	(0.00115)
2 and 2.5 hours	-0.00502***	0.00781***
	(0.000833)	(0.00114)
2.5 and 3 hours	-0.00993***	0.00627***
	(0.000872)	(0.00117)
3 and 3.5 hours	-0.000299	0.0110***
	(0.000878)	(0.00114)
3.5 and 4 hours	-0.00723***	0.00713***
	(0.000919)	(0.00120)
GRP4 between 0 and 4 minutes ago	0.0108***	-0.000113
	(0.00319)	(0.00412)
5 and 9 minutes	0.00553	0.00306
	(0.00316)	(0.00407)
10 and 14 minutes	0.00739*	0.00629
	(0.00320)	(0.00413)
15 and 19 minutes	0.0133***	0.00758
	(0.00314)	(0.00402)
20 and 24 minutes	0.0144***	0.0174***
	(0.00321)	(0.00410)
25 and 29 minutes	0.0153***	0.0151***
	(0.00337)	(0.00408)
0.5 and 1 hour	0.0158***	0.00957***
	(0.00138)	(0.00178)

1 and 1.5 hours	0.00697***	0.00978***
	(0.00139)	(0.00176)
1.5 and 2 hours	0.00922***	0.0115***
	(0.00139)	(0.00180)
2 and 2.5 hours	0.00469**	0.0110***
	(0.00148)	(0.00184)
2.5 and 3 hours	0.00850***	0.0117***
	(0.00147)	(0.00190)
3 and 3.5 hours	0.00911***	0.0130***
	(0.00144)	(0.00183)
3.5 and 4 hours	0.00210	0.0121***
	(0.00143)	(0.00179)
GRP5 between 0 and 4 minutes ago	0.0160***	0.00899***
	(0.00186)	(0.00225)
5 and 9 minutes	0.0107***	0.0117***
	(0.00178)	(0.00225)
10 and 14 minutes	0.0114***	0.0138***
	(0.00182)	(0.00232)
15 and 19 minutes	0.0141***	0.0156***
	(0.00180)	(0.00227)
20 and 24 minutes	0.0129***	0.0147***
	(0.00179)	(0.00224)
25 and 29 minutes	0.0144***	0.0181***
	(0.00181)	(0.00225)
0.5 and 1 hour	0.0115***	0.0171***
	(0.000783)	(0.000948)
1 and 1.5 hours	0.0104***	0.0177***
	(0.000769)	(0.000950)
1.5 and 2 hours	0.00875***	0.0191***
	(0.000790)	(0.000952)
2 and 2.5 hours	0.00546***	0.0166***
	(0.000796)	(0.000953)

2.5 and 3 hours	0.00690***	0.0139***
	(0.000822)	(0.000937)
3 and 3.5 hours	0.00882***	0.0124***
	(0.000809)	(0.000913)
3.5 and 4 hours	0.00735***	0.0121***
	(0.000830)	(0.000920)
draw dummies	Yes	Yes
days to draw dummies	Yes	Yes
hour dummies	Yes	Yes
Observations	441223	441223
R^2	0.843	0.665

Standard errors in parentheses

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

B Computing conditional choice probability

I now discuss for given paramters, how to conpute (9), the conditional probability of buying given that the consumer has visited the website:

$$P(buy|visit) = \int \tilde{\Phi}(\tilde{u})d\bar{F}(\varepsilon_{it0}).$$

The key challenge is to take random draws from $\bar{F}(\varepsilon_{it0})$ which is the cdf of normal distribution "truncated" in certain regions. Once such draws are taken, the rest is standard: the integral can be computed by the sample average. Thus, the question boils down to taking random draws from $\bar{F}(\varepsilon_{it0})$. This is not trivial because $\bar{F}(\varepsilon_{it0})$ is a function of parameters. That is, the truncation region changes as parameter value changes. Moreover, depending on the parameter's value, $\bar{F}(\varepsilon_{it0})$ can be left truncated, right truncated or truncated on both sides. These bring two issues: first, the random draws from $\bar{F}(\varepsilon_{it0})$ should be "fixed" so that one is certain that the value of objective function changes due to parameter changes instead of new random draw. Second, the algorithm should be such that it includes all cases for different truncation regions.

Suppose one has computed $P(visit)$ so that one knows for each time period, the lower and upper cutoff points, denoted by a and b. Given this input, I propose the following algorithm:

1. Draw a set of random numbers from standard uniform distribution and keep it fixed outside the estimation loop Denote this set by S.

2. Compute the total length of truncated region, denoted by r: $r = \Phi_{\varepsilon_0}(a) + (1 - \Phi_{\varepsilon_0}(b))$.

3. Transform each point in S into the uniform distribution of length r : $\forall s \in S : s \longrightarrow s \cdot r$,

Table 10: Definition of TV channels and groups

channel name	channel number before grouping	channel number after grouping
24Kitchen	1	2
Com.Cent.F	2	2
Comedy	3	2
Discovery	4	2
FS3E	5	2
Fox	6	2
Fox Sp1	7	2
Fox Sp 2	8	2
Geographic	9	2
ID	10	2
L1 TV	11	1
MTV	12	2
NPO1	13	1
NPO2	14	1
NPO3	15	1
Ned1	16	1
Ned2	17	1
Ned3	18	1
Net5	19	2
O.Brabant	20	1
RTL4	21	2
RTL5	22	2
RTL7	23	2
RTL8	24	2
RTV Utr.	25	1
Rijnmond	26	1
SBS6	27	2
TLC	28	2
TV Drenthe	29	1
TV Flevo	30	1
TV Gelderl	31	1
TV Nh	32	1
TV Noord	33	1
TV Oost	34	1
TV West	35	1
TV Zeeland	36	1
Veronica	37	2

Notes: This table shows the lists of Dutch TV channels. The first column shows the names for each channel. The second column is the channel number before grouping. They are assigned to each channel accroding to the alphabetical order of channel name. The third column shows the new channel number after grouping. Channel 1 consists of national and local TV channels. Channel 2 indicates commercial channels.

Table 11: Definition of radio channels and groups

channel name	channel number before grouping	channel number after grouping
100% NL	1	5
Arrow Classic Rock	2	5
Freez FM	3	5
Fresh FM	4	5
Hot-Radio	5	5
Joy Radio	6	5
L1 Radio	7	4
NPO 3FM	8	3
NPO Radio 1	9	3
NPO Radio 2	10	3
NPO Radio 3FM	11	3
Omroep Brabant	12	4
Omroep Zeeland	13	4
Omrop Fryslan	14	4
Open Rotterdam	15	4
Optimaal FM	16	5
Puur NL	17	5
Q-Music	18	5
Radio 10	19	5
Radio 2	20	3
Radio 3FM	21	3
Radio 538	22	5
Radio 8 FM	23	4
Radio Continu	24	5
Radio Decibel	25	5
Radio Drenthe	26	4
Radio Flevoland	27	4
Radio Gelderland	28	4
Radio M Utrecht	29	4
Radio Noord	30	4
Radio Noord-Holland	31	4
Radio Oost	32	4
Radio Rijnmond	33	4
Radio Royaal	34	5
Radio Veronica	35	5
Radio West	36	4
RadioNL	37	5
Simone FM	38	5
Sky Radio	39	5
Slam!FM	40	5
Sublime FM	41	5
Waterstad FM	42	5
Wald FM	43	5

Notes: This table shows the lists of Dutch radio channels. The first column shows the names for each channel. The second column is the channel number before grouping. They are assigned to each channel accroding to the alphabetical order of channel name. The third column shows the new channel number after grouping. Channel 3 consists of national radio channels. Channel 4 refer to those local radio channels. Channel 5 indicates commercial channels.

denote this new set by S'.

4. Transform each point in S' into the truncated region: $(-\infty, a] \cup [b, +\infty)$ using the following mapping:

$$\forall s \in S' : \begin{cases} s \longrightarrow s + (b-a) & \text{if } s > a \\ s \longrightarrow s & \text{otherwise} \end{cases},$$

denote the new set by S''.

5. The desired draws from $\bar{F}(.)$ are draws from $\Phi^{-1}_{\varepsilon_0}(S'')$.

Notice that the proposed algorithm overcomes the aforementioned issues. First, step 1 implies that although the real value of draws changes for each parameter value, the *relative position* of each random draw on the finite length line segment keeps the same. This is important since those random draws are "fixed" from the estimation point of view. Second, the proposed algorithm includes every case of truncation (left, right or two-sided). For example, if it is left truncated, then the upper threshold b tends to infinity. This allows me to avoid case by case inspection so that the computational time is saved.

C The probit model: a more general case

Consider a model described in figure 9. The vector of utility shock is jointly multivariate normally distributed: $\{\varepsilon_{it0}, \varepsilon_{it1}, \varepsilon_{it2}\} \sim N(0, \Sigma)$, where $\Sigma = \begin{bmatrix} \sigma_0^2 & \sigma_{01} & 0 \\ . & \sigma_1^2 & 0 \\ . & . & \sigma_2^2 \end{bmatrix}$. The idea is that the model imposes that ε_{it0} and ε_{it1} are correlated while ε_{it2} is independent. In the following, I first discuss the lower layer of the model: purchasing decision stage and then discuss the upper layer of the model: visiting decision stage.

C.1 Purchasing stage decision

It follows from result of multivariate normal distribution that conditional on ε_{ijt0}, the joint distribution of $\{\varepsilon_{ijt1}, \varepsilon_{ijt2}\}$ is also (bivariate) normally distributed: $\{\varepsilon_{ijt1}, \varepsilon_{ijt2}\}|\varepsilon_{ijt0} \sim N\left(\begin{bmatrix} \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0} \\ 0 \end{bmatrix}, \Sigma_{12}\right)$ where $\Sigma_{12} = \begin{bmatrix} \sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$. Notice that conditional on ε_{ijt0}, Σ_{12} implies $\varepsilon_{ijt1}$is independent from ε_{ijt2}. Joint normality of error terms means the purchasing stage decision is a standard binary probit model with uncorrelated error terms and thus can be estimated without numerical integral.

C.2 Visiting stage decision

The key challenge in the upper model is to evaluate the option value term, I_{ijt}, which is a function of ε_{ijt0}. For notational purpose, define $V_0 \equiv 0$ and $V_1 \equiv -p + \delta^{T-t}a + \Gamma_2(\cdot)$ so that $I_{it} = \mathbb{E}[max\{V_0 + \varepsilon_{ijt1}, V_1 + \varepsilon_{ijt2}\}|\varepsilon_{ijt0}]$. It follows immediately from independence in the previous subsection that $\varepsilon_{ijt1}|\varepsilon_{ijt0} + A \sim N\left(V_0 + \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0}, \sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2}\right)$ and $V_1 + \varepsilon_{ijt2} \sim N(V_1, \sigma_2^2)$. Using result of multivariate normal distribution, $I_{ijt} = \left(V_0 + \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0}\right)\Phi\left(\frac{V_0 - V_1 + \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0}}{\sqrt{\sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} + \sigma_2^2}}\right) + V_1\Phi\left(\frac{V_1 - V_0 - \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0}}{\sqrt{\sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} + \sigma_2^2}}\right) + \sqrt{\sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} + \sigma_2^2}\,\phi\left(\frac{V_0 - V_1 + \frac{\sigma_{01}}{\sigma_1^2}\varepsilon_{ijt0}}{\sqrt{\sigma_1^2 - \frac{\sigma_{01}^2}{\sigma_0^2} + \sigma_2^2}}\right)$ where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the pdf and the cdf of the standard normal distribution.

In general, depending on the slope of ε_{ijt0} and parameters that affect the difference in flow utilities, it could have 0,1 or 2 intersections across ε_{ijt0}. Figure 13 shows all other 4 cases. Whether the option value will cross the 45 degree line from above depends on the slope before ε_{ijt0}: $\frac{\sigma_{01}}{\sigma_1^2}$.

D Details on the econometric implementation

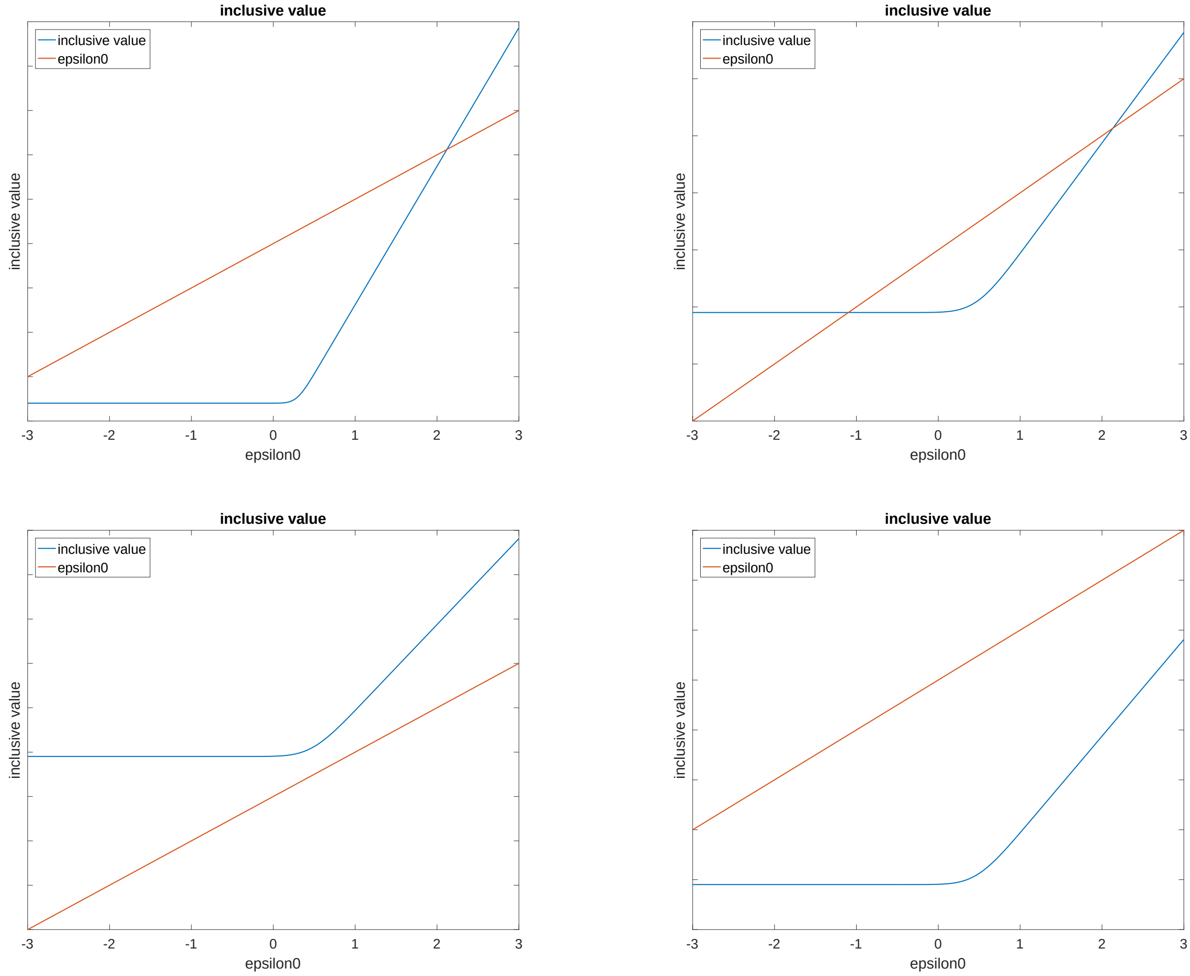
I provide further details on the econometric implementation.

D.1 Empirical setup

The data contain information on ticket sales, online traffic and advertising activities for 16 draws. Since I 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. Furthermore, since the online traffic at the beginning of each draw is very high because consumers want to check if they have won the lottery. These amount of online traffic obviously has nothing to do with purchasing decision, I exclude the first 3 days of data for each draw. Thus, the total number of periods is $\tau = 6564$ (τ is not to be confused with T, which we have defined in the context of our model). I divide them up into sub-periods, one for each draw. I account for the fact that they differ with respect to the total number of hours (T in the model) and the value to holding a ticket (ψ in the model), and of course with respect to the realized advertising activity.[12] The ticket price is constant over time and across draws.

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

Figure 13: Option value: all other cases



D.2 Method of simulated moments

The set of structural parameters that do not change across draws is $\{\lambda, \sigma_{01}, \sigma_1, \delta, \gamma_1, \gamma_2\}$. In addition, I estimate 16 values $\psi_1, \ldots, \psi_{16}$ to holding a ticket at the time of the draw and 16 cost of visiting the website $c_1, ..., c_{16}$. Thus the full set of structural parameters to be estimated is $\theta \equiv \{\lambda, \sigma, \delta, \gamma_1, \gamma_2, \psi_1, \ldots, \psi_{16}, c_1, ..., c_{16}\}$.

Let z_t be a vector of exogenous variables constructed from the data (specified in Section D.3 below) and $\hat{q}^s_t(\theta) \equiv q^s_t - \tilde{q}^s_t(\theta)$ be the difference between actual demand q^s_t in the data and the model prediction $\tilde{q}^s_t(\theta)$. Similarly, define $\hat{q}^v_t(\theta) \equiv q^v_t - \tilde{q}^v_t(\theta)$ be the difference between actual website visits q^v_t in the data and the model prediction $\tilde{q}^v_t(\theta)$. I stack the two quantity together and define $\hat{u}_t(\theta) \equiv q_t - \tilde{q}_t(\theta)$ with $q_t \equiv \begin{bmatrix} q^s_t \\ q^v_t \end{bmatrix}$. ($\tilde{q}_t$ is defined similarly.) In Section D.3 below I specify a set of moments $\mathbb{E}[m(z_t, \hat{u}_t(\theta))] = 0$ (where the left hand side is a column vector and the right hand side is a vector of zeros and the expectation is taken over hours). The (technical) condition for identification is that they hold if, and only if, we evaluate the function m at the true parameters θ (see for instance Newey and McFadden, 1994).

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

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

where W is a positive definite weighting matrix.

Under the assumption that the prediction error is orthogonal to the variables in z_t, $\hat{\theta}$ 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

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

D.3 Moments and weighting matrix

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

Specifically, I stack all $\hat{u}_t(\theta)$ into a vector $\hat{u}(\theta)$ of dimension $2\tau \times 1$ and define a $2\tau \times 2M$ matrix of exogenous variables Z in the following way: $Z = \begin{bmatrix} z & 0 \\ 0 & z \end{bmatrix}$ and $z = \begin{bmatrix} z_0 & z_1 & z_2 & z_3 \end{bmatrix}$ with z_0 being a vector of 1-s, z_1 containing the times until the draw dummies in the columns, z_2 containing GRP's and lags thereof in the columns, and z_3 being 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 $M = 359 * 2 = 718$ columns, meaning that there are 718 exogenous variables.[13] Using this, I calculate

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

I choose the weighting matrix W to be

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

[13] z_0 has 1 column. z_1 originally has 30 columns. z_2 contains GRP's from 5 groups of channels and 3 lags thereof, so it has 5*4=20 columns. z_3 has 332 columns. Most columns in z_1 are linear combinations of columns in z_3. After dropping those, z_1 has 7 columns left. Thus, there are in total $1+6+20+332 = 359$ columns.