Advertisements as barriers to entry: Evidence from the chocolate confection industry

By Chinmay Lohani*[1] and Eliana Sena[2]

[1]University of Pennsylvania
[2]University of Pennsylvania

Advertising is a highly significant marketing activity for most firms, yet there is limited evidence for it's effect on direct sales. In this paper, we explore another important reason why firms may advertise extensively in concentrated markets- to deter entry of competing products. We construct a merged dataset of TV advertising and sales for the US chocolate industry and find direct evidence for reduced entry. Reduced-form evidence using quasi-experimental variation from coarse spatial targeting highlights two facts- direct impact of advertising for own sales is small but a 10% *increase in total competitor ads reduces sales by* 1.6%. *Motivated by this evidence, we build a dynamic model of product entry and advertising. We estimate this model on my data to identify how much of firm advertising can be explained by entry-deterrence motives and run counterfactual policies that limit ad-behavior to analyse their welfare consequences.*

I. Introduction

Advertising is a very large industry in the United States, where firms spent $284.3 billion on ad-related expenditures in 2021. Most consumer-facing industries spend a substantial amount of money every year in advertising their products. This is especially true of Consumer Packaged Good (CPG) companies. For instance, P&G and Nestle spent $5.1 billion and $2.8 billion in advertising in the FY 2021- 22. Advertising and marketing of products represents very significant business activities for these firms.

However, as far as returns on these expenditures go, the evidence is mixed. Lodish et al. (1995*b*) analyse experimental data for groups which received more TV advertising and see persistent and positive effect on sales made to these customers. On the other

* University of Pennsylvania, clohani@sas.upenn.edu. We are grateful to our advisors Juan Camilo Castillo, Ulrich Doraszelski, and Aviv Nevo for their continued guidance and support. We also thank Francesco Agostinelli, Juan Pablo Atal, Sherrie Cheng, Ying Fan, Javiera Garcia, Ashley Schwanebeck, Katja Seim, Zach Weingarten, and participants of the Penn IO brown bag and Empirical Micro lunches for helpful feedback. Calculations are based on data from The Nielsen Company (US), LLC and marketing databases provided by the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen data are ours alone and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. All errors are our own.

hand, Blake, Nosko and Tadelis (2015) find limited effectiveness of paid advertising on e-Bay sales and Shapiro, Hitsch and Tuchman (2021*a*) find that TV advertising has zero to negative returns on investment for most brands. These studies together paint a meagre picture on the direct impact of advertising on consumer sales. If advertising doesn't change sales significantly, why do companies advertise? While the economics of advertising offers various theories that justify advertising, we study the industry's self prescribed view on why advertising is important. Advertising is deemed to be a long term 'brand-building' activity. According to a 2020 McKinsey report, CPG companies achieved their stature through, "Mass-market brand-building and product innovation, generating stable growth". Industry reports on marketing suggest that advertising lets brands be recognizable and stay clear of competition (McKinsey, 2020). We interpret these intuitions to specify a plausibly important mechanism of scale advertising- ads could be used by firms to maintain their position in a favorable market structure.

In this paper, we test whether, and how much of large scale advertising can be justified through the rationale of deterring product entry by competitors. We use the US chocolate industry as the focus of my empirical analysis. This is a well-established industry with large scale TV advertising, stable revenues and reasonable entry of new products during my sample, 2010-2020. We use reduced-form estimates and a structural model to understand how much of firm advertising can be attributed to entry-deterrence motives and what policies can be effective at increasing consumer welfare.

We show the empirical relevance of TV advertising in the chocolate industry using quasi-experimental variation generated by institutional features of the advertising market. To do this, we follow Shapiro, Hitsch and Tuchman (2021*a*) and assemble a brand–level panel that links monthly TV advertising to retail sales for the U.S. chocolate category from 2010 to 2020. Sales come from Nielsen RMS store–week UPC records that we aggregate to brand–market–month using fixed brand definitions that consumers recognize at the shelf. Advertising comes from Nielsen Ad Intel occurrence and viewership files, which we convert to monthly GRPs by Designated Market Area (DMA). We harmonize brand names across the two systems with a filtered fuzzy–match procedure and manual verification, and we map corporate parent identifiers to recover multi–brand ownership. GRPs are accumulated into an ad–brand stock using a standard depreciation rule so that short TV flights translate into a persistent brand capital measure seen by consumers, following Dube, Hitsch and Manchanda (2005), Shapiro, Hitsch and Tuchman (2021*b*), and Bar and Haviv (2021). The final dataset covers more than 4,000 brands across roughly 29,000 stores, with ad stocks and prices observed at the brand–DMA–month level.

The reduced–form analysis exploits institutional features of TV buying. Firms primarily purchase national and DMA–level inventory, local spot placements fill residual slots, and ad delivery is coarse in space but high frequency in time. Following Shapiro, Hitsch and Tuchman (2021*a*), we exploit DMA borders to construct quasi-experimental shifts in TV ad exposure and adopt their viewership–imputation approach for non-sweeps months. Identification comes from advertising discontinuities at borders while local demand fundamentals vary smoothly. The main specification relates log sales to own ad–brand and total rivals' ad–brand, with rich fixed effects, and isolates advertising variation using border contrasts so that local demand fundamentals vary smoothly across the boundary.

Three facts emerge. First, own TV advertising has a small but positive elasticity of sales, consistent with recent evidence on modest average returns (Shapiro, Hitsch and Tuchman, 2021*a*). Second, rivals' advertising is economically meaningful, a 10 percent increase in competitors' ads reduces a brand's sales by about 1.6 percent in the baseline border sample, which is suggestive of business stealing and is consistent with entry pressure working through advertising. Third, advertising is large in levels and highly seasonal, top advertisers maintain higher baseline exposure throughout the year while smaller brands concentrate around holidays, a pattern that is useful when disciplining the dynamic model.

Subsequently, we estimate a model of entry deterrence in this market. We estimate a demand model where consumer's purchases depend on the advertising they see. On the supply side of my model, firms choose two actions- pricing and advertising. Firms decide prices for their goods through a Nash-Bertrand model of competition. Firms also decide how much to advertise. We allow this decision to be made for reasons other than entry deterrence. When estimating the model on data, we use a 2-step procedure to identify advertising policy of firms. This will let us identify entry cost parameters- which depend on how entrant firms expect incumbents to advertise- to not be based on a model where firms only advertise to deter entry. We use our demand estimates and cost estimates to isolate the portion that can be justified by entry deterrence.

We use our estimated model to run counterfactuals where we shut off advertising incentives to maintain favorable market structure. We then estimate the effect of policies which limit the amount of advertising available to firms. Both of these counterfactuals will speak to the importance of entry-deterrence channel in this industry, and how policy prescriptions can affect consumer welfare.

A. Related literature

This project is related to a large literature on the economics of advertising. This literature has historically progressed by proposing various mechanisms through which advertising could influence market outcomes (see review by (Bagwell, 2007)). Subsequent papers have pushed the literature by testing these mechanisms quantitatively or empirically. Clark, Doraszelski and Draganska (2009) consider the effects of advertising on consumer awareness and perceptions of quality, while Doraszelski and Markovich (2007) study dynamic considerations and how "awareness" versus "goodwill" advertising implies different dynamics post-entry. On the empirical side, Lodish et al. (1995*a*) uses experimental variation in TV ads exposure to detect how advertising impacts sales. They find that significant impacts of advertising on sales which last for two additional years post advertising. Shapiro, Hitsch and Tuchman (2021*b*) uses quasi-experimental variation and finds meagre returns on investment for advertising by most brands. Our paper contributes directly to this strand of literature by empirically evaluating the effects of TV advertising for sales in the chocolate industry between 2010-20 and quantifying how much of firm advertising can be attributed to sales and entry-deterrence.

This paper also builds on the literature on endogenous market structure. The literature has explored various ways in which profit-maximising firms undertake actions to maintain favorable market structure. Goolsbee and Syverson (2005) studies how threat of entry by competitors can induce alternate pricing to prevent entry. Numerous non-price mechanisms- quality enhancements, R&D expenditures etc that are employed by firms to deter entry have also been studied. (Besanko, Doraszelski and Kryukov (2014), Ellison and Ellison (2007), Cockburn and MacGarvie (2006)). This paper explores the role of advertisements as an entry-deterrence instrument, in the spirit of Schmalensee (1978) and Sutton, 2007. As Sutton (2007) argues, firms in oligopolistic industries can use advertising to raise sunk costs of potential entrants substantially. Our model and empirical setting let us analyze the empirical relevance of entry deterrence for the market structure of the chocolate industry in the US.

While there is a substantial body of work that analyses advertising empirically, and a large body of work on entry deterrence and endogenous market structure, there is very little research that specifically analyses entry-deterrence through advertising in an empirical setting. Rizzo and Zeckhauser (1990) find evidence that physician advertisements could be justified as successive entry deterrents in local markets. Bar and Haviv (2021) is perhaps the closest paper to ours in terms of its scope. They use the anticipated entry of *Lays Stax* to estimates the value of a brand. The paper shows how *Pringles* might have used its advertisements to delay entry and suppress sales of *Lays Stax* in the US. Our paper tries to understand the overall entry-deterrence impact from advertising by all incumbents in an industry. By evaluating the extent of entry deterrence and its implications for consumer surplus, we can calculate the full welfare effects of advertisements as entry-deterrents in the chocolate industry.

II. Toy model

In this section, we write down an illustrative model of an industry where firms advertise to deter entry of new products. Each firm sells a single product. Firms live for 2 periods. For the sake of simplicity, we assume that products are identical (and therefore product characteristics are not relevant for choices made between these products), except for their prices and ad-brand. Ad-brand is a capital-like stock for each product. It increases when firms show ads (advertising is a flow variable) and depreciates across periods. Prices are assumed to be determined every period, and are therefore static decisions.

For this 2-period model, N_1 incumbent firms enter the first period with Ad-brand $((A_i)_{\{i=1..N_1\}})$. Firms set prices to maximize period profits in a Nash-Bertrand equilibrium. They choose the level of advertising (a_i) to increase their stock in period 2 such that firms maximizes their lifetime value. For simplicity's sake, we assume that incumbent firms don't exit [1], and that there are M potential entrants. Firms have expectations over facing $N_2 \in \{N_1, .., N_1 + M\}$ firms in the second period at the market. Denoting the firm

[1]This is not a conceptual restriction. The argument in appendix A can be extended to allow for firm exit. However, this substantially complicates Step 2 of the proof given there, given that the probability of facing a market with N firms can arise out of many combinations of firms entering and exiting.

i's lifetime value at the beginning of period 1 as $V(A_i, A_{-i}, N_1)$:

$$V(A_i^1, \boldsymbol{A_{-i}^1}, N^1) = max_{p_i^1, a_i} \pi(A_i^1, \boldsymbol{A_{-i}}^1, \boldsymbol{P}^1, N^1) - ca_i + \xi_i^1 + \beta \mathbb{E}[max_{p_i^2} \pi((1-\delta)A_i^1 + a_i, \boldsymbol{A_{-i}^2}, \boldsymbol{P^2}, N^2)]$$

A few notes on notation: c denotes per unit cost of advertising, (1-δ) is the depreciated Ad-brand over a period and β is how the firm discounts the future. $\boldsymbol{P}^j$ denotes the price vector in this market in period j, A_{-i}^j represents competitors' $-i$ ad-brands in period j. ξ_i^1 is a shock on a firm's per period profit. So firms have expectations over the distribution of ξ.

More firms can potentially enter the market in period 2. A firm maximizes lifetime profit, so they enter only when the expected profits from their product offset the cost of entry C:

$$V(A_j^1, \boldsymbol{A_{-j}}^1, N^1) = max_{\mathbb{1}(enter)} \mathbb{1}(enter)\Big[-C - ca_j + \beta \mathbb{E}[max_{p_j^2} \pi((1-\delta)A_j + a_j, \boldsymbol{A_{-j}^2}, \boldsymbol{P^2}, N^2)]\Big]$$

With stochastic ξ for all firms, this setup implies a derived distribution of the number of firms in this market in period 2. We denote this as $\mathbb{P}(N)$, which depends on the distribution of ξ and the firms' advertisement policies.

In this model, advertisement decisions can change market structure in the second period. If the expected profits of an entrant firm go down with more advertising by incumbent firms, it makes them less likely to enter.

We write down the first order condition for the advertising decisions of an incumbent:

$$c = \beta \frac{d\mathbb{E}[\pi_1(a_i, ., .)]}{da_i} + \underbrace{\beta \sum_{N_2 \in \{N_1, .., N_1+M\}} \big[\mathbb{E}\pi(., ., N_2). \frac{d\mathbb{P}(N_2, a_i)}{da_i}\big]}_{T_2}$$

THEOREM II.1: *If $\pi(A_i, A_{-i}, N)$ is strictly increasing in A_i, and strictly decreasing in A_{-i} and N for any point in the support of π, then T_2 is positive*

PROOF:

In appendix A

COROLLARY II.1.1: *A^* is higher in setup above, when $\frac{d\mathbb{P}(N)}{da_i} \neq 0$ than when $\frac{d\mathbb{P}(N)}{da_i} = 0$.*

Note that this corollary implies that advertising is higher when firms can impact future market structure. Thus, there are entry-deterrence motives in this model which lead to higher firm advertising.

We can focus on the FOC above to derive intuition for why we get deterrence in this model. Incumbents advertise to increase profits directly with respect to their competitors through increased demand from customers. Additionally, the second term underlines why firms advertise to deter entry- increasing advertising reduces the probability of entry because the potential value of entrants goes down.

The above argument only requires firms to have $\pi(A_i, A_{-i}, N)$ increasing in own advertisement A_i, decreasing in N and decreasing in A_{-i}. To drive our point further, we will solve this model explicitly by choosing some parametric assumptions.

For the remainder of the section, we assume that firms have a constant profit for every percentage share they occupy of the market (so markups are fixed). Firm shares depend on prices and ad-capital through the logit form. This gives per period profit as-

$$\pi(A_i, A_{-i}) = \Pi . \frac{e^{\alpha A_i - \beta p_i}}{1 + \sum_{j \in -i} e^{\alpha A_j - \beta p_j}}$$

There is one potential entrant. It has a deterministic entry cost $\mathcal{C}$ and an idiosyncratic shock to this cost drawn from a normal distribution with variance σ. Firm profits and ad-policies determine the probability of entry:

$$\mathbb{P}(N_1 + 1) = \Phi(\frac{\pi(a^*_{N_1+1}, A^*_{-N_1+1}, N_1 + 1) - C}{\sigma})$$

We can now further solve the first order condition for the incumbent firm as:

$$c = \alpha \Pi s_N (1 - s_N) + \frac{\beta\alpha}{\sigma}\Big[(\pi_N - \pi_{N+1}) e^{A_{N+1}} . s_N)\Big]$$

where $s_N := \frac{e^{\alpha A^*_i - \beta p^*_i}}{1 + N . e^{\alpha A^*_i - \beta p^*_i}}$, p^*_i, A^*_i determined in symmetric steady state equilibrium.

Thus, for sufficiently high difference between profits of the incumbent with or without the additional entrant, a large portion of firm level advertising can be justified to deter entry. This is increasing in the share of incumbent firms s_N, effectiveness of advertising in moving demand α and decreasing in the variance of entry shocks σ. We simulate a fully deterministic version (without shocks ξ) as a proof of concept in Appendix E.

III. Data

The main sources of data for this project are the Nielsen AdIntel and Nielsen Retail Management Scanner (RMS). We proceed with our analysis by constructing a data merge between these data to have a month level dataset on sales and advertising by a chocolate brand from 2010 to 2020.

A. Nielsen RMS data

The Nielsen RMS (Retail Measurement Services) data include weekly store-level information on prices and quantities sold at the UPC level. The RMS data include information for about 40,000 stores, including grocery stores, drug stores, mass merchandisers, and convenience stores. It represents a non-random subset of these stores, but has a wide coverage. This data covers more than half of all market-level spending in grocery stores and one-third of all spending at mass merchandisers. We use RMS sales data from 2010-2020, and we aggregate up from UPCs to the brand level. We define a brand as all forms of the same consumer good, as indicated by the brand code or brand name in the RMS data. The reasons for this are two-fold: (i) RMS brands are the right level of generality for a product to be identified as an advertising brand. This is conceptually meaningful, as consumers likely associate advertisement to the brand rather than to a specific variant of the brand that has different size or packaging, all of which would be recorded as a different UPC product in the data. Moreover, this is also practically the best level of aggregation to match specific advertising entities to specific RMS brands. (more details in Appendix B) (ii) This paper focuses on entry. UPCs are noisily recorded in the data, and using new UPCs to denote product entry might introduce lots of measurement error and potential biases. In our final sample, we have 4000+ brands present in over 29,000 stores.

B. AdIntel data

We use advertisement data from Nielsen Ad Intel database for the years 2010-2020. This advertising is recorded as occurrences, where an occurrence is defined as the specific instance of a product's ad being shown for a brand for a $(channel, DMA, time)$ combination. Occurrence data is used along with viewership data, which gives estimates of how many impressions were created from airing an ad, i.e., how many people were reached. These impressions are recorded consistently across time for top 25 DMAs, but for the rest, they are recorded using diaries filled by households than Nielsen surveys. These diary data are only recorded in the four "sweeps months," February, May, July, and November. Following Shapiro, Hitsch and Tuchman (2021*b*), we impute the viewership for other months using a weighted average for viewership between the closest recorded months for these DMAs.

The TV advertising data comes from four different TV media types in the dataset- Cable, Network, Syndicated, and Spot. Cable ads are aired nationally and their viewership is only available at that level (so it induces across time level differences in the data). Network and Syndicated ads have national occurrences which are mapped to local viewership data. Overall, the cross-sectional variation in the aggregate ads data comes from Spot impressions and Network and Syndicated ads.

We use impressions data along with Universe estimates, which are a measure of TV viewing households in a given region, to create GRPs. GRPs are defined as the fraction of TV viewing households reached by a given ad. These GRPs are the flow variable into our defined Ad-brand, which is the stock variable in our analysis. We aggregate GRPs for a given month by adding up the GRPs for a given ad-brand in a DMA.

C. AdIntel- RMS merge

We merge the RMS data at the brand level (subsequently called RMS brand) and advertisement data at brand level (subsequently called ad-brand) through a filtered fuzzy string match process with data pre-selection. This is necessary because there is no direct linkage between the two datasets. Furthermore, the two brands can and do differ in who they refer to, necessitating a manual vetting procedure at the end. We create two sets of matches- specific matches where ad-brand and RMS-brand refer to the same object, and general matches- where general Ad-brands are matched to multiple RMS brands under their umbrella (e.g.: "Hershey's chocolates"). We include details for this data merge in appendix B.

D. Geocoded data

Finally, we use geographic information from open source approximations to DMA boundaries and data from US Census to identify stores which are within counties that lie on DMA borders.

IV. Empirical analysis

We highlight some key features of this market that emerge from this data.

US chocolate industry is large and has steady revenue

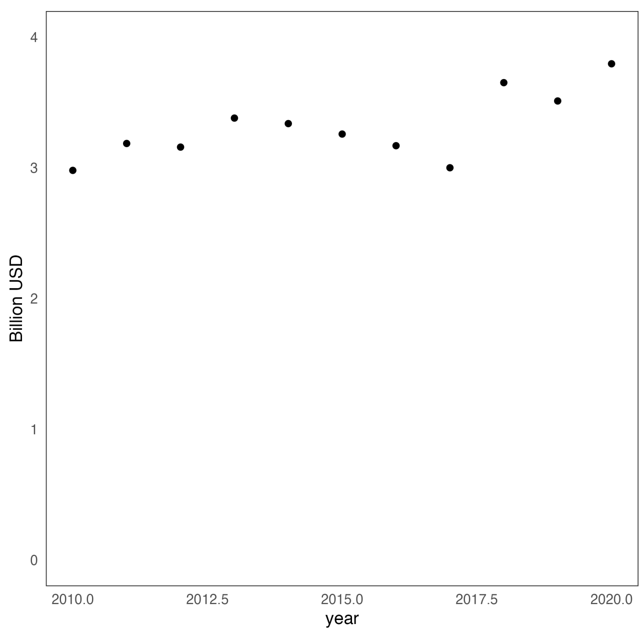


Figure 1. : Chocolate revenues are steady (accounting for inflation)

We calculate the total revenue registered for these products every month and adjust for inflation by using a deflator that scales nominal dollars to 2012 dollars. Adding these up gives yearly estimates of revenue for the chocolate industry from the sample of Nielsen stores. Contrary to some CPG goods industries, chocolate enjoys a steady revenue in these markets.

Entry trends- decreased entry and entry is slower

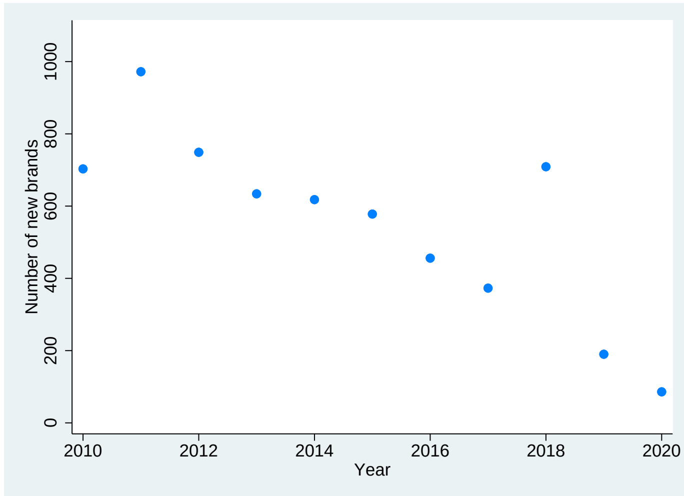


Figure 2. : New brand variants are fewer

Within the chocolate industry, we find that entry of new brands decreases over time. This is in line with similar trends across the CPG goods industry (see appendix F). Note that, we define entry here as the occurrence of a new RMS brand variant in any store in our sample anywhere. While the RMS data we use for subsequent analysis starts in 2010, we use an extended sample of all chocolate products in the raw Nielsen files from 2006 to define entry. This alleviates concerns that our definition might be over-counting entry, because we allow for a substantial number of established products to show up in the data from 2006-10. We check the sensitivity of these trends by changing the size of this pre-period and the figure shows no qualitative change. Reduced entry is consistent with higher entry costs or lower expected revenues within our model.

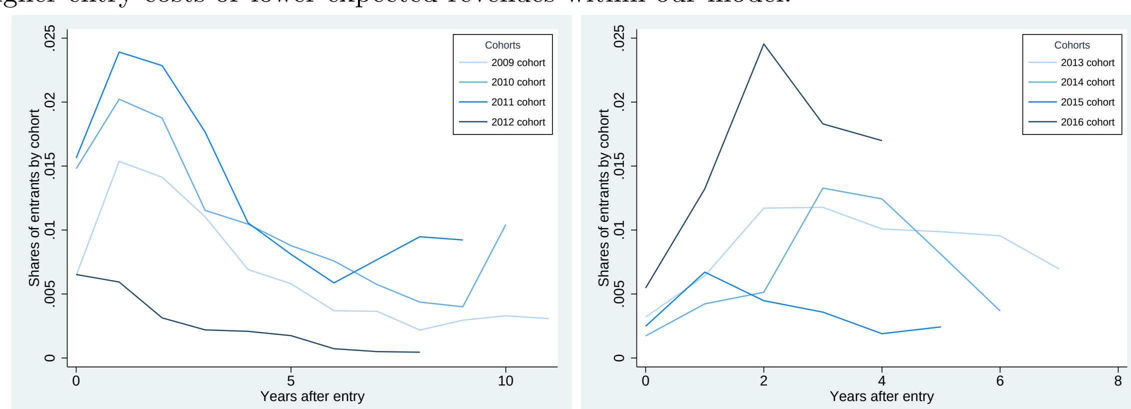


Figure 3. : New cohorts enter more gradually

We analyse trends in entry for a given product over time. To do this, we calculate the share of the market occupied by a new brand every year after it enters. We average these shares for all brands which entered in the same year, creating successive cohorts of post-entry data. We plot shares for each year after entry for these different cohorts. What emerges is that later cohorts take longer to occupy their peak market shares. Our model of entry can be used to think about entry across geographies. A secular, higher cost of entry across markets is consistent with this pattern. Here, each successive market that a product enters will take longer and happen only when a suitably large idiosyncratic shock strikes.

Advertising policy is similar across time

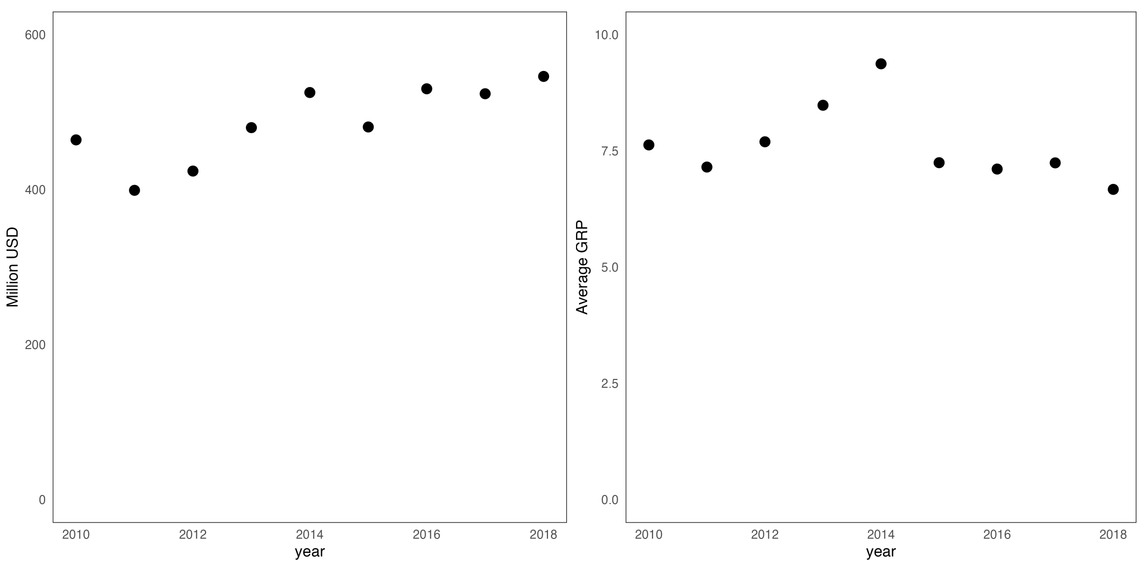


Figure 4. : Advertising program is similar across years

We plot the total money spent by firms in the chocolate industry in every year of our sample period (in 2012 dollars) and find that the aggregate expenditure is similar across years. We also plot the average GRP rating for all TV advertising by this industry in each year for this period. The average GRP appears stable, except for a slight increase near the middle of our sample. This could be because of more effective advertising during non-election years, as has been shown previously by Sinkinson and Starc (2018)

Advertisement is seasonal, spending is large

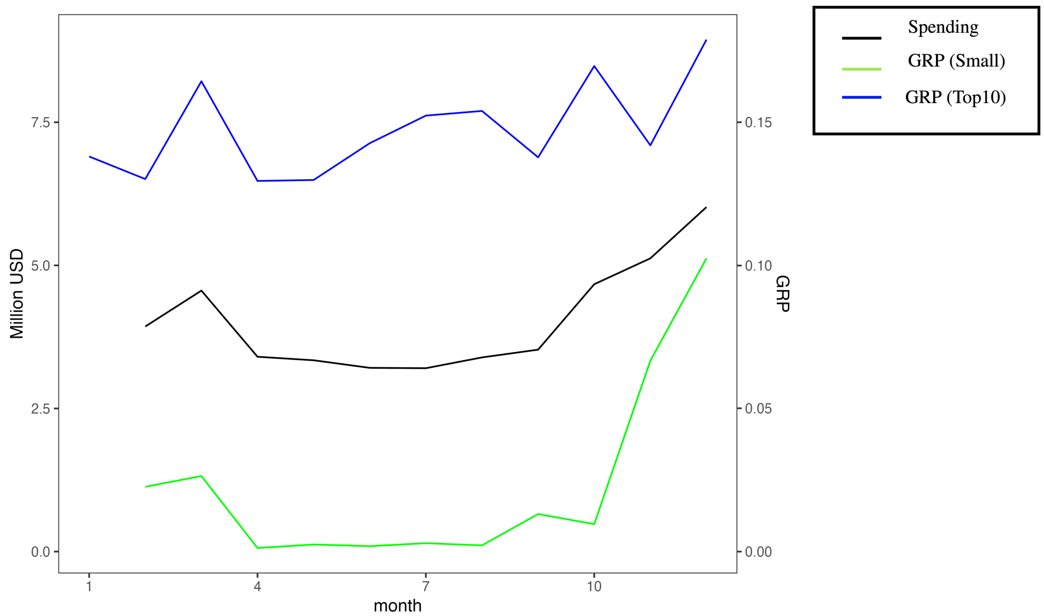


Figure 5. : Average GRPs and spending exhibits strong seasonal patterns

We analyse the advertising data to understand within year variation in advertising by chocolate companies. We plot expenditure in each month averaged across all years in our sample, and find strong seasonal patterns which seem to follow cultural holidays which generate higher chocolate demand (Easter, Valentine's day and Christmas). We also calculate average GRPs for ads in each month. We split this average for two groups: the top 10 advertisers in our sample (by total expenditure across years) and by all other firms. Besides secular higher GRPs for the top firms across all months, we also find that top firms' advertising is relatively less seasonal than other firms. This aligns well with a brand-building motive: top firms advertise less in response to demand shocks but instead maintain a large level of advertising over the entire year.

V. Reduced-form effects of advertising

We find that advertising (1) is a large fraction of revenue in the chocolate industry, (2) tracks demand patterns within a year for many firms. In this section, we want to understand how advertising generated ad-brand impacts sales.

Following the literature (Dube, Hitsch and Manchanda (2005),Shapiro, Hitsch and Tuchman (2021*b*), Bar and Haviv (2021)), we define Ad-brand as a stock variable composed of discounted sum of GRPs as the investment variable. This is to capture lasting effects of advertising over time. We broadly follow prior literature in building the main specification for our ad-brand: we shut off impacts for a given ad after a year, and choose a linear discount factor of 0.7 per month. This closely corresponds to a factor of 0.9 per week, which is used by the previous literature (Dube, Hitsch and Manchanda (2005)Shapiro, Hitsch and Tuchman (2021*b*), Bar and Haviv (2021)). This gives steady levels of ad-brand 6. We will use this as our main treatment variable which varies ad exposure.

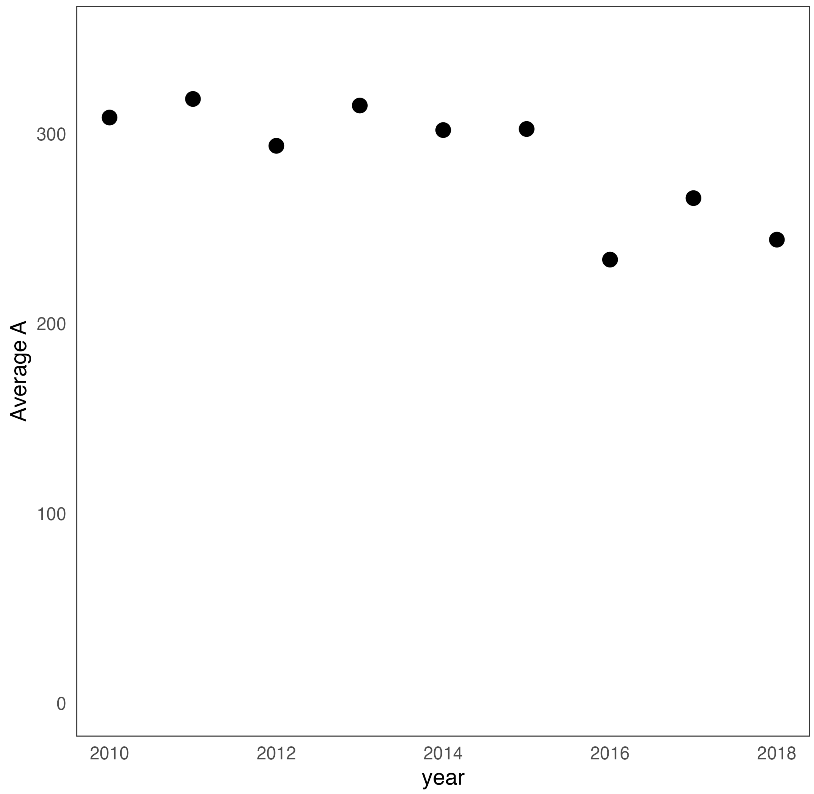


Figure 6. : Ad-brand with $\delta = 0.7$ month-on-month

A. Estimation strategy

We estimate a linear fixed effects models to identify the effects of advertising on sales. For a product i in store s for time t, we use the following model as our main specification-

$$log(q_{ist}) = \beta log(A_{ist}) + \phi log(A_{-ist}) + \alpha' log(\boldsymbol{p_{st}}) + \gamma_s + \delta_t + \eta_i + \xi_{ijt} + \epsilon_{it} \tag{1}$$

Here, A_{ist} is ad-brand of the product i, A_{-ist} represents total ad-brand of the rest of the products in the market. [2]

Ex-ante, one would expect there to be considerable endogeneity in this setting that could bias our estimates. Particularly, local demand shocks which are foreseeable could

[2] We run specifications with top 5 ad-brands added to this regression and find similar results.

be correlated with advertising policies of firms. Firms could be advertising to increase their relative shares during times of generally higher demand, which would induce positive bias in our elasticity estimates. Similarly, if cost of advertising is highly correlated with broader demand shocks, some firms may advertise less in response to these shocks giving me a negative bias. We use ξ_{ijt} to represent these unobserved product specific shocks that may bias our estimates.

We use the institutional features of this setting to select a sub-sample of stores that has quasi-random variation in advertising. Following (Shapiro, Hitsch and Tuchman, 2021*a*), we use the fact that since companies can only buy TV ads up to the DMA level, this is the finest level of targeting they can achieve for their ads. Another institutional detail that helps me is that the bulk of advertising is bought in the upfront ad-market early on in the year. This is forecasted, seasonally dependent variation and is plausibly predictable with location and time fixed effects. The remaining variation in advertisements comes from local Spot ads, which are difficult to place precisely in time. These often fit into residual timeslots available to local TV stations. Since there is large variation in the number of people reached through these timeslots, it induces variation in the GRPs.

There is remaining concern that if there are local demand shocks which can be coarsely targeted in time through local channels, advertising would still be potentially correlated with local shocks across time. In order to effectively address this, our main specification will construct a sub-sample of stores that faces comparable local shocks. We select stores which are on the boundary of DMAs. Our linear model uses the variation in ad-brands for stores lying which face otherwise similar local conditions but lie across DMA borders.

Identifying assumption

The identifying assumption for this model is that the unobserved shocks to product sales vary smoothly across DMA boundaries. Across DMA variation in ad-brand is discrete across these borders, and are uncorrelated with changes in these local shocks.

Given that we focus on across DMA stores, unobserved location-product shocks should be nearly similar across stores. After controlling for month and location fixed effects, the leftover variation in advertising varies discretely across states and should be uncorrelated with these local shocks.

B. Estimates

We estimate this model first on the entire sample of stores, and then on the subset of stores which are in counties that lie on DMA borders. [3]

[3] Since we only see county level geographic information on the stores, this is the finest selection of stores we can achieve.

	$log(y_i)$			
	Full sample		Border sample	
	(1)	(2)	(3)	(4)
$log(p_i)$	-0.3420 **	-0.3198**	-0.3421**	-0.3263**
	(0.0068)	(0.0021)	(0.0086)	(0.0022)
$log(A_{brand(i)})$	0.0019	0.0017	0.0018	0.0014
	(0.0012)	(0.0090)	(0.0012)	(0.0012)
$log(A_i)$	0.0017**	0.0026**	0.0020**	0.0028**
	(0.0008)	(0.0009)	(0.0009)	(0.0009)
$log(A_{brand(i)}) \times \mathbb{1}(new)$		0.0081**		0.0075**
		(0.0018)		(0.0017)
$log(A_i) \times \mathbb{1}(new)$		-0.0036		-0.0034
		(0.0027)		(0.0022)
$log(\mathbf{p_{-i}})$	0.973**	0.883**	1.000**	0.9168**
	(0.0320)	(0.0281)	(0.0336)	(0.0373)
$log(A_{-i})$	-0.2408**	-0.2073**	-0.1639**	-0.1638**
	(0.0579)	(0.0591)	(0.0592)	(0.0592)
$log(A_{-i}) \times \mathbb{1}(new)$		0.0107**		0.0108**
		(0.0011)		(0.0011)
N	408328461	408328461	207512796	207512796
Fixed-effects				
Brand	Yes	Yes	Yes	Yes
Store	Yes	Yes	Yes	Yes
Month of year	Yes	Yes	Yes	Yes

Table 1—: Table of log(quantity)- log(ads) estimates. Column (1) and (2) use the entire sample, Column (3) and (4) use the sample of stores which are on DMA boundaries. (2) and (4) have effects of advertising separately identified for new brands (< 2 years old). Standard errors are clustered at the Store×Year level.

Since this is a log-log model, we can interpret the coefficient on advertising variables as elasticities. We replicate findings similar to (Shapiro, Hitsch and Tuchman, 2021*a*) with respect to own-advertising elasticity- the average industry specific elasticity of advertising is low at 0.002. Since we use advertising by all the players in the industry, rather than top brands and top competitors, this speaks to general advertising effectiveness within the chocolate industry. These elasticity estimates are positive and significant for all our specifications.

Importantly, we find a large, negative and significant estimate for total advertising by other brands. This implies that increases in total advertising by competing firms reduces own sales. This speaks directly to the incentive to advertise to reduce competitors sales. Even if a large amount of advertising is not meaningfully important for own sales, it can contribute to deter entry for competing products.

We investigate whether the effects are different for new firms. Specifically, we separately identify these coefficients for new products that enter our data within 2 years of their entry. These estimates, when significant, show mildly smaller effects of ads on sales by new

products. This suggests that advertising by firms on TV doesn't have very different effects of sales of new products.
As robustness checks, we re-estimate this model with different specifications. We estimate the model with stores from state borders removed. In order to check that the results are not driven by stock buildup, we estimate the model only with GRPs: the flow variable. We also relax the functional form assumption by estimating a linear-linear model. We find similar results across all these specifications. For details, refer to appendix C.

VI. Model

Motivated by these reduced-form results, which show that large volume ads by competitors can indeed decrease sales of products, we write down a model that captures the demand and supply in these markets. This model identifies demand parameters which capture substitution patterns between products based on relevant exogenous and endogenous (price, advertisements) parameters. The supply model identifiers features of the strategic interaction between firms in determining product prices, and advertising.

A. Demand

We model product demand using a nested random coefficient hedonic model (Berry, Levinsohn and Pakes (1995),Nevo (2001) ,Petrin (2002)).

Nesting structure.

We use a single, mutually exclusive partition of products into nests $g \in G$ (either $G = \{\text{Standard}, \text{Premium/Gifting}, \text{Dietary/Specialty\&Baking}\}$ or $G = \{\text{Standard}, \text{Premium/Gifting}, \text{Dietary/Specialty}, \text{Baking}\}$). Let J_g denote the set of products in nest g for market m, time t.

Utility.

For product $j \in J_{g(j)}$ in market m and period t,

$$u_{ijmt} = \underbrace{X_{jt}\beta + \alpha p_{jmt} + \rho_A A_{jmt} + Z_{jmt}\theta}_{\delta_{ijmt}} + \varepsilon_{jmt}, \qquad \varepsilon_{jmt} \text{ nested T1EV with parameter } \lambda \in (0,1),$$

where X_{jmt} are core observables (size, cocoa type, etc.), p_{jmt} is price, A_{jmt} is ad-brand capital, and Z_{jmt} are format/seasonality flags (e.g., minis/bites, wafer/cookie, seasonal). The single inclusive-value parameter λ governs within-nest correlation.
Define the within-nest inclusive value $IV_{gmt} \equiv \sum_{k \in J_g} \exp\{\delta_{kmt}/\lambda\}$.
Then:

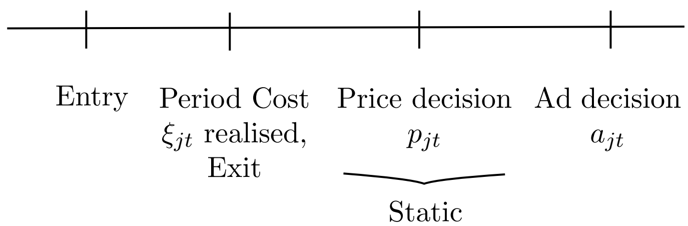
$$s_{j|g} = \frac{\exp(\delta_{jmt}/\lambda)}{\sum_{k \in J_g} \exp(\delta_{kmt}/\lambda)} = \frac{\exp(\delta_{jmt}/\lambda)}{IV_{gmt}}, \qquad S_g = \frac{(IV_{gmt})^{\lambda}}{\sum_{g' \in G} (IV_{g'mt})^{\lambda}}.$$

The market share of product j is the product of the nest share and the conditional share:

$$s_{jmt} = S_{g(j)} \cdot s_{j|g(j)} = \frac{\exp(\delta_{jmt}/\lambda)\ (IV_{g(j)mt})^{\lambda-1}}{\sum_{g'\in G}(IV_{g'mt})^{\lambda}}.$$

With these shares, and market sizes M_{mt}, firm prices yield consumer demand.

B. *Supply*



Static Price Setting

Firms set prices in a Nash-Bertrand static game.
Given the mean-utility vector $\boldsymbol{\delta}_{mt}$ and rivals' prices $\mathbf{p}_{-f,mt}$, firm f chooses its product prices to maximize static profit:

$$\max_{\{p_{jmt}\}_{j\in\mathcal{J}_f}} \big(p_{jmt} - mc_{jmt}\big)\, s_{jmt}\big(p_{jmt}, \mathbf{p}_{-fmt}; \boldsymbol{\delta}_{mt}\big)\, M_{mt}$$

Equilibrium prices are those that solve this problem for all firms simultaneously.

Advertising Policy

We allow firms to advertise for reasons other than ad-deterrence. There are various possible theories for why firms may advertise outside of influencing consumer demand or shaping favorable market structure (Bagwell, 2007). For instance, managers in larger firms may be risk averse against experimenting away from established advertising patterns, as posited by Shapiro, Hitsch and Tuchman (2021*b*). This essentially results in a principle-agent problem, where the optimal action is not undertaken by a manager who is sufficiently exposed to risky outcomes (Shavell, 1979). Another reason in this market could be related to vertical bargaining between firms and wholesalers. If firm bargaining positions are influenced by their advertising capabilities, this might result in better payment schemes and product positioning for firms' products *vis-a-vis* wholesalers (Morton and Zettelmeyer, 2004).
Motivated by these arguments, and the empirical patterns observed for advertising in this market, we model firm advertising policy through additively-separable components related to demand & market structure (through entry deterrence via lower future demand), A_d^* and additional size and time dependent policies $\tilde{A}$. This is a function of the relevant state s:

$$A(s) = A_d^*(s) + f(r_j).\zeta_t + \mu_{jt}$$

Here, other supply related reasons to advertise are captures through a flexible function $f(.)$ of firm size (in revenue) r_j. In order to capture seasonal changes in these advertising supply strategies, we identify scaling factors ζ_t. μ_{jt} is a shock drawn from a normal distribution with variance σ_s^2 that explains other variation in prices.
$A_d^*(s)$ is determined within the model as the optimal, profit maximizing level of advertising by a firm in state s. This is the level of advertising firms would arrive at if $A(s) = A_d^*(s)$ for all firms. In general, firms form rational expectations about $A(s)$ for all players in the future, bearing in mind that potential entry by competitors will change market structure and future profits.

$$V(s) = max_{a_j} \pi(\mathbf{\Phi}, A_j, \boldsymbol{A_{-j}}, \boldsymbol{P}) - ca_j + \beta \mathbb{E}[V(s')]$$

Here $\mathbf{\Phi}$ represents a vector of mean utilities of each product, $\boldsymbol{P}$ represents the vector of prices and A_{-j} represents the vector of competitors' ad-brands.

Entry

A potential entrant arrives every period, with a mean value drawn from a lognormal distribution with mean μ_{pe} and variance σ_{pe}^2.

$$\nu \sim \text{Lognormal}(\mu_{pe},\, \sigma_{pe}^2),$$

so that $\ln \nu \sim \mathcal{N}(\mu_{pe}, \sigma_{pe}^2)$.
The entrant enters if this value exceeds the entry cost $\mathcal{C}$. For a potential entrant, the value of the product is:

$$V(s) = -\mathcal{C} + \pi(\mathbf{\Phi}, A_j, \boldsymbol{A_{-j}}, \boldsymbol{P}) + \mathbb{E}[V((s'))]$$

Firm have rational expectations about competitors' advertising policies and evolution of the state variable s.

C. Equilibrium.

A *Markov Perfect Rational Expectations Equilibrium (MPREE)* consists of: (i) incumbent policy functions $(p_j(s), a_j(s))_{j \in \mathcal{J}}$, (ii) an entry rule for potential entrants, and (iii) a law of motion $\mathcal{P}(s'|s)$, such that:

1) **Incumbent optimality.** For each j and every state s, $(p_j(s), a_j(s))$ maximizes

$$\pi_j(\mathbf{\Phi}(s), A_j, \boldsymbol{A_{-j}}, \boldsymbol{P}(s)) - C\big(a_j(s)\big) + \beta\, \mathbb{E}\big[V_j(s') \,\big|\, s,\ a_j(s),\ a_{-j}(s)\big]\,,$$

taking rivals' Markov policies $(p_{-j}(\cdot), a_{-j}(\cdot))$ as given.

2) **Static pricing.** Given s and rivals' prices, $p_j(s)$ solves firm f's Nash–Bertrand static profit problem with demand $s_{jmt}(\cdot)$ implied by $\boldsymbol{\delta}(s)$.

3) **Entry optimality.** A potential entrant with draw $\nu \sim \text{Lognormal}(\mu_{pe}, \sigma_{pe}^2)$ enters if and only if

$$V^{\text{ent}}(s) \;\geq\; \mathcal{C}, \quad \text{where} \quad V^{\text{ent}}(s) = \pi^{\text{post}}(s) - \mathcal{C} + \beta\, \mathbb{E}\big[V^{\text{post}}(s') \,\big|\, s\big]\,.$$

4) **Consistency of beliefs.** The transition $\mathcal{P}(s'|s)$ is generated by the ad–stock law $A'_j = (1-\delta)A_j + a_j(s)$, the entry rule, exogenous shock processes, and the policy profile $\{(p_j(\cdot), a_j(\cdot))\}_{j\in\mathcal{J}}$. Agents' expectations are correct: perceived and actual $\mathcal{P}$ coincide.

5) **Market feasibility.** Shares s_{jmt} integrate to market size M_{mt} and the realized product set evolves as $\mathcal{J}' = \mathcal{J} \cup \mathcal{E}$.

VII. Estimation and Identification

We estimate the model in two stages. First, we estimate the static demand system described above, recovering how prices and advertising affect product demand and substitution patterns. Second, taking these demand parameters as given, we estimate the dynamic component of firm behavior related to advertising and entry deterrence.

A. *Consumer Demand*

Demand estimation follows directly from the nested logit structure in the Model section. Using the standard inversion for one–level nested logit, the estimating equation is

$$\ln s_{jmt} - \ln s_{0,mt} = X_{jmt}\beta + \alpha\, p_{jmt} + \rho_A A_{jmt} + Z_{jmt}\theta + (1-\lambda)\ln S_{g(j)mt} + \xi_{jmt},$$

where s_{jmt} is the observed share, $s_{0,mt}$ is the outside share, $S_{g(j)mt}$ is the observed nest share, and ξ_{jmt} captures unobserved demand shocks.
A key identification challenge is that both prices p_{jmt} and advertising A_{jmt} may be correlated with ξ_{jmt}. We address this endogeneity using instruments Z_{jmt} that shift prices and advertising but do not directly affect demand shocks:

$$\mathbb{E}[\xi_{jmt} \mid Z_{jmt}] = 0.$$

For prices, we use standard differentiation instruments based on rival characteristics, as well as variation induced by media-market boundaries that affects competitive conditions and pricing incentives. For advertising, we use media-market boundary variation that shifts advertising exposure and cost but does not directly change consumer tastes. These exclusions require that advertising reach and price discontinuities at DMA borders affect demand only through firms' advertising and pricing decisions, not through preferences.
We estimate the demand parameters by GMM using these instruments.

VIII. Supply Estimation

We now turn to the supply side. The goal is to recover how firms choose advertising in a dynamic environment where current advertising affects both current demand and future market structure through entry deterrence. Observed advertising reflects three forces: (i) own–demand returns, (ii) entry–deterrence incentives, and (iii) other systematic motives summarized by $f(r_j)\zeta_t$. Our objective is to separate these components in the data.
We estimate supply in two steps. Step 1 recovers the static pricing behavior implied by the estimated demand system, and estimates a reduced–form entry hazard that links incumbents' ad stocks to future entry. Step 2 uses local revealed–preference comparisons, in the spirit of **?**, to decompose observed advertising into own–demand, deterrence, and residual components, without solving the full dynamic game.

Step 1: Entry hazard

We want to identify the effect of incumbent advertising on the timing of entry. We estimate a reduced–form entry hazard that tells us how increase in ads today shifts the probability that a new product appears sooner rather than later. Identification comes from media–market (DMA) border variation that changes ad exposure while holding local demand conditions smooth across the border.

Discrete–time hazard.

For market m and period t, let $E_{m,t+1} \in \{0, 1\}$ indicate that a new product enters between t and $t+1$. Conditional on no prior entry in the spell, define the hazard

$$h_{mt} \;=\; \Pr\bigl(E_{m,t+1} = 1 \,\big|\, \text{no entry by } t,\ s_{mt}\bigr).$$

We parameterize

$$h_{mt} \;=\; \Lambda\bigl(\gamma_0 \;+\; \gamma_A\, \mathcal{A}_{mt} \;+\; X'_{mt}\gamma \;+\; \psi_m \;+\; \tau_t\bigr)\,,$$

where $\Lambda(\cdot)$ is logit or complementary log–log, $\mathcal{A}_{mt}$ is a flexible function of incumbents' ad stocks (e.g., levels and moments), X_{mt} are controls, ψ_m are market fixed effects, and τ_t are time fixed effects.

Border instruments and exclusion.

To address endogeneity of $\mathcal{A}_{mt}$, we use DMA–border instruments Z^A_{mt} that shift ad exposure discretely across borders while preferences and costs vary smoothly:

$$\mathbb{E}\bigl[\eta_{mt} \,\big|\, Z^A_{mt}, X_{mt}, \psi_m, \tau_t\bigr] = 0,$$

with a first–stage

$$\mathcal{A}_{mt} \;=\; \pi_0 \;+\; \pi_1\, Z^A_{mt} \;+\; X'_{mt}\pi \;+\; \psi_m \;+\; \tau_t \;+\; u_{mt}, \qquad \pi_1 \neq 0.$$

From hazard to transitions.

The estimated hazard $\hat{h}_{mt}$ gives the probability of entry in the next period. Combined with the observed law for other state variables, this yields the transition kernel $\widehat{\mathcal{P}}(s_{m,t+1} \mid s_{mt}, a_{mt})$ that we use in the dynamic supply step.

A. *Step 2: Advertising Motives*

Our goal is to decompose observed advertising into three parts at a given state s: (i) the level justified by demand-side returns, (ii) the extra justified by entry deterrence, and (iii) the residual "other motives." We proceed in a revealed-preference spirit: use the estimated demand system and the estimated entry hazard to ask whether small deviations from the observed ad choice would raise or lower the firm's value. Critically, we build the decomposition *sequentially*: first rationalize ads via demand alone; then, starting from that demand benchmark, layer in the value of deterring entry; finally, compare the combined benchmark to what firms actually do.

Step 1: Demand-only benchmark

Define the demand-side marginal benefit of advertising at (j, m, t):

$$\mathrm{MB}_j^{\mathrm{dem}}(s; A) \;=\; \frac{\partial \pi_{jmt}}{\partial A_{jmt}} \;+\; \beta \, \frac{\partial A_j'}{\partial A_{jmt}} \cdot \frac{\partial V_j^{\mathrm{own}}(s')}{\partial A_j'},$$

where π_{jmt} uses the estimated demand system (including ρ_A) and A_j' is the next-period ad stock (given the ad-stock law already specified above).
Current-period term. Using the chain rule,

$$\frac{\partial \pi_{jmt}}{\partial A_{jmt}} \;=\; \big(p_{jmt} - mc_{jmt}\big)\, M_{mt} \qquad \underbrace{\frac{\partial s_{jmt}}{\partial A_{jmt}}}_{=\;\rho_A \cdot \Psi_{jmt}(\hat{\theta}; s)} \qquad .$$

Here $\Psi_{jmt}(\hat{\theta}; s) \;:=\; \frac{\partial s_{jmt}}{\partial \delta_{jmt}}$ is the (known and computable) share derivative with respect to mean utility under the estimated nested-logit system; we multiply by ρ_A because $\partial \delta_{jmt}/\partial A_{jmt} = \rho_A$.
Continuation term (demand channel only). Given the ad-stock law, the marginal effect of raising A_{jmt} by one unit today propagates to period $t + \tau$ as $(1 - \delta)^{\tau}$. Holding rivals' policies fixed (as in BBL), the continuation derivative is

$$\frac{\partial V_j^{\mathrm{own}}(s')}{\partial A_j'} \;=\; \mathbb{E}\left[\sum_{\tau \geq 1} \beta^{\tau-1} \left\{ \big(p_{jmt+\tau} - mc_{jmt+\tau}\big)\, M_{mt+\tau}\, \rho_A\, \Psi_{jmt+\tau}(\hat{\theta}; s_{t+\tau}) \right\} (1 - \delta)^{\tau-1} \;\middle|\; s \right],$$

where the expectation is taken over demand shocks and the (estimated) state evolution, keeping rivals' observed policies fixed.

The *demand-only* optimal advertising level $A_j^{\text{dem}}(s)$ solves the scalar first-order condition

$$c \;=\; \text{MB}_j^{\text{dem}}\big(s; A_j^{\text{dem}}(s)\big).$$

In practice we compute $A_j^{\text{dem}}(s)$ by a local search (BBL): evaluate the value difference for small deviations $\{A_{jmt} \pm \varepsilon\}$ around the observed state s, with rivals' policies held fixed; using bisection to find the A at which $\text{MB}_j^{\text{dem}}(s; A)$ crosses c. This yields the demand-rational benchmark that we will start from in the deterrence step.

Step 2: Add entry deterrence on top of demand

Let $h(s)$ be the estimated entry hazard from Step (1) (see VIII) and let $\Delta V_j(s^+) := V_j^{\text{no entry}}(s^+) - V_j^{\text{entry}}(s^+)$ be the continuation-value gap between paths with and without a new entrant, computed under the observed policy profile.
The marginal benefit from entry deterrence, evaluated at the demand benchmark, is

$$\text{MB}_j^{\text{det}}\big(s; A_j^{\text{dem}}(s)\big) \;=\; \beta \left[-\, \frac{\partial h(s)}{\partial A_{jmt}} \right]_{A=A^{\text{dem}}} \cdot \Delta V_j(s^+).$$

We then find the *demand+deterrence* optimal level $A_j^{\text{dem+det}}(s)$ as the solution to

$$c \;=\; \text{MB}_j^{\text{dem}}\big(s; A_j^{\text{dem+det}}(s)\big) \;+\; \text{MB}_j^{\text{det}}\big(s; A_j^{\text{dem+det}}(s)\big),$$

starting the search from $A_j^{\text{dem}}(s)$.
This respects non-additivity: deterrence is evaluated and re-equated jointly with demand-side returns, conditional on the demand benchmark.
We compute $\text{MB}_j^{\text{det}}\big(s; A_j^{\text{dem}}(s)\big)$ in three steps.
(i) Hazard slope. Since entry is modeled via a logit hazard,

$$h(s) = \Lambda\Big(W(s)^\top \widehat{\beta}\Big), \quad \frac{\partial h(s)}{\partial A_{jmt}} = \widehat{\beta}_A \cdot h(s)\big(1 - h(s)\big),$$

where $\widehat{\beta}_A$ is the estimated coefficient on incumbent advertising exposure in the hazard.
(ii) Continuation-value gap. To obtain $\Delta V_j(s^+)$, we construct two next-period states that differ only in whether a new product is present and simulate forward profits under the observed policy profile. For each horizon $\tau \geq 0$, we compute

$$\pi_{j,t+\tau}^{\text{no entry}} \quad \text{and} \quad \pi_{j,t+\tau}^{\text{entry}}$$

using the estimated demand system and the Nash–Bertrand pricing mapping from Section **??**. Future profits are discounted at β, and the sequence is truncated at a horizon

T such that additional discounted terms are negligible; the tail is approximated by an autoregressive continuation factor estimated from the simulated path.

(iii) Demand+deterrence optimal advertising. Given the demand benchmark $A_j^{\text{dem}}(s)$, the optimal advertising level accounting for both demand and entry deterrence, $A_j^{\text{dem+det}}(s)$, solves

$$c = \text{MB}_j^{\text{dem}}\big(s; A_j^{\text{dem+det}}(s)\big) + \text{MB}_j^{\text{det}}\big(s; A_j^{\text{dem+det}}(s)\big).$$

We solve this scalar condition by a one-dimensional search beginning at $A_j^{\text{dem}}(s)$; at each trial point we recompute total marginal benefits and update the advertising level until marginal benefit equals marginal cost.

This two-stage construction ensures that deterrence is evaluated *conditional on* the demand-rational advertising level, and that the final $A_j^{\text{dem+det}}(s)$ equates the full marginal value of advertising—current demand, future demand via ad stock, and reduced probability of future entry—to marginal cost.

Step 3: Residual "other motives"

Define the residual component at state s as

$$\tilde{A}_j(s) \;:=\; A_j^{\text{obs}}(s) \;-\; A_j^{\text{dem+det}}(s).$$

In the model, we summarize this residual with

$$\tilde{A}_j(s) \;\approx\; f(r_j)\,\zeta_t \;+\; \mu_{jt},$$

and characterize it by regressing $\tilde{A}_j(s)$ on firm size r_j, time factors ζ_t, and other observables. Large, systematic $\tilde{A}_j(s)$ indicate advertising that cannot be rationalized by demand returns or by entry deterrence, consistent with brand-positioning, agency, or bargaining motives on the supply side.

B. *Identification*

This section summarizes how the key components of the supply model are identified: (i) the entry hazard, and (ii) the decomposition of advertising motives. Each object is identified using variation that shifts the relevant incentive while holding confounding factors fixed. The core idea is that the entry process is identified from exogenous differences in advertising exposure across media–market borders, and the supply decomposition uses revealed–preference inequalities that rely on the consistently estimated demand and hazard objects.

Entry Hazard

The probability that a new product enters a market is

$$h_{mt} = \Lambda\big(\gamma_0 + \gamma_A \mathcal{A}_{mt} + X_{mt}'\gamma + \psi_m + \tau_t\big)\,,$$

where $\mathcal{A}_{mt}$ summarizes incumbent ad stocks.
The coefficient γ_A captures how advertising shifts the likelihood of entry. The identifying variation is cross–border exposure differences: media markets broadcast across county lines, generating discrete changes in observed advertising that are not correlated with smooth changes in local demand fundamentals.
Identification of γ_A therefore relies on

$$\mathbb{E}\big[\eta_{mt} \mid Z^A_{mt}, X_{mt}, \psi_m, \tau_t\big] = 0, \quad \pi_1 = \frac{\partial \mathcal{A}_{mt}}{\partial Z^A_{mt}} \neq 0,$$

which ensure the instrument shifts incumbent advertising exposure but not the unobserved drivers of entry. This delivers a consistent estimate of $h(s)$ and its sensitivity $\partial h(s)/\partial A$, which are then used in the supply decomposition.

Advertising Motives

The final object of interest is how observed advertising decomposes into three components: (i) demand-driven, (ii) entry–deterrence, and (iii) residual motives $f(r_j)\zeta_t$. Crucially, this step does not require new instruments. Once the demand system and entry hazard are consistently estimated, we recover behavior using revealed–preference inequalities:

- The demand benchmark $A_j^{\text{dem}}(s)$ equates marginal cost with the marginal value of own–demand advertising.
- The deterrence adjustment adds the marginal value of reducing entry probability, using the estimated $\partial h(s)/\partial A$ and continuation-value gap $\Delta V_j(s^+)$.
- The difference between observed advertising and the rational $A_j^{\text{dem+det}}(s)$ is attributed to other motives.

Formally, the core identifying condition is that marginally profitable deviations would have been taken:

$$\mathbb{E}\Big[V_j(s; A_j^{\text{obs}}(s)) \ \geq \ V_j(s; A_j^{\text{obs}}(s) \pm \varepsilon) \mid s\Big],$$

with value functions constructed using the already–identified demand parameters and hazard. Thus, identification of the supply decomposition leverages observed optimality behavior and the consistency of the first–stage estimates rather than new exclusion restrictions.

Together, these strategies identify each piece of the model from distinct variation: taste–driven product substitution for demand, exogenous border exposure differences for entry, and revealed–preference comparisons for supply motives.

REFERENCES

Bagwell, Kyle. 2007. "Chapter 28 The Economic Analysis of Advertising." In . Vol. 3 of *Handbook of Industrial Organization*, , ed. M. Armstrong and R. Porter, 1701–1844. Elsevier.

Bar, Ron N., and Avery Haviv. 2021. "Brand Building to Deter Entry and Its Impact on Brand Value (May 10, 2021)."

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–890.

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov. 2014. "The Economics of Predation: What Drives Pricing When There Is Learning-by-Doing?" *The American Economic Review*, 104(3): 868–897.

Blake, Thomas, Chris Nosko, and Steven Tadelis. 2015. "CONSUMER HETEROGENEITY AND PAID SEARCH EFFECTIVENESS: A LARGE-SCALE FIELD EXPERIMENT." *Econometrica*, 83(1): 155–174.

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

Cockburn, Iain M, and Megan J MacGarvie. 2006. "Entry and Patenting in the Software Industry." National Bureau of Economic Research Working Paper 12563.

Doraszelski, Ulrich, and Sarit Markovich. 2007. "Advertising Dynamics and Competitive Advantage." *The RAND Journal of Economics*, 38(3): 557–592.

Dube, Jean-Pierre H., Guenter J. Hitsch, and Puneet Manchanda. 2005. "An Empirical Model of Advertising Dynamics." *Quantitative Marketing and Economics*, , (3): 107–144.

Ellison, Glenn, and Sara Fisher Ellison. 2007. "Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration." National Bureau of Economic Research Working Paper 13069.

Goolsbee, Austan, and Chad Syverson. 2005. "How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." National Bureau of Economic Research Working Paper 11072.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. 1995*a*. "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments." *Journal of Marketing Research*, 32(2): 125–139.

Lodish, Leonard M., Magid M. Abraham, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. 1995*b*. "A Summary of Fifty-Five In-Market Experimental Estimates of the Long-Term Effect of TV Advertising." *Marketing Science*, 14(3): G133–G140.

McKinsey. 2020. "A New model for the Consumer Goods Industry."

Morton, Fiona Scott, and Florian Zettelmeyer. 2004. "The Strategic Positioning of Store Brands in Retailer-Manufacturer Negotiations." *Review of Industrial Organization*, 24(2): 161–194.

Nevo, Aviv. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307–342.

Petrin, Amil. 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705–729.

Quirk, James P., and Rubin Saposnik. 1962. "Admissibility and Measurable Utility Functions*." *The Review of Economic Studies*, 29(2): 140–146.

Rizzo, John A., and Richard J. Zeckhauser. 1990. "Advertising and Entry: The Case of Physician Services." *Journal of Political Economy*, 98(3): 476–500.

Schmalensee, Richard. 1978. "A Model of Advertising and Product Quality." *Journal of Political Economy*, 86(3): 485–503.

Shapiro, Bradley T., Günter J. Hitsch, and Anna E. Tuchman. 2021*a*. "TV Advertising Effectiveness and Profitability: Generalizable Results From 288 Brands." *Econometrica*, 89(4): 1855–1879.

Shapiro, Bradley T., Günter J. Hitsch, and Anna E. Tuchman. 2021*b*. "TV Advertising Effectiveness and Profitability: Generalizable Results From 288 Brands." *Econometrica*, 89(4): 1855–1879.

Shavell, Steven. 1979. "Risk Sharing and Incentives in the Principal and Agent Relationship." *The Bell Journal of Economics*, 10(1): 55–73.

Sinkinson, Michael, and Amanda Starc. 2018. "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals." *The Review of Economic Studies*, 86(2): 836–881.

Sutton, John. 2007. *Sunk Costs and Market Structure: Price Competition, Advertising, and the Evolution of Concentration.* Vol. 1 of *MIT Press Books*, The MIT Press.

Mathematical Appendix

THEOREM A.1: *If $\pi(A_i, A_{-i}, N)$ is strictly increasing in A_i, and strictly decreasing in A_{-i} and N for any point in the support of π, then T_2 is positive.*

PROOF:
We will prove this in 2 steps. In Step 1, we will show that given $\pi(.,.,N)$ is strictly decreasing in N, then the expression T_2 is positive if increases in a create first order stochastic dominated shifts in the probability $\mathbb{P}(N, a)$ over N. For step 2, I will show that increasing a induces first order stochastic dominated shifts in the probability $\mathbb{P}(N, a)$ over N. Both of these put together imply that T_2 is positive.
Step 1:

$$T_2 := \beta \sum_{N_2 \in \{N_1,..,N_1+M\}} \big[\mathbb{E}\pi(.,.,N_2).\frac{d\mathbb{P}(N_2,a)}{da}\big]$$

Define $\varphi(a)$ as

$$\varphi(a) := \beta \sum_{N_2 \in \{N_1,..,N_1+M\}} \big[\mathbb{E}\pi(.,.,N_2).\mathbb{P}(N_2,a)\big]$$

By this definition, $\frac{d\varphi}{da} = T_2$. Thus, in order to show that T_2 is positive, it is sufficient to show that $\varphi(a)$ increases in a.
Now, we will show that if a induces first order stochastic dominated shifts in the probability $\mathbb{P}(N, a)$ over N, then $\varphi(a)$ is increasing in a. Observe that, by the assumption that $\pi(.,.,N)$ is decreasing in N, we have that the sequence $\{\pi(.,.,N)\}_{N\in\{N_1,..,N_1+M\}}$ is decreasing. Finally, I re-interpret $\varphi(a)$ as an integral over a decreasing function $f(n) := \pi(.,.,n)$ on the discrete space $\{N_1,..,N_1+M\}$, by viewing $\mu(n) = \mathbb{P}(n)$ as a discrete measure on $\{N_1,..,N_1+M\}$.

$$\varphi(a) = \int_{n\in\{N_1,..,N_1+M\}} f(n)\mu(n)$$

Then, it follows from Theorem (1) in Quirk and Saposnik (1962) that $\varphi(a)$ is increasing in a. ■
The intuition here is that increases in a increase the weights $\mathbb{P}(N)$ for smaller N. Since the sequence $\pi(,..,N)$ is decreasing in N, this essentially increases the weighted sum φ.

Step 2: Now, to show that increasing a induces first order stochastic dominated shocks in $\mathbb{P}(N, a)$ over N, I will show that the sign of the derivative $\frac{d\mathbb{P}(N,a)}{da}$ is only negative first and then only positive thereafter.
We use the model to write the expression for the probability $\mathbb{P}(N_1 + K)$ that there are N_1+K products in period 2. This means that K products entered the market, and $M-K$ products did not enter. Since we assumed the shocks ξ for each individual product are independently drawn, we can write down the expression for entry of K firms as a composite event. Let $F(.)$ denote the probability distribution over ξ. Define $e_K := F(\pi(a,.,N_1+K))$,

which is the probability for the event that a product enters among the cohort of K firms in period 2. Then, $\mathbb{P}(N_1 + K)$ is related to e_K as:

$$\mathbb{P}(N_1 + K, a) = \binom{M}{K} e_K^K (1 - e_K)^{M-K}$$

Then, the derivative of $\mathbb{P}(N_1 + K, a)$ is

$$\frac{d\mathbb{P}(N_1 + K, a)}{da} = \binom{M}{K} e_K^{K-1} (1 - e_K)^{M-K-1} (K - Me_K) \frac{de_K}{da}$$

By our assumption that $\pi(a, ., N_1 + K)$ is strictly decreasing in a, $\frac{de_K}{da} < 0$. Thus, the sign of $\frac{d\mathbb{P}(N_1+K,a)}{da}$ is determined by the sign of $(K - Me_K)$.
Its easy to see that for $K = 0$, $(K - Me_K) < 0$ and for $K = M$, $(K - Me_K) > 0$. Now we need to show that this sign changes only once, and then we would have proved Step 2.
Observe that $\{e_1 > e_2 >, .. > e_M\}$ since we assumed that $\pi(a, ., N_1 + K)$ is decreasing in $N_1 + K$.
Suppose that the series $(K - Me_K)$ changes sign more than once. We know that it is negative first and positive eventually, so let's pick the smallest integers $\{K_1, K_1 + L\}$ for which the sign flips from negative to positive and then from positive to negative respectively. This implies:

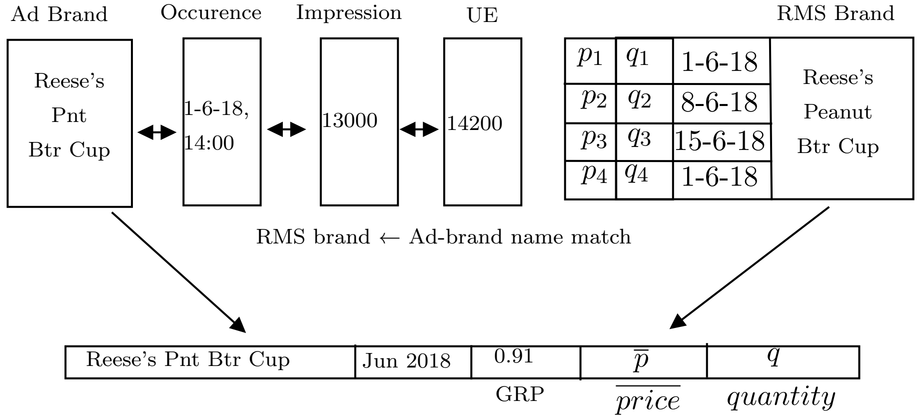
$$\begin{aligned} K_1 - Me_{K_1} &> 0 \\ Me_{K_1+L} - K_1 - L &> 0 \end{aligned}$$

Adding these inequalities, we have

$$e_{K_1+L} - e_{K_1} > L/M$$

But we know that $e_{K_1+L} - e_{K_1}$ is negative, so this cannot hold. Thus, we have a contradiction. ■

Details of data merge



Unfortunately, there are no viable identifiers to link products in the RMS panel data and the advertisements data. I use fuzzy string matching to create a crosswalk between ad-brands and RMS brands.

Since there are a large number of advertising brands, a fuzzy string matching exercise which looks for matches for RMS-brands among all the ad-brands runs into dimensionality issues and becomes infeasible. In order to solve this problem, I subset the set of ad-brands to only include advertisement product categories with marketcodes which are topically related to chocolate brands. I employ a bag-of-words logic, where a topic (chocolates) is identified with a bag-of-words which are statistically most likely to describe chocolate brands in the RMS data. I select the most frequent 15 words that occur in RMS brands. Using this word list, I identify the marketcodes among advertising products. I use the set of products with these marketcodes as the master list of ad-brands, and for each RMS-brand, I create and save the Jaro–Winkler distance to the top 10 RMS-brands for each ad-brand. Finally, I manually check these potential matches to identify which RMS-brands matched successfully to particular ad-brands. At the end of this procedure, I am able to match more than 90% of ad-brand names for our selected (advertising) marketcodes to a specific RMS brand.

The advertisement data also has advertising under a general brandname, e.g., "Hershey's chocolates". I allow advertisements with this general ad-brand to be related to all RMS brands which carry the brand label ("Hershey's", in this case). Through this process, I match all general ad-brands to multiple related RMS brands. This procedure creates a comprehensive match between advertisements and sales for almost all brands in this industry. Furthermore, since the AdIntel data has identifiers for advertising parents, I create the ownership structure in this market through this match process which will be crucial for the supply model.

Robustness checks

Exclude state border variation

Stores which lie in different states might have different demand, or more plausibly, supply decisions arising from state regulations. Since DMAs are often a sub-grid of States, across DMA variation could be confounded by unobservable differences in sales due to across-state variation. I carry out a robustness check where I reduce my sub-sample of stores to stores which are on DMA boundaries, excluding stores that lie on state boundaries. This selects for a specific subset of stores which are more likely in state centers as opposed to peripheries. In the presence of heterogeneous treatment effects, one can expect different elasticity magnitudes for this sub-sample ex-ante, but should be hopeful of the same signs from our estimates if underlying mechanisms act the same way.

	$log(y_i)$	
	(1)	(2)
$log(p_i)$	-0.3097 **	-0.2938**
	(0.0090)	(0.0024)
$log(A_{brand(i)})$	0.0013	0.0009
	(0.0011)	(0.0011)
$log(A_i)$	0.0020**	0.0028**
	(0.0011)	(0.0009)
$log(A_{brand(i)}) \times \mathbb{1}(new)$		0.0069**
		(0.0017)
$log(A_i) \times \mathbb{1}(new)$		-0.0050**
		(0.0022)
$log(\mathbf{p_{-i}})$	0.9656**	0.9032**
	(0.0339)	(0.0291)
$log(A_{-i})$	-0.1946**	-0.1653**
	(0.0543)	(0.0553)
$log(A_{-i}) \times \mathbb{1}(new)$		0.0108**
		(0.0011)
N	58,477,230	58,477,230
Fixed-effects		
Brand	Yes	Yes
Store	Yes	Yes
Month of year	Yes	Yes

Table C1—: Table of log(quantity)- log(ads) estimates for stores on a DMA border but not on a state border. (2) has effects of advertising separately identified for new brands (< 2 years old). Standard errors are clustered at the Store×Year level.

I indeed estimate similar effects of ads- own ads have small, positive and significant effect on own sales while other total ads have meaningful negative elasticities for own sales.

Alternate functional form

I test the robustness of my results with the choice of a different functional form assumption. I estimate a linear-linear model for sales and advertisements, with the same set of controls. I find similar signs of the effects.

	(y_i)	
	(1)	(2)
p_i	-72.041 **	-0.2938**
	(34.4408)	(0.0024)
$A_{brand(i)}$	-0.0783**	0.0009
	(0.0383)	(0.0011)
A_i	0.2681**	0.0028**
	(0.1095)	(0.0009)
$A_{brand(i)} \times \mathbb{1}(new)$		0.0069**
		(0.0017)
$A_i \times \mathbb{1}(new)$		-0.0050**
		(0.0022)
$\mathbf{p_{-i}}$	59.13**	0.9032**
	(3.352)	(0.0291)
A_{-i}	-0.0204**	-0.1653**
	(0.0064)	(0.0553)
$A_{-i} \times \mathbb{1}(new)$		0.0108**
		(0.0011)
N	103,095,487	103,095,487
Fixed-effects		
Brand	Yes	Yes
Store	Yes	Yes
Month of year	Yes	Yes

Table C2—: Table of quantity- ads estimates for stores on a DMA border. (2) has effects of advertising separately identified for new brands (< 2 years old). Standard errors are clustered at the Store×Year level.

While I cannot compare magnitudes of these estimates to my main specification, the relative importance of own ads and total ads flips. Own ad effects are still positive and significant and total ad effects are negative and significant. However, the effect of additional total advertising in the linear model is less in magnitude compared to own-ad effects. One way to explain this is through a heterogeneous effects framework, where log-models weight effects of smaller-level advertising more. Alternatively, it is instructive to note that total other advertising is overall very large compared to own-advertising for this industry. Given this, log models represent percentage changes in own-advertising which in GRP terms are smaller changes for many firms than total other advertising (the two differ in means by a $x30$ scale). Nevertheless, markets do see large changes in total advertising which are estimated to have an economically meaningful effect on product sales in this specification

as well.
I also check robustness against spurious correlation arising from the construction of Ad-brand. Given that ad-brands are cumulative sums of GRPs, for some values of δ they can be correlated with increasing sales trends by construction. In my data, I have shown that industry level revenues show no trends while my preferred Ad-brand construction ($\delta = 0.7$) shows statistically insignificant negative trend. Nevertheless, it is possible that individual product level sales over time could have this spurious correlation. I estimate my main specification by only using the GRPs in current period as the independent variable. [4] This yields similar results, alleviating concerns of spurious correlation.

	$log(y_i)$	
	(1)	(2)
$log(p_i)$	-0.3187 **	-0.283**
	(0.0069)	(0.0015)
$log(A_{brand(i)})$	0.0007	0.0002
	(0.0009)	(0.009)
$log(A_i)$	0.0047**	0.0049**
	(0.0009)	(0.0009)
$log(A_{brand(i)}) \times \mathbb{1}(new)$		0.0006**
		(0.0001)
$log(A_i) \times \mathbb{1}(new)$		-0.0060**
		(0.0027)
$log(\mathbf{p_{-i}})$	0.9435**	0.8689**
	(0.0295)	(0.0309)
$log(A_{-i})$	-0.0018**	-0.0021**
	(0.0021)	(0.0023)
$log(A_{-i}) \times \mathbb{1}(new)$		0.0124**
		(0.0011)
N	103,095,456	103,095,456
Fixed-effects		
Brand	Yes	Yes
Store	Yes	Yes
Month of year	Yes	Yes

Table C3—: Table of log(quantity)- log(ads) estimates for stores on a DMA border using no accumulation of ad-capital. (2) has effects of advertising separately identified for new brands (< 2 years old). Standard errors are clustered at the Store×Year level.

[4] This also corresponds to a static ad-process with a $\delta = 0$

Maps: GRPs, residuals and store restrictions

I use geographic information to subset stores which lie on DMA borders. To do this, I download shapefiles of US counties. Using the *SF* (Simple features) package in *R*, I identify the counties whose borders overlap with DMA borders. This creates the sub-sample of firms I use in my main specification. I am including a plot of US counties which are color coded- red indicates counties that are excluded and blue counties are the border counties I end up using as my geographic sub-sample.

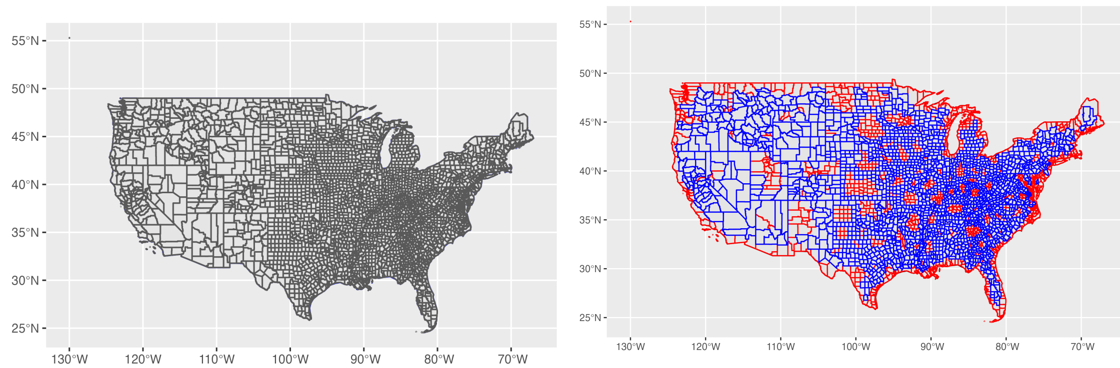


Figure D1. : 1) Raw map of US counties (2) Red counties are excluded, blue are included in main specification

Furthermore, I remove the counties which are on state borders for one of my robustness checks described in the previous section. I also plot the counties by their inclusion status below.

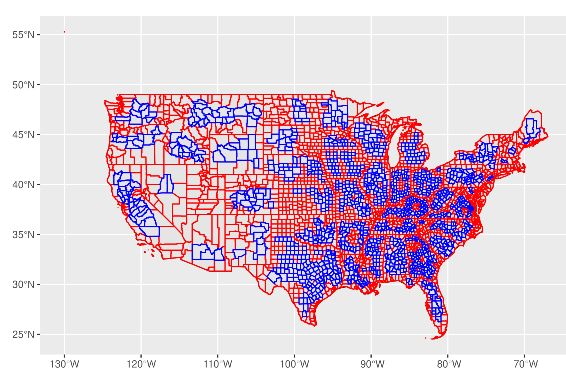


Figure D2. : Red counties are excluded, blue included in robustness check without stores on state borders

Simulation for entry deterrence

As a proof of concept, I simulate a market with otherwise identical firms where deterrence occurs in equilibrium. I create a market which has logit shares coming from identical firm value. There is no price in this simple simulation, only advertising is the firms choice which generates difference in shares. Finally, there is a fixed cost which determines the maximum number of firms that can exist in a steady state equilibrium. Below E1 is a plot of this maximum firm count by fixed cost.

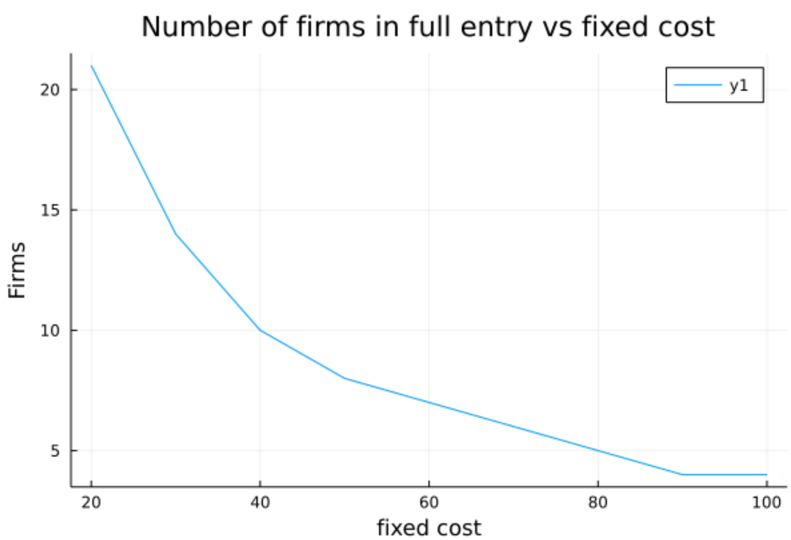


Figure E1. : Max firm count N^* by fixed cost. Note that in each equilibrium, firms select a profit maximising identical level of advertising endogenously.

Now, as a minimal example of deterrence, I find a level of (identical) advertising by $N^* - 1$ incumbent firms so that an N^*th prospective firm has zero lifetime profit. If the N^*th firm were to enter, I plot below in E2 the subsequent ad-capital and generated market share paths in a dynamic equilibrium for these firms. Both of these eventually converge to steady state levels.

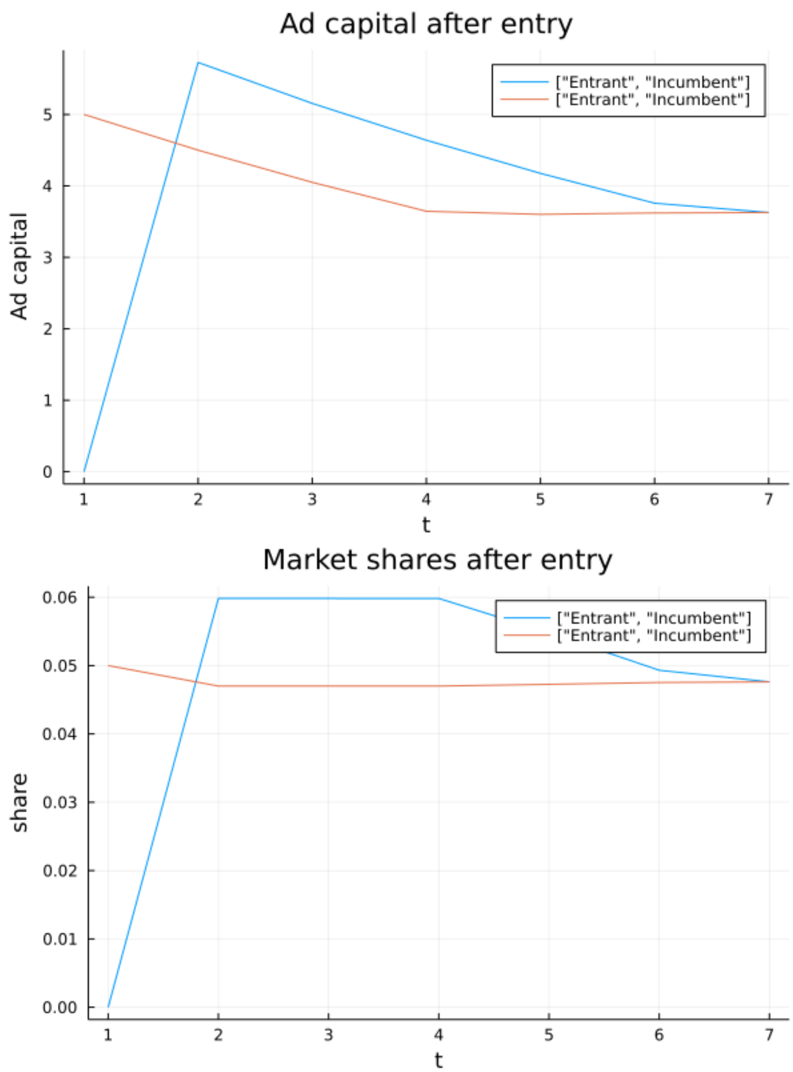


Figure E2. : 1) Ad-capital path (2) Shares, for incumbents in orange and entrant in blue.

Main takeaways here are that by keeping the level of advertising slightly higher than the steady state level (Y intercept of the orange line), incumbent firms can actually keep entrants out. In this particular case, I chose the profits and fixed costs such that it is actually profitable for $N^* - 1$ firms to deter entry of the N^*th firm.

Broader entry trends in CPG goods industries

Decrease in entry counts is a broader pattern than extends beyond the particular industry that I focus on in this paper. I look at the count of new UPCs that show up year after year, and compare them to the counts in 2007. This count has decreased substantial for all broad categories of dry groceries found in the Nielsen data.

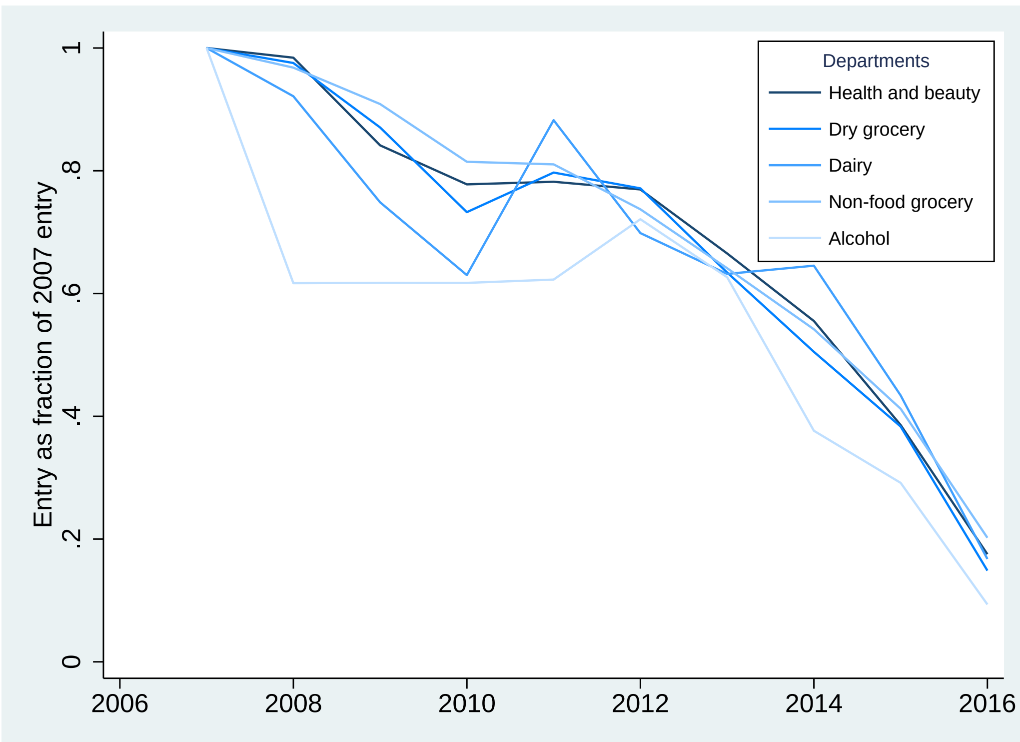


Figure F1. : Number of new UPCs entered as a fraction of 2007 count

Furthermore, I find similar patterns of staggered entry over time for products in all of these markets. Firms enter markets gradually and then reduce their presence from many stores. For each year after a product enters the data, I calculate it's presence as a share of its maximum reach after entry. I average this for all entrants within broad product categories to arrive at the following figure:

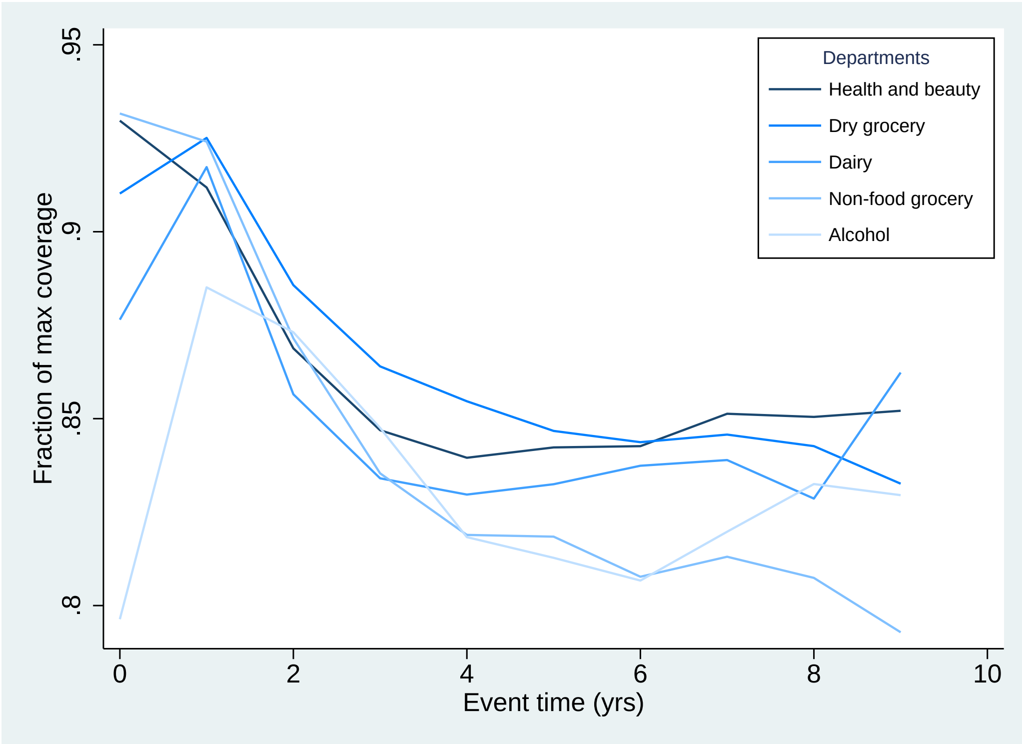


Figure F2. : Products enter stores gradually over time