

NBER WORKING PAPER SERIES

GENERALIZABLE AND ROBUST TV ADVERTISING EFFECTS

Bradley Shapiro
Günter J. Hitsch
Anna Tuchman

Working Paper 27684
<http://www.nber.org/papers/w27684>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2020

All three authors contributed equally although not listed in alphabetical order. We acknowledge the superb research assistance of Jihong Song and Ningyin Xu. We thank Liran Einav, Paul Ellickson, Jeremy Fox, Wes Hartmann, Carl Mela, Matt Shum, and Sha Yang for helpful comments. We also benefited from the comments of seminar participants at Amazon, Bates White, Columbia, CUHK, HKUST, Johns Hopkins, NUS, Rice, UNC, UCSD, Yale, Marketing Science, the MSI Young Scholars Conference, the Wash U. Junior Faculty Development Forum, the NYC Media Seminar, the (IO)² Zoom Seminar and the 12th Workshop on the Economics of Advertising and Marketing. Calculated (or derived) 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 those of the researchers and do not reflect the views of Nielsen, nor of the National Bureau of Economic Research. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2020 by Bradley Shapiro, Günter J. Hitsch, and Anna Tuchman. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Generalizable and Robust TV Advertising Effects
Bradley Shapiro, Günter J. Hitsch, and Anna Tuchman
NBER Working Paper No. 27684
August 2020
JEL No. B41,C18,C52,C55,C81,L00,L15,L81,M31,M37

ABSTRACT

We provide generalizable and robust results on the causal sales effect of TV advertising for a large number of products in many categories. Such generalizable results provide a prior distribution that can improve the advertising decisions made by firms and the analysis and recommendations of policy makers. To provide generalizable results, we base our analysis on a large number of products and clearly lay out the research protocol used to select the products. We characterize the distribution of all estimates, irrespective of sign, size, or statistical significance. To ensure generalizability, we document the robustness of the estimates. First, we examine the sensitivity of the results to the assumptions made when constructing the data used in estimation. Second, we document whether the estimated effects are sensitive to the identification strategies that we use to claim causality based on observational data. Our results reveal substantially smaller advertising elasticities compared to the results documented in the extant literature, as well as a sizable percentage of statistically insignificant or negative estimates. Finally, we conduct an analysis of return on investment (ROI). While our results show that many brands perform better with their observed advertising than they would without advertising, we document considerable over-investment in advertising at the margin.

Bradley Shapiro
Booth School of Business
University of Chicago
5807 South Woodlawn Avenue
Chicago, IL 60637
and NBER
Bradley.Shapiro@chicagobooth.edu

Anna Tuchman
Northwestern University
Kellogg School of Management
2211 Campus Dr
Evanston, IL 60208
anna.tuchman@kellogg.northwestern.edu

Günter J. Hitsch
University of Chicago
Booth School of Business
5807 South Woodlawn Avenue
Chicago, IL 60637
guenter.hitsch@ChicagoBooth.edu

A data appendix is available at <http://www.nber.org/data-appendix/w27684>
A Online interactive results is available at <https://advertising-effects.chicagobooth.edu/>

1 Introduction

We study the causal effect of television advertising on sales, with a focus on the generalizability of the results across products in different categories. Further, we document the robustness of the results to different approaches to construct the data and to different empirical strategies to obtain causal advertising effect estimates.

Generalizable knowledge on the causal effect of advertising is of importance both to the economic analysis of advertising and to business analytics. First, an important literature in economics has investigated the effect of advertising on market structure, competition, and concentration. Sutton’s (1991) endogenous sunk cost theory of market structure and concentration, applied to the case when advertising creates vertical product differentiation, assumes that advertising affects consumer demand. The degree to which advertising is effective and has a long-run impact on demand or “brand equity” (Borkovsky et al. 2017) determines if entry deterrence is possible (Ellison and Ellison 2011, Bar and Haviv 2019). Second, a long line of research in economics investigates if advertising is primarily persuasive, informative, or effectively a complement to product consumption (see the survey by Bagwell 2007). This study of the mechanism by which advertising affects demand would be moot if advertising were ineffective. Third, in traditional broadcast markets content and advertising are bundled. As such, advertising acts as a price paid by consumers for content, and the long-run viability of this business model depends on the effectiveness of advertising (the estimates in Brynjolfsson et al. 2019 imply that consumers would be willing to pay \$135 billion per year to avoid losing TV access). The size of the industry also speaks to its economic importance. Total U.S. advertising spending in 2019 was \$256 billion, with \$66 billion spent on TV advertising.¹

Furthermore, from a normative point of view, a key task for business and marketing analytics is to predict the profitability or return on investment (ROI) from incremental advertising spending in order to facilitate good advertising planning decisions.

To provide economists and industry practitioners with a general understanding of the effectiveness of TV advertising, our study must ensure that the results are not driven by idiosyncratic examples. Generalizable results ensure the external validity of the findings and provide a prior distribution for decision-making. In the case of advertising, a prior distribution of the advertising elasticity among similar products allows a firm or policy maker to assess a likely range of advertising ROIs even without conducting its own analysis. Once specific advertising elasticity estimates are obtained, for example using an internal analysis conducted by the firm’s data science team or using an external analysis by a marketing consulting firm, the prior serves as a benchmark to assess the credibility of these estimates. An unbiased prior is important not only to firms, but also to researchers. One of the basic goals of science is to create knowledge that can be replicated and aggregated to form a broad understanding of what we have learned. Such an aggregation of what we know informs subsequent research questions, methodological advances

¹Winterberry Group: *The Outlook for Data Driven Advertising & Marketing 2020*

and the search for new and better data.

In this paper, we provide a generalizable distribution of television advertising elasticities and estimates of advertising return on investment (ROI). The analysis is based on a sample of 288 consumer packaged goods (CPG) that are selected using a clear research protocol.² We estimate the advertising elasticities and ROIs brand by brand, carefully controlling for confounding factors. We report all results, irrespective of size, sign, or statistical significance. Because the data are available for researchers through the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business, the analysis can be replicated and the sample selection process can be verified.

We obtain generalizable results using this research process because it circumvents publication bias. Most advertising research in empirical industrial organization and marketing has used a case study approach. Generalizations are obtained either by using a systematic meta-analysis or through a more informal summary of published results. However, if published results are not representative due to publication bias, a meta-analysis will not yield an estimate of the true population distribution of effects. If our knowledge-base is biased, firms could make sub-optimal advertising investments and research could proceed in sub-optimal ways.

Robustness of the results to the specific assumptions and choices made in the analysis is—in addition to the research protocol used—an important component to obtain generalizable results.

To ensure robustness, we first provide a detailed discussion of the approach and assumptions made to construct the final data, in particular the data on the intended advertising exposure level, from the raw data sources. This part of our work should be of interest to other researchers or analysts who use the Nielsen Ad Intel data as a source of advertising occurrence and exposure measures.

We intend to provide generalizable results on the *causal* effect of advertising on sales. In general, advertising is not randomly assigned, and thus, in the presence of unmeasured confounders, the estimated advertising effects do not have a causal interpretation. We employ two identification strategies that rely on the specific institutions of the ad-buying process. First, in the *baseline specification* we posit that advertisers can target demand differences across markets or seasons and also aggregate, time-varying demand shocks. The geographic and seasonal demand differences in particular are relatively easy to predict. The identifying assumption is that the residual variation in advertising is driven by slot availability and cost factors. We view this assumption as plausible given the institutional features of the ad-buying and scheduling process. This identifying assumption will be violated, however, if some advertising is targeted in a more sophisticated way to local, transient demand factors. Hence, we also use a second approach, the *border strategy*, which exploits the institutional constraint that local media markets (DMAs)

²We focus on CPG products because CPG firms like Procter and Gamble are among the largest advertisers on TV and because demand data on multiple products in other categories are not readily available. Based on estimates from the U.S. Bureau of Economic Analysis, approximately 10% of personal consumption expenditures are on CPG products.

are the smallest geographic unit at which ad-buys can be made. Assuming that consumers who live close to a DMA border are similar, this constraint is the source of plausibly quasi-random variation in advertising at a DMA border (Shapiro 2018).

In addition, we conduct extensive sensitivity analyses to evaluate alternative modeling assumptions. As it would be impractical to include all sensitivity checks in the paper, we created an interactive web application that allows the reader to explore all model specifications. The web application is available at <https://advertising-effects.chicagobooth.edu>.

We estimate advertising stock elasticities, which are a form of long-run elasticity that represents the percentage change in current period sales volume resulting from a one-percent increase in current and past advertising. We find that the mean and median of the distribution of estimated long-run own-advertising elasticities are 0.025 and 0.014, respectively, and more than two thirds of the elasticity estimates are not statistically different from zero. The main results are robust, both to the main identification strategies and to the exact specifications within each identification strategy. In particular, the fixed effects and border discontinuity strategies yield similar estimates for the brands in our sample and virtually indistinguishable advertising effect distributions. The results are robust to controls for own and competitor prices or feature and display advertising, and the advertising effect distributions are similar whether a carryover parameter is assumed or estimated. The estimates are also robust to functional form and do not appear to be driven by measurement error.

A key finding of this work is that the magnitudes of the estimated advertising elasticities are considerably smaller compared to the results in the extant literature. This is consistent with both publication bias and over-estimated elasticities in the literature due to confounding factors.

The advertising elasticity estimates are not only small in comparison to results in the prior literature, but also imply a negative return on investment (ROI) at the margin. In particular, the ROI on advertising in a given week is -79.0% for the median brand and negative for more than two-thirds of the brands in our sample.³ Hence, the majority of brands over-invest in advertising. This result does not imply that all advertising is wasted: For many brands, the observed level of advertising is more profitable than not advertising at all.

This paper highlights the need for generalizable results and proposes a multi-product research design that allows us to study the fundamental questions of ad effectiveness and ad profitability with a wide-angle lens. We first discuss how our work relates to the existing literature in Section 2. Section 3 introduces the research design to estimate causal advertising effects. Section 4 describes the data used in the empirical analysis, and Section 5 provides summary statistics and documents key variation in the data. In Section 6 we present the estimation results. Section 7 discusses the economic implications for profits and ROIs, and Section 8 concludes.

³These results are obtained using plausible assumptions on retail and manufacturer gross margins.

2 Literature review

Our work is closely related to a set of papers that perform meta-analyses of published advertising elasticities with the objective of drawing generalizable conclusions about advertising effectiveness. While a variety of measures can be used to quantify advertising effectiveness, elasticities are especially useful when drawing comparisons across studies and product categories because the metric is unit-less. Assmus et al. (1984) analyzes 128 advertising elasticity estimates reported in 22 studies published between 1962 and 1981. The average short-run elasticity is 0.22 with a standard deviation of 0.26. In a more recent follow-up study, Sethuraman et al. (2011) augments the sample used by Assmus et al. (1984) with additional studies of advertising effectiveness that were published between 1981 and 2008. They report a mean short-run advertising elasticity of 0.12 and a mean long-run elasticity of 0.24, concluding that advertising effectiveness has decreased over time.⁴ Similarly, the meta-analysis by Henningsen et al. (2011) documents mean short and long-run elasticities of 0.09 and 0.19, respectively.⁵

This type of work has two main limitations. First, it relies on published estimates of advertising effectiveness. Second, differences in the analytic approach may create spurious differences across studies of ad effectiveness. In other words, the conclusions drawn from a meta-analysis are only as strong as the quality and comparability of the underlying data and models. In our study, we overcome this limitation by using a single source of data and the same model across estimated TV ad elasticities.

Most closely related to our study is the seminal work by Lodish et al. (1995), which summarizes advertising elasticity estimates for 141 brands. The estimates are based on matched household-level advertising exposure and purchase data from IRI’s BehaviorScan household panel. In the BehaviorScan markets, IRI conducted split-cable experiments in which advertising treatments were randomized across households. For the 89 established products in the tests, conducted between 1982 and 1988, Lodish et al. (1995) documents an average advertising elasticity of 0.05 and reports that 33% of the elasticity estimates are statistically different from zero using a one-tailed test at a 20% level.⁶ These results provide a relevant comparison to our work, because (i) the Lodish et al. (1995) results were almost certainly not selected based on size, sign or statistical significance, (ii) robustness is ensured given the split-cable RCT design, and (iii) the population of consumer packaged goods is likely similar to our population.⁷

⁴Sethuraman et al. (2011) defines a “long-term” ad elasticity as the percentage change in a brand’s current and future period sales for a 1% change in the brand’s current advertising. Some of the studies they analyze report long-term elasticities directly. Others report the short-term elasticity and a “carryover” coefficient on an included lagged dependent variable, which they convert into a long-term elasticity by computing (short-term elasticity / (1 - carryover coefficient)) (Clarke (1976)). As we discuss in Section 3.1 and Appendix A, the long-term elasticities reported in Sethuraman et al. (2011) are comparable to our estimated advertising stock elasticity.

⁵Henningsen et al. (2011) provide a database of published elasticity estimates.

⁶Lodish et al. (1995) compute elasticities by dividing the estimated percentage change in brand volume over the test year by the percentage change in GRP weight applied.

⁷In a follow-up paper, Hu et al. (2007) analyze the results of 241 TV advertising tests carried out between 1989 and 2003. The authors document larger effects compared to Lodish et al. (1995), primarily due to the effect sizes in tests conducted after 1995. We focus our discussion on Lodish et al. (1995) because we cannot rule out

We aim to build on the work of Lodish et al. (1995) in several ways. First, the BehaviorScan test markets are no longer in use and cannot be used for advertising measurement today. In contrast, our work evaluates television advertising effects using currently available data and methods that are widely employed in the industry and by researchers. Second, although not reported by Lodish et al. (1995), the power of the tests was likely low. (Abraham and Lodish (1990) reports a total of about 3,000 households in the BehaviorScan markets, and thus 3,000 is the maximum sample size in each test.) Compared to Lodish et al. (1995), our study covers a longer time series and many more markets, through which we obtain better statistical power and greater external validity.

Our work is also complementary to some cross-category studies that relate television advertising to various outcomes using observational data. For example, Clark et al. (2009) and Du et al. (2018) examine the relationship between advertising and survey measures of consumer brand awareness and attitudes. Deng and Mela (2018) studies the effects of micro-targeting using a model that jointly estimates the utility from television viewing with a purchase utility model. Our work builds on these studies by focusing on the effect of advertising on sales as an outcome.

Our work also relates to some recent multi-product studies of online advertising (Goldfarb and Tucker (2011), Johnson et al. (2016), Kalyanam et al. (2018)). Just as these studies help us assess the generalizability of online ad effects, our analysis extends our understanding of the full distribution of TV ad effects. Further, while these studies may suffer from selection bias stemming from willingness to conduct experiments, our study has no such selection.

Also related to this study of advertising is the work on pricing and promotion strategies in retail grocery channels by Hitsch et al. (2019), which provides generalizable results by estimating store-level price elasticities for the top 2,000 CPG brands (based on sales revenue).

3 Research design

3.1 Basic model structure

Our goal is to measure the effect of advertising on sales. For each product or brand, we specify a constant elasticity model with advertising carryover. The basic model structure, not including fixed effects and other covariates that we will introduce below, is:

$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \mathbf{A}_{d(s)t}) + \boldsymbol{\alpha}^T \log(\mathbf{p}_{st}) + \epsilon_{st}. \quad (1)$$

Q_{st} is the quantity (measured in equivalent units) of the product sold in store s in week t . $\mathbf{A}_{d(s)t}$ is a vector of own and competitor advertising stocks, also referred to as goodwill, in DMA $d(s)$

that the results in Hu et al. (2007) are affected by selection, in particular if advertisers who saw small advertising effects in the earlier split-cable tests stopped testing. We are less concerned about such a form of selection in the 1982-1988 tests, because these tests were conducted immediately after the introduction of the BehaviorScan test markets and before the overall small degree of advertising effectiveness was publicized in the Harvard Business Review (Abraham and Lodish 1990) and in the academic work by Lodish et al. (1995).

in week t . \mathbf{p}_{st} is a corresponding vector of own and competitor prices. We specify the advertising stock (goodwill) as:

$$\mathbf{A}_{d(s)t} = \sum_{\tau=t-L}^t \delta^{t-\tau} \mathbf{a}_{d(s)\tau}. \quad (2)$$

$\mathbf{a}_{d(s)t}$, also a vector, is the flow of own and competitor advertising in DMA $d(s)$ in week t , and δ is the advertising carryover factor. L indicates the number of lags or past periods in which advertising has an impact on current demand. In our empirical specification we set $L = 52$. We assume that $\mathbf{A}_{d(s)t}$ captures all dynamics associated with advertising, including the standard carryover effect (current advertising causes future purchases) and structural state dependence (current purchases caused by current advertising cause future purchases). Variation in current advertising that affects future sales will be captured via the distributed lag structure in (2), regardless of the specific mechanism.

We measure own advertising using two *separate* variables. The first own advertising variable captures advertising messages that are specific to the focal product. Such advertising is likely to have a non-negative effect on sales.⁸ The second own advertising variable captures advertising messages for affiliated products that, *ex ante*, could have either a positive effect through brand-spillovers or a negative effect through business stealing. For example, an increase in advertising for Coca-Cola soft drinks could increase demand for regular Coca-Cola, but it could also decrease demand for regular Coca-Cola if sufficiently many consumers substitute to Coke Zero or Diet Coke. We will discuss the corresponding data construction approach more thoroughly in Section 4. We also include advertising and prices for up to three competitors in the model.⁹

As the demand function is specified as a log-log model, $\boldsymbol{\alpha}$ includes the own and cross-price elasticities of demand. The coefficients in $\boldsymbol{\beta}$ have an approximate elasticity interpretation. For simplicity, we drop the store and market indices and focus on one component of \mathbf{A}_t , i.e., the advertising stock for one specific product. The corresponding advertising stock elasticity is given by

$$\frac{\partial Q_t}{\partial A_t} \frac{A_t}{Q_t} = \beta \frac{A_t}{1 + A_t} \approx \beta. \quad (3)$$

Thus, β is an approximation of the advertising stock elasticity. As shown in Appendix A, the advertising stock elasticity captures the long-run effect of a change in advertising on demand, and in particular measures the percentage change in demand resulting from an increase in current and past advertising by one percent. See Appendix A for alternative interpretations of the advertising stock elasticity as a long-run advertising effect and its relationship to the short-run advertising elasticity.

⁸It is possible to construct models, such as the consideration set model in Sahni (2016), where an increase in own advertising can *reduce* own demand.

⁹The competing brands are selected based on size (total revenue). Not all brands are sold at all stores. Hence, if a competing brand that is included in the model is not sold at a store, all observations for that store need to be excluded from the analysis. Therefore, for each brand we determine the number of competitors that are included in the model based on the percentage of observations that would be lost if we added one additional competitor.

As our particular default specification may work better for some applications than others, we provide robustness to the functional form specification, both allowing δ to vary by brand and allowing the shape of the ad response curve to vary by brand.

Demand model choice and functional form

To obtain generalizable and robust advertising effect estimates we need to scale the computations to obtain the estimates across a large number of brands and a large number of different model specifications. The log-linear demand model makes these computations feasible (parallelized across many nodes in a state-of-the-art computing cluster it takes approximately one month to obtain all results).

Absent these computational constraints, we would ideally estimate the relationship between advertising and sales using a micro-founded, structural demand model, such as Berry et al. (1995). Indeed, for the purpose of specific policy evaluations, such as the effect of a merger on equilibrium prices (e.g., Nevo 2000) and advertising levels, or to assess the welfare effect of new product introductions (e.g., Petrin 2002), a structural demand model would be indispensable. The main goal of this paper, however, is to document the distribution of the overall effectiveness of TV advertising across many brands, and we do not conduct policy evaluations that require a prediction of the change in equilibrium advertising due to, for example, a merger or the entry of a new competitor in a market. We consider our demand specification as a log-linear approximation to a micro-founded, structural demand model. To assess the robustness of our results to the specific functional form, we also present flexible semi-parametric estimates that are regularized using the Lasso. Our main results are unchanged by the additional flexibility.

3.2 Identification strategies

The main challenge when estimating model (1) is that advertising is not randomly assigned. Firms may target their advertising in DMAs and periods when they believe that advertising will be most effective. Correspondingly, firms may advertise more in markets and periods where consumers are positively disposed towards the product even in the absence of advertising. There may also be unobserved and hence omitted factors that are correlated with both advertising and sales. In the presence of such confounding factors, the statistical relationship between advertising and sales does not have a causal interpretation. Hence, to ensure that we estimate the causal effect of advertising on sales, we need a plausibly random source of variation in advertising.

3.2.1 Institutions of the ad-buying process

Our identification strategies utilize the particulars of the ad-buying process. Television ads are purchased through negotiations between advertisers (or ad agencies) and television stations. As much as 80% of advertising is purchased well in advance of the ad being aired in an “upfront”

market.¹⁰ In addition to being purchased in advance, there is considerable bulk buying. That is, an agency will buy a large quantity of advertising to be divided between many clients in exchange for discounts from the stations. The remaining advertising inventory is sold throughout the year. The so-called “scatter market” allows for last minute purchases of individual ads, typically sold at higher rates than upfronts (Hristakeva and Mortimer 2020). Additionally, local networks sometimes sell unsold remnant advertising space to the national networks or other aggregators, which bundle the ads and sell them to advertisers at a discount.

These institutions of the ad-buying process make precise targeting difficult. In the upfront market, advertisers may target demand based on differences across local markets, seasonal factors, and trends in demand that can be predicted in advance. Advertisers may also attempt to target demand based on more concrete information about local demand factors that becomes available over time. However, as the majority of inventory is sold up front, ad buys in the scatter and bundled remnant markets close to a target date are constrained by slot availability. Hence, if ad slots in a given week are unavailable in some local markets, the advertiser may buy air time in a previous or subsequent week or not buy additional ad slots in these local markets at all. Even if ad slots are available, the cost of advertising may differ across local markets. In particular, in some local markets advertising inventory may be available in the relatively cheap bundled remnant market, whereas in other markets ad slots may only be available in the relatively expensive scatter market. Further, when purchasing in the bundled remnant market, an advertiser may incidentally purchase an ad slot that was of little interest due to the fact that it was bundled with a more desired ad slot. Because of these cost differences, ad buys may occur in the relatively cheap but not in the expensive markets.

Variation in advertising over time or across markets may also occur due to technical or scheduling factors that may cause ads to get displaced from their originally planned slots. For example, a sporting event may go on longer or shorter than originally planned, altering the planned schedule for ads both during and after the event.

These institutions of the ad-buying process suggest two plausible sources of quasi-random variation in advertising: (i) variation in the exact timing of advertising in a local market, and (ii) variation in the amount of advertising across local markets within a time period. Among our two main identification strategies, the *baseline specification* relies entirely on (i) and (ii). The *border strategy* utilizes an additional source of quasi-random variation in advertising, the institutional constraint that advertising is bought at the DMA level and hence cannot be targeted to a smaller geography within a DMA.

3.2.2 Baseline specification

The baseline specification includes various fixed effects and controls:

$$\log(Q_{st}) = \beta^T \log(1 + \mathbf{A}_{d(s)t}) + \alpha^T \log(\mathbf{p}_{st}) + \gamma_s + \gamma_{\mathcal{S}(t)} + \gamma_{\mathcal{T}(t)} + \boldsymbol{\eta}^T \mathbf{x}_{st} + \epsilon_{st}. \quad (4)$$

¹⁰<https://digiday.com/marketing/upfrontses-wtf-upfronts/>

γ_s is a store fixed effect that subsumes persistent regional differences in demand, $\gamma_{\mathcal{S}(t)}$ is a week-of-year fixed effect that captures seasonal effects, and $\gamma_{\mathcal{T}(t)}$ captures aggregate changes or trends in demand. In our preferred specification, $\gamma_{\mathcal{T}(t)}$ is a month fixed effect corresponding to week t , but we also estimate specifications with quarter or week fixed effects and a specification where $\gamma_{\mathcal{T}(t)}$ represents a linear time trend. \mathbf{x}_{st} is a vector of other controls at the store-week level, including feature and display advertising in some of the model specifications.¹¹ Also, recall that $\mathbf{A}_{d(s)t}$ and \mathbf{p}_{st} are vectors that include the advertising stocks and prices of competing brands, s , which may be correlated with the focal brand’s advertising activity.

Advertisers are likely to choose different levels of advertising across markets and seasons based on systematic, predictable differences in demand. The baseline specification incorporates store and season fixed effects to adjust for the corresponding confounds. Advertisers may also predict more idiosyncratic fluctuations or trends in demand at the aggregate, national level. We adjust for the resulting confounds using the time fixed effects or trend. Furthermore, advertising campaigns may be coordinated with national or local feature and display advertising, and thus we include data on displays and features in some of the model specifications.

The baseline specification allows for targeting of national demand fluctuations, but assumes that advertisers do not engage in more sophisticated targeting of transient demand shocks at the local market level. Under these assumptions, the residual variation in advertising, conditional on all fixed effects and controls, can plausibly be considered quasi-random, given the institutions of the ad-buying process discussed above in Section 3.2.1 that result in variation in the exact timing of advertising within and variation in the amount of advertising across local markets due to costs and slot availability.¹²

3.2.3 Border strategy

The baseline specification assumes that advertisers can target aggregate, national demand fluctuations. However, advertisers may also target demand shocks that occur at the local market level using ad buys in the scatter market and the bundled remnant market. If such micro-targeting occurs and is systematically related to local, transient demand shocks, the time fixed effects or trends in the baseline specification will not be sufficient to yield a causal advertising effect. To address this challenge, we use the border strategy, a research design introduced by Shapiro (2018).¹³

The border strategy focuses on consumers who live on different sides and close to DMA

¹¹Feature and display advertising is only recorded for 17% of all stores in the data. Hence, our preferred specifications omit these variables. See (<https://advertising-effects.chicagobooth.edu/>) for the results that include feature and display advertising.

¹²The baseline specification results presented in this paper use monthly time fixed effects, even though the data are at the weekly level. We also estimated specifications using week and quarter fixed effects. Utilizing week fixed effects considerably decreases the statistical power for many brands, due to the large share of national advertising, and it also makes the week-of-year dummies redundant.

¹³Shapiro (2018) studies the effect of television advertising on antidepressant demand. The border strategy has also been used in Tuchman (2019) to study e-cigarette advertising, as well as in Spenkuch and Toniatti (2018) and Wang et al. (2018) to study political advertising.

borders. The central assumption is that these consumers are similar, in the sense that consumers on different sides of a DMA border exhibit identical trends or transient fluctuations in their purchasing behavior. Essentially, this is an assumption of parallel trends.¹⁴ However, while the consumers close to the DMA border are assumed to exhibit identical trends in their purchasing behavior, the overall transient demand fluctuations in each of the two bordering DMAs may be different.

The border strategy utilizes the institutional constraint that DMAs are the smallest geographic unit at which ad-buys can be made. Hence, advertising that is targeted to local, DMA-specific demand shocks will be based on the overall, population-weighted *average* demand shocks at the DMA level. Consequently, if the average demand shocks across two bordering DMAs differ, then advertising across the DMAs will differ, too. In particular, the level of advertising at the border will generally be different from the optimal, equilibrium level of advertising that firms would set if they could micro-target more locally within a DMA.

The institutional constraint on ad-buying and the assumption of identical trends at the DMA border imply that temporal differences in advertising across the bordering DMAs are not confounded with temporal differences in border-specific demand shocks. In particular, if we generalize the baseline strategy and allow for border-specific time fixed effects, the residual variation in advertising, conditional on all fixed effects and controls, will be quasi-random.

Additional identifying variation in advertising is provided by the institutions of the ad-buying process that lead to variation in the incidence and timing of advertising across local markets, as discussed in Section 3.2.1. The resulting quasi-random variation in advertising provides robustness, in particular in the extreme case where the average demand shocks across two DMAs were identical in every period, resulting in *identical optimal advertising levels* on each side of the DMA border. In this extreme case, absent the variation that is due to the ad-buying process institutions, the residual variation in advertising at the border would be zero.

The border strategy also requires the stable unit treatment value assumption (SUTVA) that consumers who are exposed to advertising do not cross the border and purchase in the neighboring DMA.¹⁵ If this assumption is violated, the advertising effect estimate will be biased towards zero. This assumption has empirical support based on Tuchman (2019), who analyzes e-cigarette purchases made by Nielsen Homescan panelists and only finds a 3% incidence of cross-DMA transactions.

The implementation of the border strategy has two components. First, we restrict our sample to the set of stores that are located in counties that share a border with a county located in a different DMA. In total, there are 183 borders between the 123 DMAs in the contiguous United States where we observe each of the major television networks in the data. Figure 1 shows the location of the borders and border counties on the U.S. map. Second, we adapt the baseline

¹⁴Shapiro (2018) and Tuchman (2019) discuss the parallel trends assumption in the border strategy in detail and provide evidence that the assumption is likely to hold in their empirical applications.

¹⁵This assumption also needs to hold in the baseline specification, although it is less relevant when the sample is not restricted to stores close to the border.

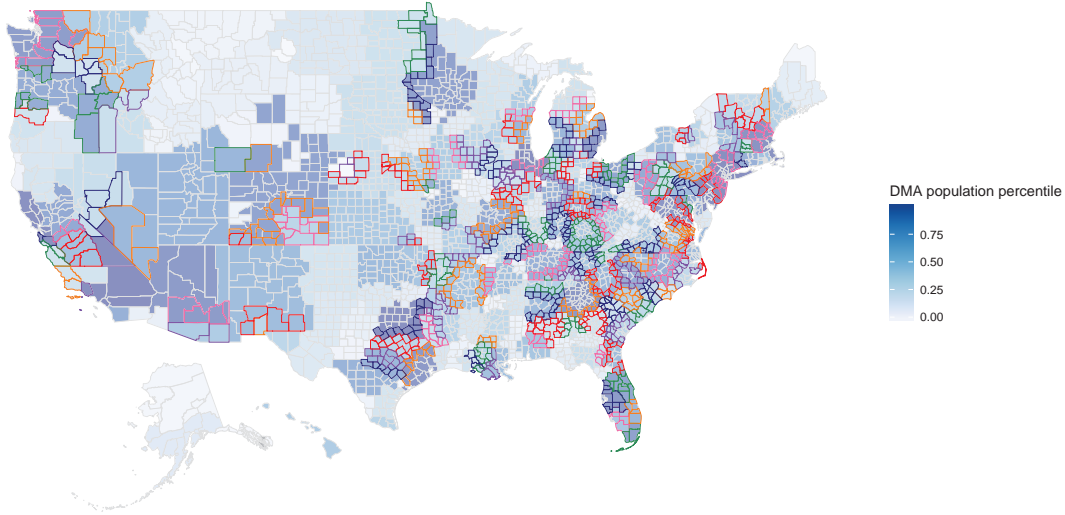


Figure 1: Border Counties in the United States

specification (model (4)) to include border-specific time fixed effects, $\gamma_{\mathcal{B}(s,t)}$:

$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \mathbf{A}_{d(s)t}) + \boldsymbol{\alpha}^T \log(\mathbf{p}_{st}) + \gamma_s + \gamma_{\mathcal{S}(t)} + \gamma_{\mathcal{B}(s,t)} + \boldsymbol{\eta}^T \mathbf{x}_{st} + \epsilon_{st}. \quad (5)$$

Our preferred specification uses border-month fixed effects, but we also estimate specifications using border-quarter and border-week fixed effects. We consider these different specifications because the unobservables may be spatially and temporally correlated in different ways, and we want to explore the robustness of our results to alternative assumptions about these correlations. We report the alternative specifications in our interactive online appendix (<https://advertising-effects.chicagobooth.edu/>).¹⁶ As in the baseline specification, the store fixed effects γ_s account for persistent regional differences in demand, and the week-of-year fixed effects $\gamma_{\mathcal{S}(t)}$ capture seasonality.

3.2.4 Comparison of baseline specification and border strategy

Both the baseline specification and the border strategy utilize the quasi-random variation in advertising due to the institutions of the ad-buying process. In addition, the border strategy utilizes the institutional constraint that micro-targeting within a DMA is not possible as a key source of quasi-random advertising variation across two neighboring DMAs. The border

¹⁶This appendix allows the user to add and subtract control variables, to change the main specification, to alter the fixed effects and to restrict the sample in various ways. For example, the appendix shows the distribution of estimates for the border strategy implemented using border-week fixed effects rather than border-month fixed effect. The user can also choose to restrict the sample to only those brands that have positive and significant effects, or to the subset of brands with 50% ex ante power to detect a 0.05 advertising elasticity. In this way, the reader may transparently observe the sensitivity of the distribution to a very large number of alternative specifications. Please see the appendix for instructions.

strategy provides robustness against the possibility that there is local targeting of DMA-level demand shocks. The variation in advertising due to the institutional ad-buying process provides robustness against the possibility that there is no cross-border variation in advertising that is driven by demand targeting.

If advertising effects are geographically homogeneous, a comparison of the results of the baseline and border approaches is informative as to the prevalence of local targeting of DMA-level demand shocks. In particular, if advertisers target local transient demand shocks, the baseline strategy will be biased.

If the baseline strategy is assumed to be valid, the comparison of the two specifications is instructive about potential geographic heterogeneity in advertising effects.¹⁷ Any differences between estimates in the baseline and border strategies can then be interpreted as differences in advertising effectiveness in the country as a whole and in the DMA border regions, in particular.

3.2.5 Standard error calculations

When estimating the regression models, standard errors are clustered to account for correlation in the error terms. The clustering varies by specification as different specifications induce different forms of residual variation in advertising, which induces different correlation structures across the error terms. In the baseline specification, where the monthly time fixed effects correspond to a different time interval than the weekly data, we two-way cluster the standard errors by DMA and week. This accounts for (1) the serial correlation in error terms that can arise when analyzing panel data with repeated observations over time and (2) the correlation in error terms induced by correlation in the advertising treatment across markets. In particular, since time fixed effects are at the month level, there may be correlation within month and between weeks induced by the fact that every market receives the same amount of national advertising. In the border strategy specification, we two-way cluster standard errors by border-side and by week.

3.2.6 Statistical power

Using only observations from the border counties significantly decreases the sample size. However, the net effect on statistical power is ambiguous. To see this, note that the border strategy affects statistical power in three ways. First, each border-specific time fixed effect is an additional control variable. This reduces residual variance in the dependent variable, which, all else equal, increases statistical power.¹⁸ However, the additional fixed effects also reduce the residual variance in the advertising stock, which all else equal, reduces statistical power. Finally, focusing on the border counties reduces the sample size, which reduces statistical power. Hence, the net

¹⁷As a large majority of advertising is purchased in advance and micro targeting based on the residual inventory is costly, we believe the assumptions underlying the baseline strategy are reasonable.

¹⁸For example, demand for lotion during winter may increase more in the Northeast than in the South. The border-specific time fixed effects are able to explain these differential trends, while common time fixed effects cannot.

effect of the border strategy on statistical power depends on the relative magnitudes of these three factors.

3.3 Other identification strategies

Other papers have proposed identification strategies that use instrumental variables to estimate a causal effect of advertising on sales. For example, Gordon and Hartmann (2013) uses market-level advertising prices as instruments, and Sinkinson and Starc (2019) proposes to use the timing of political campaigns as an exogenous shifter of brand advertising. Thomas (2020) and Li et al. (2019) propose instrumental variables approaches that take advantage of the fact that advertising decisions are made at a more aggregate level than the measurement of demand in the data. Finally, Shashoua et al. (2018) analyzes household panel data and proposes using a model with correlated random coefficients and lagged purchase and advertising observations as instruments to estimate causal advertising effects.

We chose not to implement these strategies because instruments are case-specific and, hence, impractical for a study that estimates advertising effects for 288 brands. Furthermore, whether the instruments are sufficiently strong to avoid weak instruments bias will also vary across applications.¹⁹

4 Data

To estimate the effect of advertising on sales we use data on purchase volumes, advertising intensities, and other components of marketing, in particular prices. We construct a data set by merging market (DMA) level TV advertising data with retail sales and price data at the brand level. The data and our matching procedure are described in more detail below. Our study is the first to provide generalizable and comprehensive results on the effectiveness of TV advertising using the wealth of information in the Nielsen Ad Intel and RMS scanner data. Merging these two large data sets is difficult. The brand identifiers in the Ad Intel data do not match up perfectly with the brand descriptions in the RMS data. Often the advertised brand name is either more or less specific than the brand name associated with a UPC code. Another complication is that the advertising data come from a number of measurement devices at the local and national level that must be reconciled in order to produce a coherent television timeline. Appendix B shows, in detail, how to re-create our data construction process.

4.1 RMS retail scanner 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.

¹⁹See, for example, Moshary et al. (2020), which shows that the strength of the first stage varies tremendously across product categories when using political advertising as an instrument for product advertising.

Despite covering a large number of stores and retailers, the data available for research from the Kilts Center for Marketing constitute only a non-random subset of all retail chains in the U.S. Typically, the data cover more than 50% of all market-level spending in grocery and drug stores and one-third of all spending at mass merchandisers.²⁰

The sample used in our analysis includes data from 2010 to 2014. We focus our analysis on the top 500 brands in terms of dollar sales. These brands account for 45.3% of the total observed RMS revenue, even though there are more than 300,000 brands in the data.²¹ We define a brand as all forms of the same consumable end product, as indicated by the brand code or brand name in the RMS data. That is, Coca-Cola Classic includes any UPC that was composed entirely of Coca-Cola Classic, including twelve ounce cans, two-liter bottles, or otherwise. Because advertising is generally at the brand level, rather than the UPC level, we aggregate across UPCs, calculating total volume sold in equivalent units and average price per equivalent unit. After dropping some smaller stores and stores that are located in counties that switch DMAs over time, we are left with 12,671 stores in the final estimation sample.

The price of a UPC is only recorded in weeks when at least one unit of the UPC was sold. To impute these prices that are missing from the data, we follow the approach detailed in Hitsch et al. (2019). This approach uses an algorithm to infer the base price, i.e., regular, non-promoted shelf price of a product, and assumes that weeks with zero sales occur in the absence of a promotion, such that the unobserved price corresponds to the base price.

4.2 Homescan household panel data

The policy experiments and ROI calculations in Section 7 make use of the Nielsen Homescan household panel data as an additional source of purchase information. The Homescan data capture household-level transactions, including purchase quantities and prices paid. Data for more than 60,000 households are available each year. Nielsen provides weights called “projection factors” for each household. Using these weights, transactions can be aggregated across all households to be representative at the national level, i.e., estimate a product’s total purchase volume for in-home use. We utilize these estimates of total sales for the policy experiments and ROI calculations because the RMS data do not capture all transactions and would hence underestimate the incremental value of advertising.

4.3 Advertising data

Product-level television advertising data for 2010–2014 come from the Nielsen Ad Intel database. The advertising information is recorded at the occurrence level, where an occurrence is the placement of an ad for a specific brand on a given channel, in a specific market, at a given day and time. Four different TV media types are covered in the data: Cable, Network, Syndicated,

²⁰More details are provided in the *Retail Scanner Data set Manual* that is available from the Kilts Center for Marketing.

²¹See Hitsch et al. (2019) for a detailed analysis of the relationship between cumulative revenue and UPC rank.

and Spot. Occurrences for each of these different media types can be matched with viewership data, which then yields an estimate of the number of impressions, or eyeballs, that viewed each ad. In the top 25 DMAs, impressions are measured by set-top box recording devices. For all other DMAs, impressions are measured using diaries filled out by Nielsen households. These diary data are only recorded in the four “sweeps months,” February, May, July, and November. We impute the impressions for all other months using a weighted average of the recorded impressions in the two closest sweeps months.

For Cable ads, which are aired nationally, viewership data are available only at the national level. Spot ads are bought locally, and viewership measures are recorded locally, separately for each DMA. Network and Syndicated ads are recorded in national occurrence files that can be matched with local measures of viewership in each DMA. Thus, in our data, variation in a brand’s aggregate ad viewership across markets is due to both variation in occurrences across markets (more Spot ads were aired in market A than in market B) and variation in impressions (eyeballs) across markets (a Network or Syndicated ad aired in both markets A and B, but more people saw the ad in market A than in market B).

Using the occurrence and impressions data, we calculate gross rating points (GRPs), a widely used measure of advertising exposure or intensity in the industry. We first calculate the GRP for a specific ad occurrence, defined as the number of impressions for the ad as a percentage of all TV-viewing households in a DMA (measured on a scale from 0 to 100). To obtain the aggregate, weekly GRPs in a given DMA, we obtain the sum of all occurrence-level GRPs for a brand in a given week in the DMA.

4.4 Matching advertising and retail sales data

We merge the advertising and sales data sets at the store-brand-week level. Our merging procedure warrants some discussion because the brand variables in the Ad Intel and RMS data sets are not always specified at the same level. For example, we have to decide if an advertisement for “Coca-Cola” should only be joined to the sales data of regular Coca-Cola or also to Diet Coke. We include three types of advertising variables in our models. First, we include advertising that directly corresponds to the RMS product in question. Second, we create a variable that captures advertising for “affiliated” brands, including potential substitutes, that may affect the focal RMS product. Third, we include advertising for the top competitor. For example, for the Diet Coke brand, own advertising includes ads for Diet Coke, whereas affiliated advertising includes advertising for Coca-Cola soft drinks, Coke Zero, Coca-Cola Classic, and Cherry Coke. Furthermore, we include advertising for Diet Pepsi, the top competitor of Diet Coke.

We separately estimate the effect of own brand and affiliated brand advertising because own brand advertising is likely to have a positive effect on sales, whereas the sign of the effect of affiliated brands’ advertising is ambiguous. Hence, lumping own and affiliated brand advertising together might result in small and uninterpretable elasticity estimates.

Full details of the matching approach are provided in Appendix B and the estimated affiliated

brand and competitor ad effects are reported in Appendix C.

5 Data description

5.1 Brand-level summary statistics

Using the process described in Section 4.4 and Appendix B, we were able to match 288 of the top 500 brands in the RMS data to TV advertising records in the Ad Intel database. These products are typically established products, and hence the results from our empirical analysis need not apply to new products.²²

Table 1 provides brand-level summary statistics. Total yearly revenue is larger when based on the spending records in the Homescan data compared to the measured revenue in the RMS retail sales data. This is expected, because the RMS data only contain information on a subset of all retail chains, and furthermore, the reported RMS revenue is calculated using the subset of stores used in our estimation sample. The Homescan revenue, on the other hand, is predicted using the transaction records and household projection factors supplied in the Nielsen data, and is thus designed to be representative of total national spending.

The data reveal a large degree of heterogeneity in advertising spending. Total yearly TV advertising spending for the median brand is 10.5 million dollars, with a 90% range of 2.2 to 61.3 million dollars. Table 1 and Figure 2 also document the variation across brands in average weekly GRPs at the DMA level, calculated as $\bar{a}_j = \frac{1}{M \cdot T} \sum_{m=1}^M \sum_{t=1}^T a_{jmt}$, where M is the number of DMAs and T the number of weeks. The median of this measure of advertising activity is 35.5, with a 90% range from 4.7 to 184.8. A similar degree of cross-brand heterogeneity is evident in the advertising/sales ratio, with a median of 2.8 and a 90% range from 0.5 to 17.8.²³

5.2 Temporal and cross-sectional variation at the brand level

The degree of temporal and cross-sectional variation in brand-level advertising is of particular relevance for the goal of estimating advertising effects on demand. The corresponding empirical analysis relies on variation in advertising levels and ad stocks both across markets and over time. We document the extent of this variation in the data. First, separately for each brand,

²²The results from the split-cable experiments in Lodish et al. (1995) indicate that new-product advertising is typically much more effective than advertising for established products.

²³The Dorfman-Steiner theorem says that an optimizing firm should set its advertising-to-sales ratio equal to the product of its gross-margin percentage times its advertising elasticity of demand (Dorfman and Steiner (1954)). Thus, we can use the observed advertising-to-sales ratios and average industry margins to calibrate expected elasticities *assuming firms are advertising optimally*. Because brand-specific gross-margins cannot be inferred from our data, we utilize the following median gross margin percentages that are sourced from a technical report prepared by the Grocery Manufacturers Association, the Food Products Association, and PricewaterhouseCoopers: 34% for food companies, 44% for beverage companies, and 50% for companies selling household goods and personal care products (). Plugging in the median advertising-to-sales ratio and the median gross margin percentage for different classes of CPG products, *if firms are advertising optimally*, then the advertising elasticity should be about $\varepsilon_A = 2.8\%/34\% = 0.08$ for food companies, $\varepsilon_A = 2.8\%/44\% = 0.06$ for beverage companies, and $\varepsilon_A = 2.8\%/50\% = 0.06$ for companies selling household goods and personal care products. These calculations provide a benchmark against which we compare our estimates in Section 6.

Table 1: Brand Level Summary Statistics

	Median	Mean	Percentiles							
			1%	5%	10%	25%	75%	90%	95%	99%
RMS revenue	113.1	170.8	28.8	51.2	61	75.4	190	332.4	522.4	698.9
Homescan revenue	341	546.3	74.3	119.7	150.8	220	647.7	1046.3	1544.6	3190.4
Advertising spending	10.5	18.6	0.9	2.2	3.6	5.9	22.2	44.7	61.3	106.7
Mean weekly GRPs	35.5	59.4	2.2	4.7	8.4	19	71.8	149.7	184.8	333.8
Adv./sales ratio (%)	2.8	5.3	0.2	0.5	0.8	1.4	5.6	12.4	17.8	33.6
% of Adv. Spending										
<i>Cable</i>	50.9	52.8	10	20.9	29.9	38.4	65.6	81	88.2	99.8
<i>Network</i>	34.5	34.1	0.9	4.2	8.6	19.6	47.5	56.6	66.6	82.4
<i>Spot</i>	3.7	8.7	0.1	0.4	0.7	1.5	8.6	20.9	37.8	82.5
<i>Syndicated</i>	5.3	6.6	0	0	0	1.6	9.8	16.6	19.1	22.9

Note: The sample includes 288 brands. Revenue and advertising spending are expressed in millions of dollars. The advertising/sales ratio is calculated using Homescan revenue.

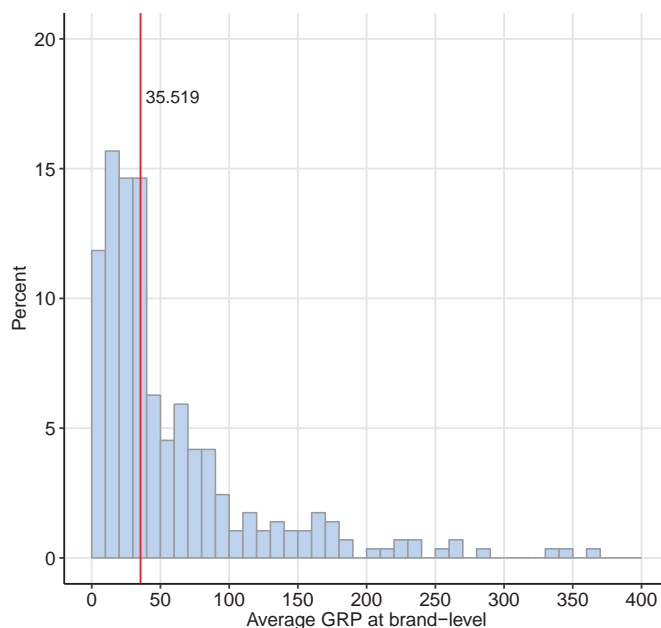


Figure 2: Variation in Advertising Intensity Across Brands

Note: Mean advertising is calculated as the average of GRPs across all DMAs and weeks, separately for each brand.

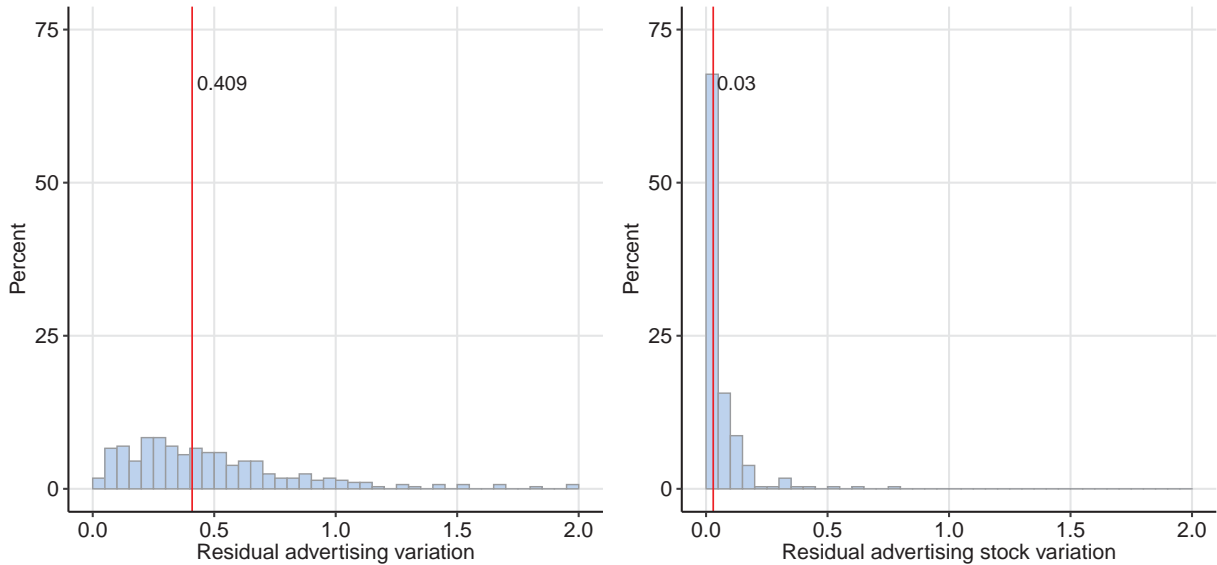


Figure 3: Residual Variation in Advertising

Note: The residual variation measures are based on the residuals from a regression of advertising or advertising stock ($\delta = 0.9$) on DMA, time (month), and seasonal (week-of-year) fixed effects, own and competitor prices, and competitor advertising. The residual variation is the ratio of the standard deviation of these residuals relative to the mean advertising or advertising stock. The measure is calculated separately for each brand, and these graphs show the distribution across brands.

we regress weekly DMA-level advertising, measured in GRPs, on a set of DMA, week-of-year (season), and month fixed effects. Additional covariates included in this regression are own and competitor prices, and competitor advertising. We then calculate the ratio of the residual standard deviation from this regression relative to average DMA/week advertising. This measure is similar to a coefficient of variation and serves as a parsimonious way of quantifying the degree of variation in advertising that is not explained by the fixed effects and the other covariates.

Figure 3 presents a histogram of the measure across brands. Figure 3 also shows an analogous measure of the residual variation in advertising *stock* relative to the average DMA/week advertising stock.²⁴ The advertising stock is calculated assuming a carryover parameter of $\delta = 0.9$. The “coefficient of variation” of advertising flows is 0.41 for the median brand. I.e., for the median brand, the standard deviation of the residuals is 41% of average weekly advertising. Hence, we observe relatively large deviations from average advertising levels for most brands. For advertising stocks, on the other hand, the relative residual variation is substantially smaller. For the median brand, the “coefficient of variation” in advertising stocks is 0.03. This variation is smaller than the corresponding variation in advertising levels because the advertising stock is a weighted average of the advertising flows, and thus smooths out the variability in advertising levels. Due to the small residual variation in advertising stocks it may be difficult to accurately measure the advertising stock effect on demand.

²⁴In the regression used to obtain the residuals, advertising flows are replaced with advertising stocks.

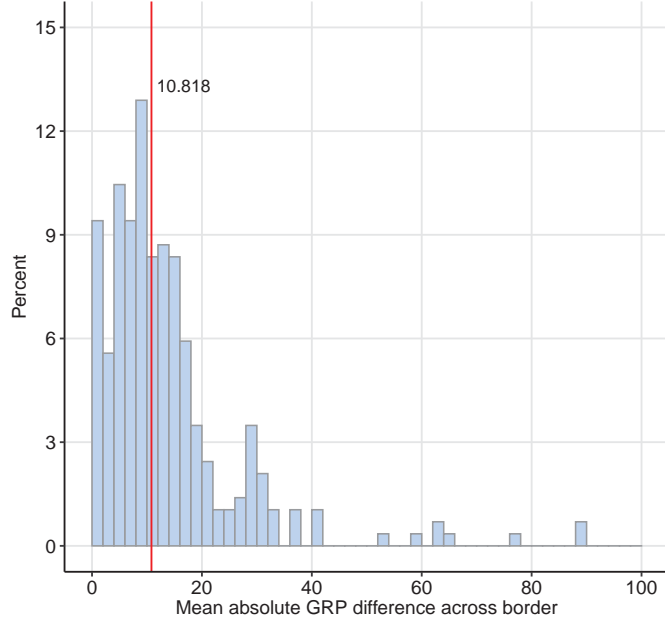


Figure 4: Weekly Absolute GRP Difference Across Borders

Note: The graph shows the distribution of brand-level mean absolute GRP differences across borders.

The border strategy discussed in Section 3.2 utilizes cross-border variation in advertising in a given period and thus imposes additional requirements on the advertising data generating process. In particular, the regressions now include border/time fixed effects, and there needs to be a sufficient degree of residual variation in the ad flows and ad stocks, net of these fixed effects and other covariates, to estimate the effect of advertising on demand. We provide two analyses to examine if such variation is present in the data. First, in Figure 4, we show the distribution of brand-level average absolute differences in GRPs across borders, where the average is taken over all border-week observations in the data.²⁵ The average absolute difference is about 14 GRPs, which appears reasonably large relative to the average weekly GRPs documented in Table 1. Second, in Figure 5, we report results on the net residual variation in advertising flows and advertising stocks corresponding to the border strategy. These results are analogous to the results in Figure 3, but the residuals are obtained from regressions where the time fixed effects are replaced by border-month fixed effects. Notably, the net variation in residuals from the border-strategy model is similar to the net variation in residuals from the more parsimonious baseline model in Figure 3.

²⁵ $\overline{\Delta a_j} = \frac{1}{B \cdot T} \sum_{b=1}^B \sum_{t=1}^T |a_{j d_1 t} - a_{j d_2 t}|$, where $d_1 = d_1(b)$ and $d_2 = d_2(b)$ are indices for the two DMAs on each side of border b .

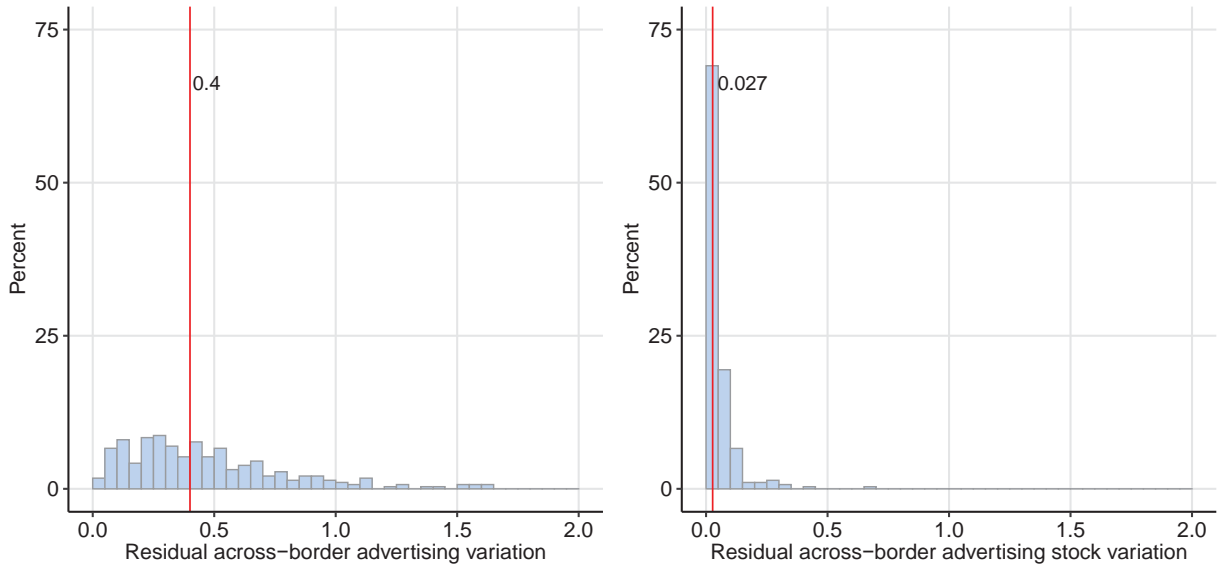


Figure 5: Residual Variation in Advertising: Border Strategy

Note: The residual variation measures are based on the residuals from a regression of advertising or advertising stock ($\delta = 0.9$) on border-month and seasonal (week-of-year) fixed effects, own and competitor prices, and competitor advertising. The residual variation is the ratio of the standard deviation of these residuals relative to mean advertising or advertising stock. The measure is calculated separately for each brand, and these graphs show the distribution across brands.

6 Results

We first present the results of the two main specifications discussed in Section 3.2, the baseline specification (model (4)) and the border strategy (model (5)). Recall that both models include store and week-of-year (season) fixed effects, but the baseline model includes common month fixed effects, while the border strategy includes border-specific month fixed effects. The estimation results are initially obtained assuming a carryover parameter $\delta = 0.9$, which is similar to other specifications in the literature.²⁶

In Section 6.2, we examine the robustness of the results to the exact specification and identification strategy, and show when the estimates “stabilize,” in the sense that the advertising effect estimates do not change significantly when we include additional covariates in the model. Specifically, we start with a *naive* specification, which contains no fixed effects, and we show how the distribution of the estimated advertising effects changes as we incrementally add controls for different potential confounders.²⁷ We also discuss the robustness of the results to calibrating and estimating the carryover parameter, δ .

The results discussed in this section are only a subset of all the models we have estimated. Please see <https://advertising-effects.chicagobooth.edu/> to explore the sensitivity of the results

²⁶For example, Dubé et al. (2005) estimate an advertising decay parameter of $\delta = 0.9$ using data on weekly ad GRPs for brands in the frozen entree category.

²⁷The naive model includes own price, competitor price, and advertising as covariates.

to alternative modeling choices.

6.1 Main results

We present the estimation results for the own-advertising stock elasticities, i.e., the coefficients corresponding to the focal brand in the vector β .²⁸ As we discussed in Section 3.1, the advertising stock elasticities can be interpreted as a form of long-run advertising elasticity. For the sake of brevity, from now on we will refer to the own-advertising stock elasticities as advertising elasticities or advertising effects.

Summary statistics for the naive, baseline, and border strategy model estimates are provided in Table 2, and the full distributions are displayed in Figure 6. The left panel in Figure 6 shows the histogram of advertising elasticities from the baseline specification with store, month and week-of-year (season) fixed effects. The right panel displays the results when we employ the border strategy.

Based on the naive model with no fixed effects, the median long-run advertising elasticity across brands is 0.0299, and the mean is 0.0415. 19.4% of estimates are negative and statistically significant, and 38.9% are not statistically different from zero. Using the baseline specification, which adjusts for potential confounders using a rich set of fixed effects, the median shrinks considerably to 0.0140, and the mean shrinks to 0.0233. Negative and significant results are reduced to 7.3% of the estimates, and 66.3% of the estimates are not statistically distinguishable from zero. The results using the border strategy are similar to the results from the baseline specification. The median of the estimates is 0.0136, and the mean is 0.0258. Furthermore, 7.3% of the estimates are negative and statistically significant, and 68.4% of the estimates are not statistically distinguishable from zero. Figure 6 shows that the distributions of advertising elasticities from the baseline specification and border strategy are nearly identical.

Discussion

We draw a few key take-aways from the analysis in this section. First, the mean of the estimated advertising elasticities is small. In particular, it is substantially smaller than the mean long-run elasticity of 0.24 reported in Sethuraman et al. (2011)'s meta-analysis of advertising effects and roughly half the size of the most comparable prior study, Lodish et al. (1995), which used data from the 1980s.²⁹ Second, the estimated advertising effects are not statistically different from zero for two-thirds of the brands in our sample. In the following section, we explore whether the relatively small effect magnitudes and high frequency of null effects are driven by a lack of statistical power, measurement error or specific modeling decisions.

²⁸Appendix C discusses the estimated affiliated brand and cross-advertising elasticities.

²⁹Furthermore, comparing our results to the predictions from the Dorman-Steiner formula reported in footnote 23, the mean estimated elasticity is also roughly half the size of the elasticities we would expect to see if firms were advertising optimally.

Table 2: Own-Advertising Stock Elasticity Estimates

	Median	Mean	% p \geq 0.05	% p < 0.05		Percentiles			
				> 0	\leq 0	10%	25%	75%	90%
Main Results									
Naive	0.0299	0.0415	38.89	41.67	19.44	-0.0713	-0.0178	0.0833	0.1827
+ Store FE	0.0218	0.0467	33.68	50.69	15.62	-0.0349	-0.0051	0.0682	0.1494
+ Season FE	0.0152	0.0251	28.82	51.04	20.14	-0.0454	-0.0107	0.0534	0.1051
+ Time trend	0.0110	0.0171	41.67	42.36	15.97	-0.0360	-0.0053	0.0381	0.0772
Baseline specification	0.0140	0.0233	66.32	26.39	7.29	-0.0406	-0.0082	0.0426	0.0919
Border strategy	0.0136	0.0258	68.40	24.31	7.29	-0.0321	-0.0055	0.0472	0.1015
Robustness									
Prices excluded	0.0158	0.0275	73.26	21.53	5.21	-0.0498	-0.0133	0.0597	0.1396
+ Own price	0.0128	0.0243	68.40	25.35	6.25	-0.0341	-0.0070	0.0418	0.0992
+ Top 1 competitors	0.0124	0.0253	69.10	24.31	6.60	-0.0312	-0.0078	0.0411	0.0993
+ Up to top 2 competitors	0.0134	0.0254	68.75	24.31	6.94	-0.0301	-0.0067	0.0415	0.0996
Feature & display included	0.0107	0.0215	73.96	19.79	6.25	-0.0335	-0.0069	0.0369	0.0943
50% power to detect 0.05	0.0073	0.0083	69.43	21.02	9.55	-0.0215	-0.0047	0.0214	0.0367
Lasso	0.0181	0.0313	-	-	-	-0.0550	-0.0159	0.0700	0.1253

Note: Descriptive statistics of estimated advertising elasticities reported for 288 brands. The naive model includes own and competitor advertising stocks and prices but no additional controls. The baseline model includes store, week-of-year (season), and month fixed effects. The border strategy restricts the sample to stores in border counties and includes store, week-of-year (season), and border-month fixed effects. All robustness results are obtained using the border strategy. Standard errors are two-way clustered at the DMA and week level in the naive and baseline specifications, and two-way clustered at the border-side and week level in the border strategy.

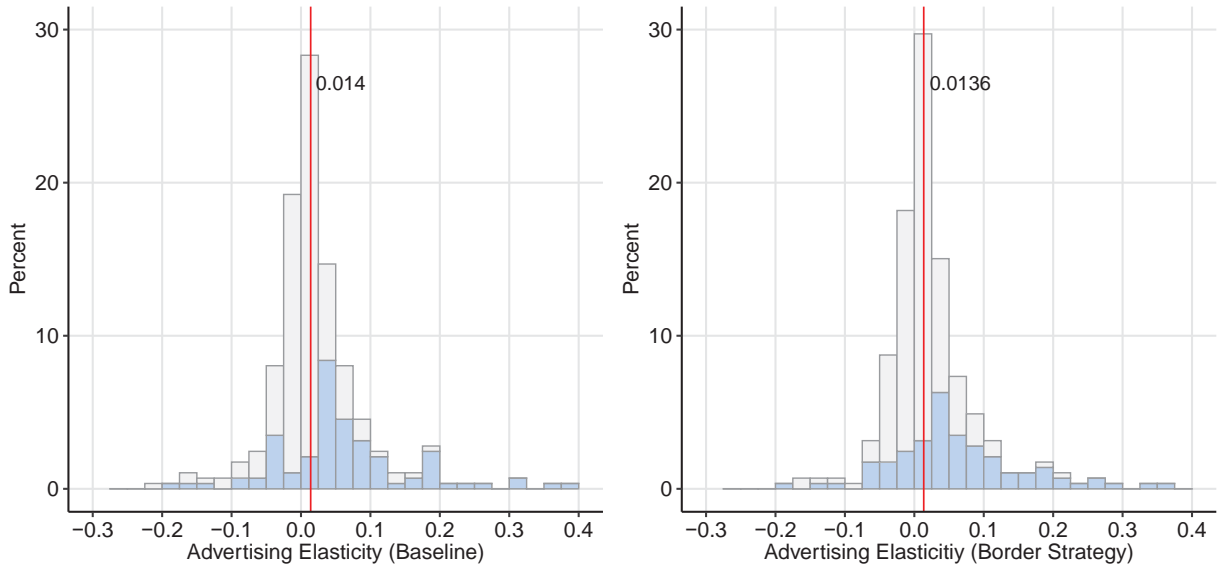


Figure 6: Main Estimation Results: Own-Advertising Stock Elasticities

Note: The estimates are obtained assuming a carryover parameter $\delta = 0.9$. Bars highlighted in blue indicate statistically significant estimates. The vertical red line denotes the median of the distribution.

6.2 Robustness

6.2.1 Incremental adjustment for confounders

In this section, starting with the naive specification, we incrementally add controls (fixed effects) to the model. The purpose of this analysis is to evaluate whether any of our models is sufficiently rich, such that the results are stable and robust to the inclusion of additional controls. We assess stability in terms of the overall distribution and in terms of the individual brand-level estimates.

The results are presented in Figures 7 and 8, and Table 2 contains additional details. The left column in the figures displays histograms of the estimated elasticities for each specification. To the right of each histogram is a scatter plot, with each point representing a brand. In each row, the scatter plot shows the estimated elasticities for the specification shown in the histogram on the y -axis, and the estimates from the specification in the row above on the x -axis. The forty-five-degree line makes it easy to compare the elasticity estimates between two “adjacent” model specifications. Each of the specifications is estimated using the assumed advertising carryover factor of $\delta = 0.9$.

We draw several lessons from this analysis. First, the distribution of estimated advertising elasticities stabilizes once market and week-of-year fixed effects are included in the model. Allowing for a parametric or flexible time trend has little effect on the median of the distribution, although the distributions become tighter. This evidence is consistent with firms adjusting their advertising spending and targeting predictable demand differences at the market- and season-level. On the other hand, there is little evidence that firms are targeting their advertising to

more specific temporal demand shocks, as additional fixed effects do not shift the distribution.

Second, the baseline specification and border strategy yield very similar estimates of not only the distributions but also the brand-level advertising elasticities. These results do not indicate that firms engage in micro-targeting of advertising to idiosyncratic, local demand shocks. The results also alleviate any concerns that consumers who live at the DMA borders differ from the overall population in their response to advertising.

The analysis in this section indicates that both the baseline specification and border strategy advertising elasticity estimates are robust. The evidence does not point to remaining confounding that either of these strategies cannot capture. Hence, subject to the estimated model being an adequate approximation of reality, the estimated advertising effects have a causal interpretation.

6.2.2 Robustness to other modeling choices

We also explore the robustness of the estimates along other dimensions and find that the results are robust to the inclusion or exclusion of various covariates, the treatment of the carryover parameter, δ , and functional form considerations. For the sake of brevity, we present the results only for the border strategy. The full set of results, including results for the baseline specification, is available at <https://advertising-effects.chicagobooth.edu/>. See the bottom panel of Table 2 for the medians, means and quantiles of the distributions depending on inclusion or exclusion of various covariates. In all variations, the estimated distributions are very similar.

Own price The basic model includes own prices. However, it is possible that changes in prices are an “outcome” of advertising, making them bad controls. It is also possible that net of the fixed effects in the model, residual prices are correlated both with the error term and with ad stock. We find that the distribution of ad effects is similar if we do not control for prices. We show in Appendix D that this should be expected, as net of the fixed effects in the model, price and advertising are uncorrelated.³⁰

Included competitor prices In our main specifications, we include the prices of up to three competing products that have the largest total sales revenue in the category.³¹ The results are unchanged whether we include one, two or three of the largest competitors. Hence, we conclude that the omission of additional competitor prices is unlikely to bias the results.

Feature and display advertising Feature and in-store display advertising by retail chains is typically funded by the brand manufacturers and may hence be coordinated with the advertising campaigns. Our main specifications do not include display and feature because these data are only recorded for a subset of 17% of stores in the RMS data. The results are unchanged when we include feature and display advertising. We show in Appendix D that this should be expected, as net of the fixed effects in the model, feature, display and advertising are uncorrelated.

³⁰We show this is true for temporary promotions as well as general price changes.

³¹Categories are defined based on the Nielsen product module code.

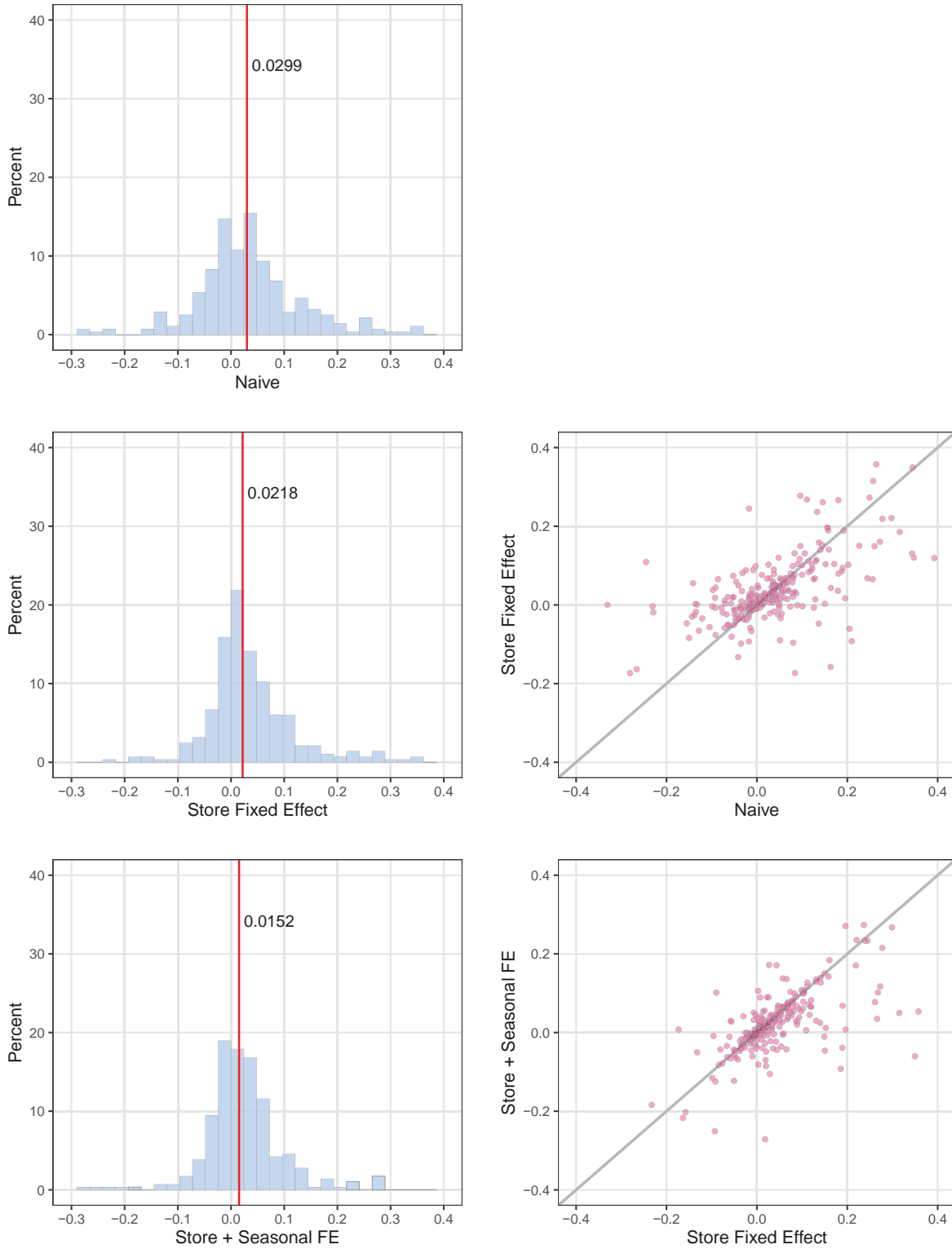


Figure 7: Advertising Stock Elasticities by Specification

Note: The results for the naive model specification are presented in the top panel, and we *incrementally* add store and week-of-year (season) fixed effects. carryover parameter: $\delta = 0.9$.

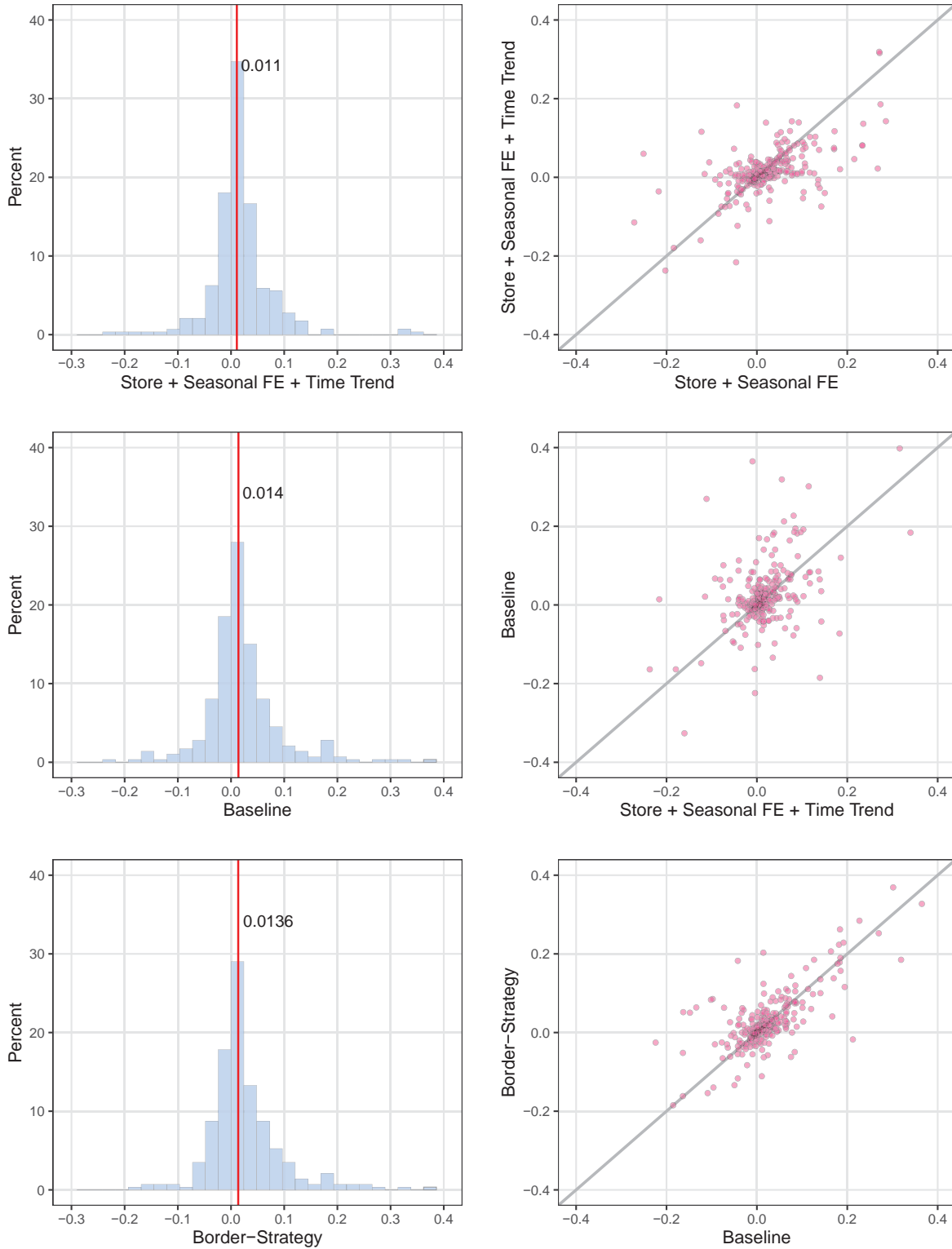


Figure 8: Advertising Stock Elasticities by Specification

Note: Following Figure 7, in the top row we add a linear time trend, whereas the baseline specification uses month fixed effects instead. Carryover parameter: $\delta = 0.9$.

Statistical power 157 of the 288 brands have at least 50% ex ante power to detect an advertising elasticity of 0.05 at the 5% level.³² Within this set, the median advertising elasticity is 0.0073, and the mean is 0.0083. 69.4% of the elasticities are not statistically significant. Hence, the large incidence of estimates that are not statistically significant in the full sample of brands is not simply due to noise. The distribution of advertising effects in the smaller set of 157 brands is compressed—all estimates are less than 0.1 in absolute value. The 90th percentile of the distribution is 0.0367, compared to 0.1015 in the full sample. Hence, the large estimates in the full sample seem to indicate a significant degree of noise rather than a truly large advertising effect.

Calibration of carryover parameter Table 3 presents the estimation results for various values of the carry-over parameter, $\delta = 0, 0.25, 0.5, 0.75, 0.9, 0.95, 1$. The mean and median of the estimated coefficients change when we change the assumed δ . However, the percentage of statistically insignificant coefficients, the percentage of positive and statistically significant coefficients, and the percentage of negative coefficients is robust to any of the assumed δ s.

Estimation of carryover parameter In our main specification we assume the carry over parameter, $\delta = 0.9$. Here we estimate δ using a grid search, using a grid from 0 to 1 in increments of 0.05. For each point in the grid, we calculate the advertising stock using equation (2) and then estimate the remaining model parameters via OLS. For each brand, the estimated δ is the carryover parameter that minimizes the predicted mean squared error.³³ The results are reported in Table 3. The distribution has a similar mean and median but exhibits a larger spread compared to the case when we set $\delta = 0.9$. This is evidenced by the fact that the 10th and 90th percentiles take on more extreme values and the percentage of both positive and negative statistically significant estimates is larger when we estimate, not calibrate, δ .

Functional form We allow for a flexible functional relationship between each component A_j of the advertising stock vector \mathbf{A} and sales using a linear basis expansion. The basis includes polynomials of A_j and $\log(1 + A_j)$ of degree up to 10, and basis B-splines of order 4 on an interval with 9 interior knots. This basis includes the main parametric model specification and a cubic B-spline as special cases. To prevent over-fitting, we estimate the model using a cross-validated Lasso. For each brand we calculate a summary measure of the advertising elasticity that we can compare to the elasticity estimates from the main parametric model specification. In particular, we first calculate the elasticity at each percentile between the 25th and 75th percentiles of the ad

³²Specifically, we identify the set of brands for which the standard error of the brand's estimated ad effect is less than or equal to $0.05/1.96$ (Gelman and Hill, 2007).

³³Estimating δ will yield more accurate advertising effects if the assumption that $\delta = 0.9$ is false or if there is heterogeneity across brands in the degree of advertising carryover. A downside is that if the advertising elasticity is zero ($\beta = 0$), then δ is not identified. In this case, if δ is not restricted, the estimates will be uniformly distributed on $(-\infty, \infty)$. However, since we impose the constraint that $0 \leq \delta \leq 1$, the estimated carryover parameter will likely be at the bounds of the grid, $\delta = 0, 1$. Similarly, in cases where the advertising elasticity β is not precisely estimated, it is likely that δ is also hard to pin down and takes values on the bounds of the grid.

Table 3: Own-Advertising Stock Elasticities by Carryover, δ

		Median	Mean	% $p \geq 0.05$	% $p < 0.05$		Percentiles				
					> 0	≤ 0	10%	25%	75%	90%	
Border Strategy											
	Specified δ	0.00	0.0019	0.0029	78.47	16.67	4.86	-0.0078	-0.0022	0.0069	0.0127
		0.25	0.0036	0.0040	75.35	19.44	5.21	-0.0103	-0.0022	0.0096	0.0153
		0.50	0.0054	0.0059	72.57	21.88	5.56	-0.0133	-0.0040	0.0127	0.0229
		0.75	0.0073	0.0112	68.40	23.61	7.99	-0.0169	-0.0043	0.0214	0.0414
		0.90	0.0136	0.0258	68.40	24.31	7.29	-0.0321	-0.0055	0.0472	0.1015
		0.95	0.0147	0.0381	67.71	23.61	8.68	-0.0560	-0.0079	0.0708	0.1519
		1.00	0.0115	0.0358	76.04	15.97	7.99	-0.0708	-0.0085	0.0817	0.1845
	Estimated δ	0.0111	0.0263	48.96	38.54	12.50	-0.0364	-0.0066	0.0420	0.1324	

Note: Descriptive statistics of estimated advertising elasticities reported for two model specifications and 288 brands. Elasticities derived from regressions of log quantity on log advertising GRP stock (own and competitor) and log prices (own and competitor). The baseline model includes store, month, and week-of-year fixed effects. The border strategy i) restricts the sample to those stores that are located in a county that shares a border with a different DMA and ii) includes store, border-month, and week-of-year fixed effects. Regressions are estimated separately for each brand. The unit of observation in each regression model is a store-brand-week. Standard errors are two-way clustered at the DMA level and the week level in the naive and baseline specifications. Standard errors are two-way clustered at the border-side level and the week level in the border strategy specification.

stock and then summarize the overall advertising elasticity using the median of these values.³⁴

The distribution of the elasticities from the flexible model and a comparison to the parametric model results are shown in Figure 9. The advertising elasticity estimates from the flexible and parametric model specifications are highly correlated. The flexible model makes the extreme negative estimates more negative and the extreme positive estimates more positive, i.e., it magnifies those elasticity estimates that, as we documented above, are particularly noisy.³⁵

6.2.3 Measurement error

If there is classical measurement error in advertising net of fixed effects, the advertising elasticity estimates will be biased towards zero. We measure advertising using GRPs, which are constructed from advertising occurrences and the corresponding television viewership, i.e., number of households who watched the program where an ad was aired. Nielsen uses an automated process involving pattern recognition technology to measure occurrences. Correspondingly, advertising

³⁴Alternatively, we could calculate the advertising elasticity at the mean or median of the observed advertising stock. However, this approach yields noisy results due to wiggles in the estimated advertising response function, i.e. deviations from the overall slope that are local around the mean and median of the advertising stock that do not reflect the slope of the overall advertising response curve.

³⁵Appendix E compares the results for two example brands. In both cases, the $\log(1 + A)$ function appears to provide a reasonable approximation.

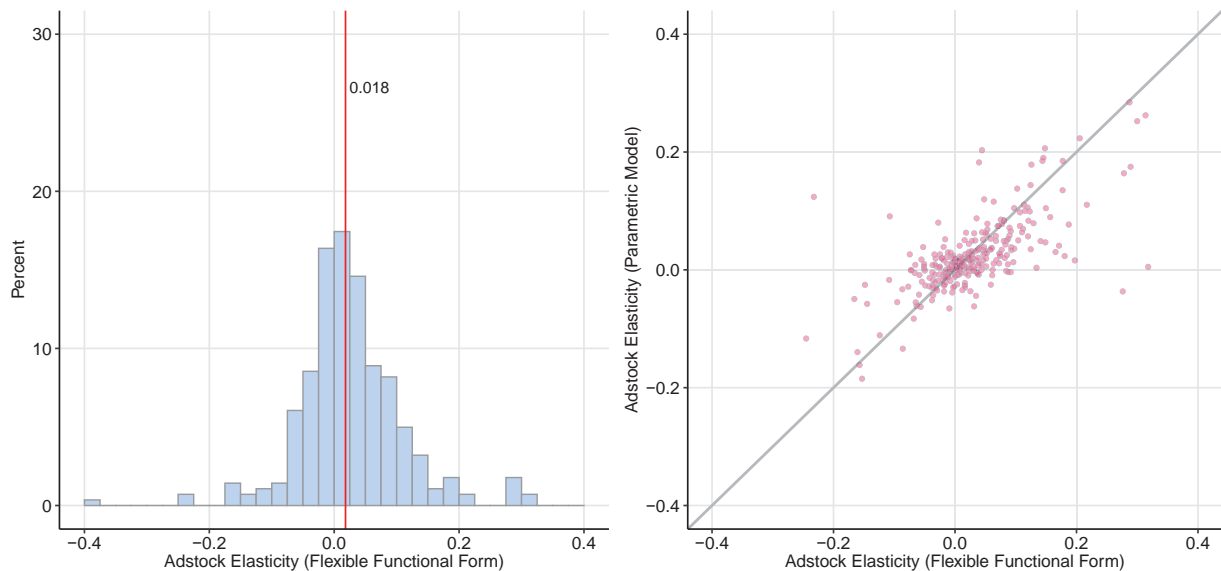


Figure 9: Flexible Functional Form for Advertising Response

Note: The left panel shows the distribution of the estimated brand-level advertising elasticities obtained using the flexible functional form for the advertising response discussed in Section 6.2.2. The vertical line denotes the median of the distribution. The right panel plots the elasticities from the parametric model versus the flexible functional form estimates (both are obtained using the border strategy), together with a 45 degree line.

occurrences and durations are likely to be measured accurately. The advertising viewership is predicted based on the viewing behavior of a sample of households, the “Nielsen Families.” In the top 25 markets, viewership is measured using electronic devices called People Meters. In all other markets, viewership information is collected in the form of self-reported diaries. Hence, the viewership data are the most likely source of measurement error in advertising.

The timing of an ad exposure relative to a shopping trip may induce measurement error. In any given week, some households will be exposed to ads after they have completed the relevant shopping trips. As a result, the measured amount of advertising in the concurrent week may overstate the amount of advertising that could have affected purchasing behavior.

We use four approaches, summarized in Table 4, to assess if the relatively small advertising elasticities that we documented are due to attenuation bias resulting from measurement error.³⁶

Advertising stocks based on durations We use only the occurrence data, which are unlikely to be measured with error, to construct the advertising stocks. The estimated advertising elasticities are similar to the original estimates that are based on GRPs.

LPM markets only Advertising impressions measured using People Meters may be more accurate than the self-reported diary entries. We hence re-estimate the advertising elasticities

³⁶Note, however, that previous studies of TV advertising effectiveness utilize similar data and are hence subject to similar concerns regarding measurement error in advertising.

Table 4: Measurement Error Analysis: Own-Advertising Stock Elasticities

	Median	Mean	% $p \geq 0.05$	% $p < 0.05$		Percentiles			
				> 0	≤ 0	10%	25%	75%	90%
Border Strategy									
GRP stock	0.0136	0.0258	68.40	24.31	7.29	-0.0321	-0.0055	0.0472	0.1015
Occurrence stock	0.0160	0.0327	68.06	24.65	7.29	-0.0367	-0.0072	0.0516	0.1270
1-week-lagged GRP stock	0.0111	0.0165	72.92	21.18	5.90	-0.0400	-0.0091	0.0365	0.0867
Baseline Specification									
GRP, All markets, Time FE	0.0140	0.0233	66.32	26.39	7.29	-0.0406	-0.0082	0.0426	0.0919
Occurrence	0.0170	0.0315	63.89	29.51	6.60	-0.0390	-0.0079	0.0482	0.1289
LPM markets	0.0118	0.0224	74.65	19.79	5.56	-0.0388	-0.0090	0.0428	0.1030
Time Trend	0.0110	0.0171	41.67	42.36	15.97	-0.0360	-0.0053	0.0381	0.0772

Note: Descriptive statistics of estimated advertising elasticities reported for six model specifications and 288 brands. Elasticities derived from regressions of log quantity on log advertising stock (own and competitor) and log prices (own and competitor).

using only the 25 LPM (Local People Meter) markets. The estimates are similar compared to the estimates using all markets.

Fixed effects and measurement error Fixed effects may exacerbate attenuation due to classical measurement error (Griliches and Hausman 1986). In Section 6.2.1 we discussed how the estimates change when we incrementally add fixed effects to the model. The most granular fixed effects are the time fixed effects and in particular the border/time fixed effects in the border strategy. The inclusion of these fixed effects does not yield smaller estimates (Table 2, Figure 8).

One-week-lagged advertising stock We estimate the model using advertising stocks that are lagged by one week. This ensures that all advertising exposures occur before a shopping trip. To obtain elasticities that are comparable to the original estimates, we divide the coefficient on the lagged ad stock by 0.9 to account for the one-week decay. The results are similar to the original estimates.

Based on these analyses, we conclude that the robust small magnitudes of estimated ad elasticities are unlikely due to measurement error.

6.3 Systematic patterns in advertising elasticities

While a distribution of ad effects serves as a useful prior, many managers and researchers may want to know if advertising elasticities are systematically predictable given data on brand or

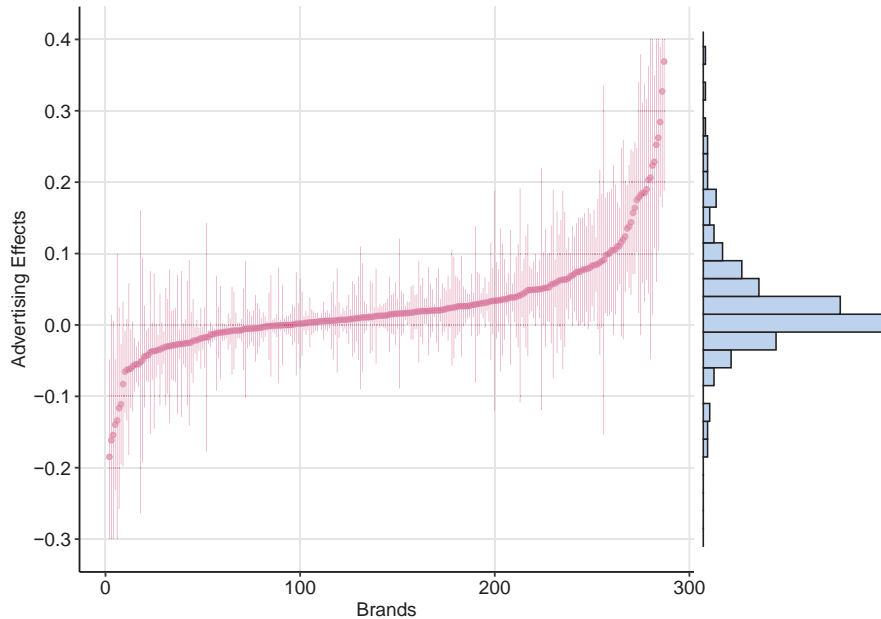


Figure 10: Advertising Effects and Confidence Intervals using Border Strategy

Note: Brands are arranged on the horizontal axis in increasing order of their estimated ad effects. For each brand, a dot plots the point estimate of the ad effect and a vertical bar represents the 95% confidence interval. Results are from the border strategy model with $\delta = 0.9$ (equation 5).

industry characteristics. As noted in the previous section, the estimates with the best statistical power to detect an advertising effect tend to be those closest to the mean of the distribution. This point is illustrated in Figure 10. We arrange the brands on the x -axis in order of the elasticity estimates. On the y -axis we display the elasticities and the corresponding 95% confidence intervals. The estimates near the mean have the smallest confidence intervals, on average, whereas the large estimates in the right and left tails tend to be imprecisely estimated. Hence, as the large estimates are particularly noisy, it will likely be difficult to systematically predict the magnitude of the advertising effects.

Despite these anticipated limitations, we study if there are systematic differences in the advertising elasticities across the 65 product categories or 8 grocery store departments in our data.³⁷ The average elasticity is significantly different from the median at the 5% level in 2 out of 65 product modules, and in one module the average elasticity is significantly different from the mean. Furthermore, the average elasticity is significantly different from the median (but not mean) in 1 out of 8 departments. Hence, as the incidence of statistically significant differences is roughly consistent with results that are obtained by chance, we do not find evidence for a systematic association between product categories or departments and advertising elasticities. This is not to say that other data, such as detailed brand characteristics, would not be able to predict systematic differences in the advertising elasticities. The ability to do so, however, is

³⁷We use the Nielsen product module code to define categories.

limited in our data.

7 Economic Implications

We now discuss the implications of the reported advertising elasticities for the economic value of advertising. Specifically, for each brand we conduct policy experiments to evaluate the change in profits that results from a change in advertising. We report the impact on profitability as the return on investment (ROI) that results from a modification of the brand manufacturer’s advertising policy.

Note that we do not attempt to address by how much advertising should be reduced, and how the corresponding advertising schedule should change. Answering this question requires solving for the dynamically optimal advertising schedule, such as in Dubé et al. (2005), which is beyond the scope of this paper.

Full technical details of how we compute ROI and the data that go into those calculations are presented in Appendix G.³⁸

7.1 ROI results

We consider two policy experiments. First, we compute the average ROI of advertising in a given week. This metric is meant to capture whether firms are (roughly) setting the correct advertising quantity, on the margin. Second, we compute the overall ROI of the observed advertising investment. This analysis is meant to capture how much better or worse off firms are from advertising versus not advertising at all. The results reported below are based on the estimated elasticities from the border strategy with carryover parameter $\delta = 0.9$.³⁹ Standard errors in each case are computed using the delta method.

7.1.1 Average ROI of weekly advertising

The goal of the first policy experiment is to estimate the average ROI of each brand’s observed level of advertising in a given week. For a given brand and week t , we set advertising to 0 across all markets. We predict the total ROI from this change according to the approach detailed in Appendix G. We repeat this process for each week with positive advertising and then average the ROIs across all weeks to compute the average ROI of weekly advertising. We then report the negative, i.e., -1 times the predicted ROI. This reported ROI predicts the return of the observed

³⁸Note that we scale our quantities to national totals using the Nielsen Homescan data. For products where on-site purchase and consumption are commonplace (for example at a fast food restaurant or at a sporting event), the Homescan data will understate total quantity. Beer and soft drinks are particularly likely to be affected by this, while other products are much less so. Separating out the 24 beer and soft drink brands does not significantly alter the distribution of ROI that is ultimately estimated. The results are available by request.

³⁹We also calculated the ROIs using different model specifications and carryover parameters. As the estimates of the advertising elasticities are quite robust to the different assumptions, we focus on a single specification here.

advertising investment relative to a counterfactual with no advertising in a given week, holding the rest of the advertising schedule fixed.⁴⁰

Figure 11 shows the distribution of the predicted ROIs. The three panels show the results separately over a range of assumed manufacturer margin-factors, m , between 20% and 40%, where m represents the manufacturer’s dollar margin as a percentage of the retail price.⁴¹ The estimates are shaded by whether the ROI estimate is statistically different from 0. While the distributions change slightly depending on the margin, the results are broadly consistent and show that the ROI of advertising in a typical week is negative for most of the brands. Table 5 provides further details on the results. At the margin $m = 0.3$, the median ROI is -87.8% , and only 1.1% of brands have a positive and statistically significant ROI (29% of brands include positive ROIs in their confidence interval). If we restrict the sample to the subset of brands with sufficient power to detect an effect of 0.05 at the 5% level, we find that 1.3% of brands have a positive and significant ROI (24% include positive ROI values in their confidence interval). Furthermore, the large percentage of brands with negative ROIs is not exclusively due to brands that are estimated to have a negative advertising effect. Among the brands with a positive advertising effect, 61.8% of the brands have a negative and statistically significant ROI, and the median ROI for these brands is -70.3% .

Table 5 also provides a sensitivity analysis, where we reduce the estimated advertising costs by 20%. The reduction in ad costs has only a small impact on the results. At the margin $m = 0.3$, the reduction increases the median ROI from -87.8% to -84.8% , and the incidence of negative ROIs is reduced from 71.2% to 62.8%. Hence, measurement error in advertising costs does not appear to have a significant impact on the results.

Thus far we documented that there is a large percentage of brands with negative ROIs of advertising in an average week, even amongst brands with a positive ad elasticity estimate. In Appendix G.3, we assess how much larger the TV advertising effects would need to be for the observed level of weekly advertising to be profitable. We find that for the median brand, the estimated advertising effect would have to be more than 6 times larger in order for the observed level of weekly advertising to be profitable (assuming $m = 0.3$).

⁴⁰In all ROI calculations, we hold constant observed prices, as well as advertising for affiliated and competitor brands.

⁴¹Because we do not observe wholesale prices and manufacturing costs for individual brands, we need to make assumptions on what margins the manufacturers earn. The considered range corresponds to a range of manufacturer gross margins between 25% and 55% and retail gross margins between 20% and 30%. To see this, note that m can be expressed as the product of the manufacturer margin and 1 minus the retail margin:

$$m = \left(\frac{w - mc}{w} \right) \left(1 - \frac{p - w}{p} \right) = \frac{w - mc}{p}.$$

Here, p is the retail price, w is the wholesale price, and mc is the marginal cost of production. The range of manufacturer gross margins that we consider aligns with industry reports of median manufacturer gross margins of 34% for food companies, 44% for beverage companies, and 50% for companies selling household goods and personal care products (Grocery Manufacturers Association and PricewaterhouseCoopers 2006).

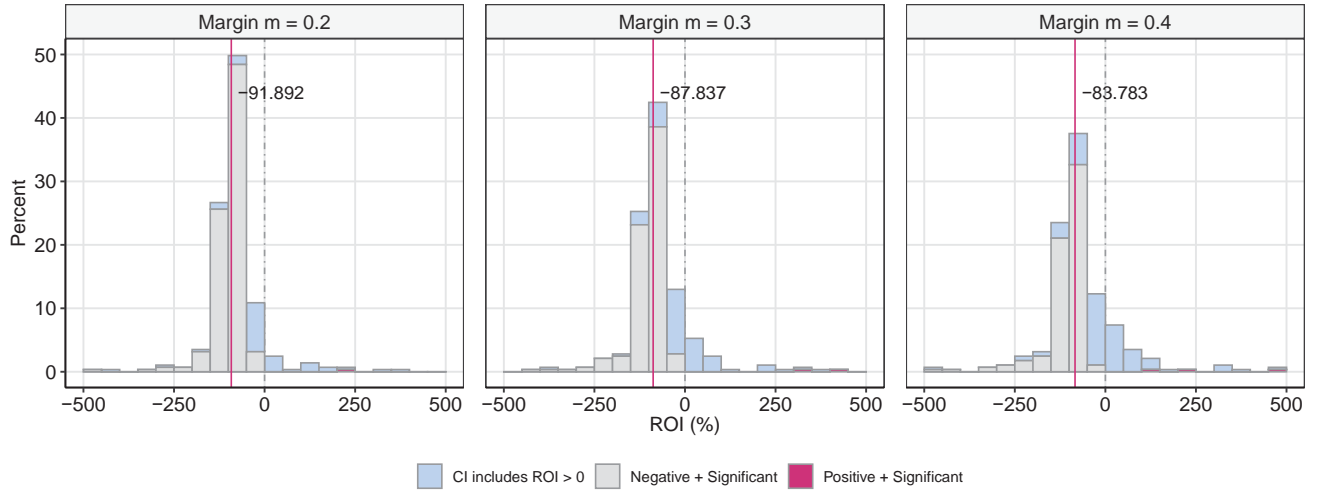


Figure 11: Distribution of Average ROI of Weekly Advertising

Table 5: Average ROI of Weekly Advertising

	Median	Mean	Insig. (%)	Significant (%)		Percentiles			
				ROI > 0	ROI < 0	10%	25%	75%	90%
Main results									
20% Margin	-91.89	-81.98	17.19	0.35	82.46	-122.78	-102.88	-68.03	-28.05
30% Margin	-87.84	-72.97	27.72	1.05	71.23	-134.16	-104.32	-52.05	7.93
40% Margin	-83.78	-63.97	36.49	1.75	61.75	-145.55	-105.76	-36.06	43.91
30% Margin									
Brands with 50% power to detect elasticity = 0.05	-88.37	-82.28	23.08	1.28	75.64	-136.00	-103.96	-61.73	-5.84
Brands with positive ad elasticity	-70.32	-34.28	36.65	1.57	61.78	-95.22	-88.00	-23.40	35.77
Main results with 80% of the estimated advertising cost	-84.80	-66.22	35.44	1.75	62.81	-142.71	-105.40	-40.06	34.91

Note: The estimates are obtained using the border strategy and assuming a carryover parameter $\delta = 0.9$.

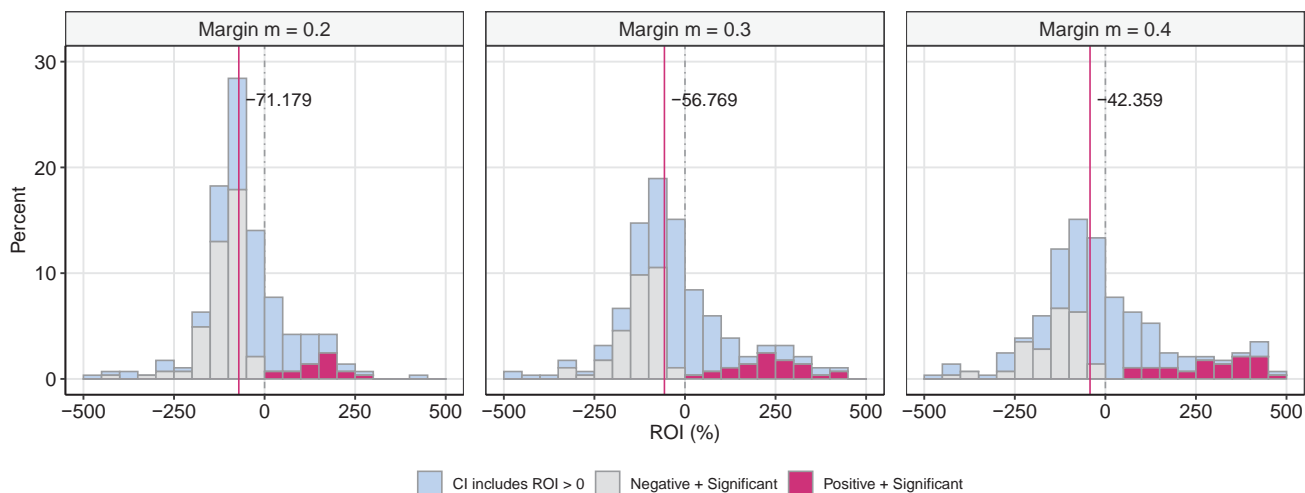


Figure 12: Distribution of ROI of All Observed Advertising

7.1.2 ROI of all observed advertising

In the previous section we showed that for a large percentage of brands, the return on advertising is negative. This statement holds for advertising *at the margin*; in particular, for the actual advertising spending in a given week *conditional* on the firm's observed advertising in the other weeks. We now conduct a different policy experiment to ask if the overall observed advertising schedule has a positive or negative return. Thus, we calculate the ROI of the observed advertising schedule relative to a counterfactual baseline with zero advertising in all periods. Note that relative to the analysis of average weekly ROIs above, the results in this section rely more heavily on the assumed functional form by which the advertising stock affects sales.

Figure 12 shows the distribution of the ROIs of all observed advertising, again separately for the different assumed manufacturer margins in the 20% to 40% range. Table 6 provides further details and summary statistics. The results indicate larger ROIs compared to the weekly advertising returns. Assuming that the manufacturer's margin is 30% ($m = 0.3$), the median ROI is -56.8% , and 11.2% of all brands have a positive and significant ROI. More than two-thirds of brands include positive ROI values in their confidence interval. Furthermore, among the brands with positive advertising elasticities, the median ROI is -3.4% , 16.8% of the brands have a positive and significant predicted return on advertising, and more than 80% of brands include positive ROI values in their confidence interval.

These results show that despite the small advertising elasticities that we documented in Section 6, the observed level of advertising could still be profitable for many brands in the sample.

Table 6: ROI of All Observed Advertising

	Median	Mean	Insig. (%)	Significant (%)		Percentiles			
				ROI > 0	ROI < 0	10%	25%	75%	90%
Main results									
20% Margin	-71.18	-84.40	50.53	7.37	42.11	-177.66	-115.35	-1.75	132.98
30% Margin	-56.77	-76.59	57.54	11.23	31.23	-216.49	-123.03	47.37	249.47
40% Margin	-42.36	-68.79	62.81	13.68	23.51	-255.32	-130.70	96.50	365.97
30% Margin									
Brands with 50% power to detect elasticity = 0.05	-66.74	-61.48	53.85	5.13	41.03	-187.76	-114.41	9.45	100.53
Brands with positive ad elasticity	-3.40	80.49	65.97	16.75	17.28	-79.38	-56.87	128.65	296.54
Main results with 80% of the estimated advertising cost	-45.96	-70.74	62.11	13.33	24.56	-245.62	-128.79	84.21	336.84

Note: The estimates are obtained using the border strategy and assuming a carryover parameter $\delta = 0.9$.

7.2 Discussion

We evaluated the economic significance of the estimated advertising stock elasticities and provided estimates of the ROI of TV advertising for CPG brands that choose to advertise. Based on the estimated advertising effects reported in Section 6.1, we predict that 11% of all brands in our data earn a positive ROI on their total advertising spending during the sample period and another 57% of brands include positive ROI values in their confidence interval. However, the vast majority of brands over-invest in advertising and could increase profits by reducing their advertising spending.

8 Conclusions

In this paper, we provide a generalizable distribution of television advertising elasticities that can serve as a prior distribution for firms and researchers. Providing generalizable estimates of TV advertising effects necessitates transparent and replicable estimation methods and an a priori relevant population of products. Our analysis is based on a sample of 288 large, national CPG brands that are selected using a clear research protocol, and our data sources (Nielsen Ad Intel and RMS scanner data) are widely used by marketing managers and academic researchers. To ensure robustness of the results, we consider a variety of specification choices and identification strategies. We find that the median of the distribution of estimated long-run advertising elasticities is (depending on the exact specification and identification strategy) between 0.0089 and 0.0144, and the corresponding mean is between 0.0102 and 0.0257. We draw two main lessons from these results.

First, the estimated advertising elasticities are small. Roughly two-thirds of all estimates

are not statistically different from zero and roughly seven percent are negative and significant. The estimates are also economically small, in the sense that more than 70% of all brands have a negative ROI of advertising at the margin. The estimates are roughly half the size of the most comparable prior study, Lodish et al. (1995), which used data from the 1980s. This difference is consistent with an overall decline in TV advertising effectiveness over the last three decades.

Second, our results are robust. The results “stabilize” once we adjust for confounding due to market (store) and season specific factors, and we find there is no systematic difference between our baseline specification and border strategy results. While individual estimates move around to some degree, the overall distribution is stable. Hence, although one can never completely rule out confounding in observational studies, it appears implausible that there are any remaining confounds in these data. Finally, our results are robust to considerations of statistical power and functional form, and do not seem to be driven by measurement error.

We caution that while we document over-investment in advertising at the margin, conditional on the observed total spending, this does not imply that all advertising is wasted. For about half of all brands, the observed advertising spending has a positive return over no ad spending.

Our results reveal that even using one of the best and most widely used data sources, advertising effects are either hard to measure or the size and direction of the effects is not always as expected. We hope this work will encourage firms to re-evaluate their advertising strategies and researchers and firms to invest in data and techniques that can improve the measurement of TV advertising effectiveness. We also hope that this work provides an unbiased prior that is useful for decision making.

References

- ABRAHAM, M. M. AND L. M. LODISH (1990): “Getting the Most Out of Advertising and Promotion,” *Harvard Business Review*.
- ASSMUS, G., J. U. FARLEY, AND D. R. LEHMANN (1984): “How Advertising Affects Sales: Meta-Analysis of Econometric Results,” *Journal of Marketing Research*, 21, 65–74.
- BAGWELL, K. (2007): “The Economic Analysis of Advertising,” in *Handbook of Industrial Organization*, ed. by M. Armstrong and R. Porter, Elsevier B.V., vol. 3, chap. 28, 1701–1844.
- BAR, R. N. AND A. M. HAVIV (2019): “Brand Building to Deter Entry and its Impact on Brand Value,” *manuscript*.
- BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63, 841–890.
- BORKOVSKY, R. N., A. GOLDFARB, A. M. HAVIV, AND S. MOORTHY (2017): “Measuring and Understanding Brand Value in a Dynamic Model of Brand Management,” *Marketing Science*, 36, 471–499.

- BRYNJOLFSSON, E., A. COLLIS, AND F. EGGERS (2019): “Using massive online choice experiments to measure changes in well-being,” *Proceedings of the National Academy of Sciences*, 116, 7250–7255.
- 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, 207–236.
- CLARKE, D. G. (1976): “Econometric Measurement of the Duration of Advertising Effect on Sales,” *Journal of Marketing Research*, 13, 345–357.
- DENG, Y. AND C. F. MELA (2018): “TV viewing and advertising targeting,” *Journal of Marketing Research*, 55, 99–118.
- DORFMAN, R. AND P. O. STEINER (1954): “Optimal Advertising and Optimal Quality,” *American Economic Review*, 44, 826–836.
- DU, R. Y., M. JOO, AND K. C. WILBUR (2018): “Advertising and Brand Attitudes: Evidence from 575 Brands Over Five Years,” *Working Paper*.
- DUBÉ, J.-P., G. J. HITSCH, AND P. MANCHANDA (2005): “An Empirical Model of Advertising Dynamics,” *Quantitative Marketing and Economics*, 3, 107–144.
- ELLISON, G. AND S. F. ELLISON (2011): “Strategic Entry Deterrence and the Behavior of Pharmaceutical Incumbents Prior to Patent Expiration,” *American Economic Journal: Microeconomics*, 3, 1–36.
- GELMAN, A. AND J. HILL (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press.
- GOLDFARB, A. AND C. TUCKER (2011): “Online Display Advertising: Targeting and Otrusiveness,” *Marketing Science*, 30, 389–404.
- GORDON, B. AND W. HARTMANN (2013): “Advertising Effects in Presidential Elections,” *Marketing Science*, 32, 19–35.
- GRILICHES, Z. AND J. A. HAUSMAN (1986): “Errors in variables in panel data,” *Journal of econometrics*, 31, 93–118.
- GROCERY MANUFACTURERS ASSOCIATION AND PRICEWATERHOUSECOOPERS (2006): “Insights into the Food, Beverage, and Consumer Products Industry: GMA Overview of Industry Economic Impact, Financial Performance, and Trends,” .
- HENNINGSSEN, S., R. HEUKE, AND M. CLEMENT (2011): “Determinants of Advertising Effectiveness: The Development of an International Advertising Elasticity Database and a Meta-Analysis,” *Business Research*, 4, 193–239.

- HITSCH, G. J., A. HORTAÇSU, AND X. LIN (2019): “Prices and Promotions in U.S. Retail Markets: Evidence from Big Data,” *manuscript*.
- HRISTAKEVA, S. AND J. H. MORTIMER (2020): “Impacts of Legacy Discounts in the Market for National Television Advertising,” *Working Paper*.
- HU, Y., L. LODISH, AND A. KRIEGER (2007): “An Analysis of Real World TV Advertising Tests: A 15-Year Update,” *Journal of Advertising Research*, 47, 341–353.
- JOHNSON, G., R. LEWIS, AND E. NUBBEMEYER (2016): “The Online Display Ad Effectiveness Funnel & Carry-Over: A Meta-Study of Ghost Ad Experiments,” *Working Paper*.
- KALYANAM, K., J. MCATEER, J. MAREK, J. HODGES, AND L. LIN (2018): “Cross Channel Effects of Search Engine Advertising on Brick & Mortar Retail Sales: Meta Analysis of Large Scale Field Experiments on Google.com,” *Quantitative Marketing and Economics*, 16, 1–42.
- LEWIS, R. AND D. NGUYEN (2015): “Display advertising’s competitive spillovers to consumer search,” *Quantitative Marketing and Economics*, 13, 93–115.
- LI, X., W. R. HARTMANN, AND T. AMANO (2019): “Identification Using Border Approaches and IVs,” *Available at SSRN 3402187*.
- LODISH, L. M., M. ABRAHAM, S. KALMENSON, J. LIVELSBERGER, B. LUBETKIN, B. RICHARDSON, AND M. E. STEVENS (1995): “How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments,” *Journal of Marketing Research*, 32, 125–139.
- MOSHARY, S., B. SHAPIRO, AND J. SONG (2020): “How and When to Use the Political Cycle to Identify Advertising Effects,” *Marketing Science*, *forthcoming*.
- NEVO, A. (2000): “Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry,” *RAND Journal of Economics*, 31, 395–421.
- PETRIN, A. (2002): “Quantifying the Benefits of New Products: The Case of the Minivan,” *Journal of Political Economy*, 110, 705–729.
- SAHNI, N. S. (2016): “Advertising Spillovers: Evidence from Online Field Experiments and Implications for Returns on Advertising,” *Journal of Marketing Research*, 53, 459–478.
- SETHURAMAN, R., G. TELLIS, AND R. BRIESCH (2011): “How Well Does Advertising Work? Generalizations from Meta-Analysis of Brand Advertising Elasticities,” *Journal of Marketing Research*, 48, 457–471.
- SHAPIRO, B. T. (2018): “Positive Spillovers and Free Riding in Advertising of Prescription Pharmaceuticals: The Case of Antidepressants,” *Journal of Political Economy*, 126, 381–437.

- SHASHOUA, M., J. FOX, AND S. HODERLEIN (2018): “Private Labels, Famous Brands, and Heterogeneous Households: Can High Ad Spending be Justified and are Households’ Advertising Elasticities Stable Across Products?” *manuscript*.
- SINKINSON, M. AND A. STARC (2019): “Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals,” *Review of Economic Studies*, 86, 836–881.
- SPENKUCH, J. L. AND D. TONIATTI (2018): “Political advertising and election outcomes,” *Quarterly Journal of Economics*, *forthcoming*.
- SUTTON, J. (1991): *Sunk Costs and Market Structure*, MIT Press, Cambridge, MA.
- THOMAS, M. (2020): “Spillovers from Mass Advertising: An Identification Strategy,” *Marketing Science* (*forthcoming*).
- TUCHMAN, A. E. (2019): “Advertising and Demand for Addictive Goods: The Effects of E-Cigarette Advertising,” *Marketing Science*, 38, 994–1022.
- WANG, Y., M. LEWIS, AND D. A. SCHWEIDEL (2018): “A Border Strategy Analysis of Ad Source and Message Tone in Senatorial Campaigns,” *Marketing Science*, 37, 333–355.