Generalizable and Robust TV Advertising Effects*

Bradley T. Shapiro
University of Chicago – Booth

Günter J. Hitsch
University of Chicago – Booth

Anna E. Tuchman
Northwestern University – Kellogg

September 17, 2019

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. A single case study cannot provide generalizable results, and hence the literature provides several meta-analyses based on published case studies of advertising effects. However, *publication bias* results if the research or review process systematically rejects estimates of small, statistically insignificant, or "unexpected" advertising elasticities. Consequently, if there is publication bias, the results of a meta-analysis will not reflect the true population distribution of advertising effects. 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. If we only select products with statistically significant and positive estimates, the mean and median of the advertising effect distribution increase by a factor of about five. The results are robust to various identifying assumptions, and are consistent with both publication bias and bias due to non-robust identification strategies to obtain causal estimates in the literature.

*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, 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. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

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. Evaluating the effect of advertising is part of an important literature in marketing and industrial organization. From a normative point of view, a key task of marketing is to predict the profitability or return on investment (ROI) from incremental advertising spending, both in the short and the long run, in order to facilitate good advertising decisions. From a positive point of view, economists and policy-makers are interested in predicting the effect of advertising on product prices, market structure, and ultimately welfare.

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 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.

In this paper, we provide a generalizable distribution of television advertising elasticities. The analysis is based on a sample of 288 consumer packaged goods (CPG) that are selected using a clear research protocol. We estimate the advertising elasticities brand-by-brand, carefully controlling for confounding factors. All results, irrespective of size, sign, or statistical significance, are reported. 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, rather than by summarizing results published in the literature, due to a concern about publication bias. Most of the advertising research in empirical industrial organization and marketing has used a case study approach. Generalizations are either obtained 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. Publication bias arises if the academic review process systematically rejects some studies based on the findings, such as the sign, size, or statistical significance of the results (see Ioannidis 2005, Franco et al. 2014, and Maniadis et al. 2014).[1][2] Advertising is likely to be particularly susceptible to publication bias because advertising effects tend to be small (e.g. Lodish et al.

[1] Andrews and Kasy (2019) show evidence that there is censoring of results in published studies, and they provide a method for correcting the results that are most likely to be over-stated. Frankel and Kasy (2018) characterize conditions on journal objectives under which publication bias could be optimal.

[2] Researchers may not complete or submit research with results that are likely to be rejected. This selection on the studies submitted to journals is frequently referred to as the file drawer problem (Franco et al. 2014).

(1995)'s study of TV advertising and Lewis and Rao (2015)'s analysis of digital advertising), and many academics have prior beliefs that advertising must be effective, since otherwise the large amount of advertising spending in the industry cannot be rationalized. Due to publication bias, a biased prior distribution of advertising effects can be self-perpetuating, for example if estimates outside the likely range of elasticities according to the biased prior distribution are rejected. To avoid publication bias we use a clear research protocol to select products into the sample, and we ensure that all results, irrespective of size, sign, or statistical significance, are reported.

Robustness of the results to the specific assumptions and choices made in the analysis is—in addition to the research protocol used to avoid publication bias—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.

Second, we intend to provide generalizable results on the *causal* effect of advertising on sales. Hence, to ensure that the estimated advertising effects have a causal interpretation, we analyze the robustness of the results to different identification strategies. 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 to obtain causal advertising effects that are easily scalable across brands. First, we consider confounding due to market, season, and more general time-specific factors that are correlated with advertising levels. To control for such confounding factors, we include a rich set of fixed effects in the estimated model. The identifying assumption is that the variation in advertising net of these fixed effects is as good as random, which is plausible given the institutional features of the ad-buying and advertising scheduling process. This identifying assumption will be violated, however, if some advertising is targeted to local, transient demand factors. Hence, we also use a second identification strategy that exploits the discontinuity in advertising across media market (DMA) borders (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 find that the mean and median of the distribution of estimated long-run own-advertising elasticities is 0.025 and 0.014, respectively, and more than two thirds of the elasticity estimates are not statistically different from zero or negative. 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 strategies yield similar estimates for the brands in our samples and virtually indistinguishable advertising effect distributions. The results are also largely 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.

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. Furthermore, the large number of statistically insignificant or negative estimates is in stark contrast to the low incidence of such estimates in the literature. This discrepancy in the results is suggestive of publication bias. Indeed, if we exclude the products with negative or statistically insignificant estimates from the analysis, the mean and median of the advertising elasticity estimates is substantially larger, 0.073 and 0.102, respectively, and more in line with the estimates in the extant literature.

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 all brands.[3] Hence, the majority of brands over-invest in advertising. This result does not imply that all advertising is wasted: For almost half of all 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 assesses the generalizability of our results, and Section 8 discusses the economic implications for profits and ROIs. Section 9 provides a further discussion of our research approach, and Section 10 concludes.

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. 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. The authors go on to explore how different characteristics of each study's data and econometric analyses are correlated with the estimated elasticities. For example, the authors find that models estimated with product-level data produced larger elasticities than studies that used brand-level data. 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

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

were published between 1981 and 2008. The expanded sample includes 751 brand-level short-term advertising elasticities coming from 56 different publications. With the augmented sample, the authors set out to identify the factors that influence advertising elasticities. These factors include product/market factors, data characteristics, and model characteristics.

Although this type of work helps us understand as researchers how the modeling assumptions we make impact the results we obtain, it has two main limitations. First, this approach relies only on published estimates of advertising effectiveness. As such, the distribution does not represent a random draw of potential studies. Assmus et al. (1984) note this as a limitation of their work, and encourage future researchers to build upon their analysis by supplementing published estimates with unpublished academic and industry measures of ad effectiveness. Second, important differences across products may be overshadowed by differences in the analytic approach. For example, Sethuraman et al. (2011) note that advertising elasticities appear to decline over time, and the authors attribute this decline to increased competition in consumer products, improved access to information through the internet, and the introduction of devices like TiVo and DVRs that allow consumers to opt-out of TV ads. While the authors do their best to control for the factors that differ across studies and time periods in their analysis, there were large changes in quality and types of data sources over the 50 year period that they consider, as well as significant innovation in modeling approaches that occurred over this period. This evolution in data and models over time makes it difficult to feel confident that the observed decrease in ad elasticities is truly being driven by changes in the marketplace, as opposed to some unobserved differences in the studies included in the sample. This speaks to the fact that 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 use a single source of data and the same model across estimated TV ad elasticities.

Another class of papers has taken a different tack that helps alleviate some of these concerns. Instead of relying on existing published estimates that derive ad elasticities from different types of source data and models, one can collect data from a single source and time period that covers a wide variety of product categories and analyze the data using the same modeling framework. This approach allows researchers to focus on the variation in ad elasticities that arises across products and explore why these differences exist without having to worry about any variation in elasticity estimates that is driven by differences in modeling approach and data quality. For example, Eastlack and Rao (1989) conducted 19 advertising experiments with the Campbell's Soup Company, the majority of which involved varying the intensity of advertising during the period of study. Only one of these "weight" tests yielded a statistically significant change in sales during the test period, and the lift from that one study was not enough to compensate for the increased ad expenditure.

Most closely related to this study is the seminal work by Lodish et al. (1995), that 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) document an average advertising elasticity of 0.05. 33% of the elasticity estimates were statistically different from zero using a one-tailed test at a 20% level. 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.[4]

In contrast to these studies, our work evaluates television advertising effects using currently available data and methods that are widely employed in the industry and by researchers. Indeed, the BehaviorScan test markets are no longer in use and hence cannot be used for advertising measurement today. Also, although not reported by Lodish et al. (1995), the power of the tests was likely low (Abraham and Lodish (1990) report a total of about 3,000 households in the BehaviorScan markets, and thus 3,000 is the maximum sample size in each test). Furthermore, BehaviorScan included only a small number of small markets, such as Eau Claire, WI and Pittsfield, MA, and the estimated advertising elasticities from these markets may not be representative of the nation as a whole. 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 related to some cross-category studies of television advertising on various outcomes using observational data. For example, Clark et al. (2009) analyze survey data on consumer brand awareness and perceived quality, while Du et al. (2018) examine the relationship between survey-measured brand attitudes and advertising. Deng and Mela (2018) study the effects of micro-targeting using a model that jointly estimates the utility from television viewing with a purchase utility model. They estimate advertising effects for 77 product categories and find mostly small and statistically insignificant average advertising effects. Our work builds on these studies by focusing on the effect of advertising on sales as an outcome. Our analysis of store-level sales data and market-level advertising data is also complementary to these studies that utilize individual-level data. Further, we pay careful attention to the causal interpretation of estimated effects as well as to the sensitivity of our results to different identifying assumptions.

Our work also relates to some recent multi-product studies of online advertising. Goldfarb and Tucker (2011) analyze data on many online-ad campaigns across many different industries, emphasizing that this multi-product approach allows them to draw more general conclusions

[4] In 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 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 or potentially even advertising their products. We are much 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).

about the average effectiveness of online advertising. Similarly, Johnson et al. (2016) conduct a meta-analysis of hundreds of online display ad field experiments and use the distribution of effects across experiments to come up with rules of thumb on relative elasticities at different parts of the purchase funnel. Kalyanam et al. (2018) study the effect of search advertising on brick and mortar sales, using a Bayesian hierarchical model to provide a generalizable distribution of advertising effects.[5] 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, we don't have to worry about selection bias stemming from which companies or brands are willing to run ad experiments.

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:

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

Q_{st} is the quantity (measured in equivalent units) of the product sold in store s in week t, $\boldsymbol{A}_{d(s)t}$ is a vector of advertising stocks (goodwill) in DMA d in week t, and $\boldsymbol{p}_{st}$ is a corresponding vector of prices. We specify the advertising stock or goodwill as:

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

$\boldsymbol{a}_{d(s)t}$, also a vector, is the flow of 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 $L = 52$.

$\boldsymbol{a}_{d(s)t}$ and $\boldsymbol{p}_{st}$ include own and competitor advertising and prices. We measure own advertising using two *separate* variables. The first own advertising variable captures advertising messages that are specific to the focal product or brand. Such advertising is likely to have a non-negative effect on sales.[6] The second own advertising variable captures advertising messages for affiliated

[5] See Meager (2018) for a discussion of the benefits of Bayesian hierarchical models for evidence aggregation. Due to the computational limitations given the large size of the data in this study, we leave the application of such methods for future work.

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

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 for up to three competitors in the model.[7]

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. Dropping the store index s for simplicity, the advertising stock elasticity is given by

$$\frac{\partial Q_t}{\partial A_t}\frac{A_t}{Q_t} = \beta\frac{A_t}{1+A_t}.$$

Thus, β can be interpreted as an upper bound on the advertising stock elasticity. The advertising stock elasticity is a form of long-run elasticity that represents the percentage change in current period sales units resulting from a one percent increase in current and past advertising. Appendix A discusses alternative short run and long run elasticity metrics that can be computed from this model.

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.

We take two approaches that—subject to specific identifying assumptions—provide causal advertising effects. First, following the intuition provided by advertising practitioners, we employ a rich set of fixed effects to control for the confounding factors that advertisers can observe or predict and hence use for targeted media buying. Second, we utilize the plausibly quasi-random variation in advertising across the borders of television markets (DMAs). We refer to the first approach as our *baseline* specification and to the second approach as the *border strategy.*

The first approach, the baseline specification, is based on different fixed effects and control variables. To control for persistent demand differences in a particular area, we employ store fixed effects. To control for aggregate trends in the demand for a product, we employ time fixed

[7] The 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.

effects. Our preferred specification uses month fixed effects, but we also estimate specifications using a parametric time trend or either quarter or week fixed effects. To control for seasonality that occurs within the months in a given year, we employ week-of-year fixed effects.[8] Furthermore, in some specifications we also include indicators for feature advertising and in-store display advertising. As already discussed, the model also includes own prices and the prices and advertising levels of competing brands, which may be correlated with the focal brand's advertising activity. The main idea of this approach is that the fixed effects and other controls capture all predictable factors that affect demand to which advertisers can respond. Hence, the remaining variation in advertising (conditional on the fixed effects and controls) does not represent planned changes in advertising that coincide with predicted demand changes. As a result, the remaining variation in advertising is quasi-random with respect to residual demand. One key factor that induces such residual variation is the institutional practice of the ad-buying process, whereby advertising agencies buy ad slots often many months in advance. Advertising agencies may follow a coarse temporal scheduling guideline such that ad buys are coordinated with predictable seasonal variation in demand, which we capture using week-of-year fixed effects. However, the identifying assumption is that ad buys are not targeted to coincide with more short-lived demand shocks. Other factors that can induce residual variation in advertising include uncertainty from the network as to programming length or alternative ads they have to run, and technical 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.

The demand model for the baseline specification includes the controls and fixed effects discussed above:

$$\log(Q_{st}) = \boldsymbol{\beta}^T \log(1 + \boldsymbol{A}_{d(s)t}) + \boldsymbol{\alpha}^T \log(\boldsymbol{p}_{st}) + \gamma_s + \gamma_{\mathcal{S}(t)} + \gamma_{\mathcal{T}(t)} + \boldsymbol{\eta}^T \boldsymbol{x}_{st} + \epsilon_{st}. \tag{3}$$

γ_s is a store fixed-effect, $\gamma_{\mathcal{S}(t)}$ is a week-of-year fixed effect that captures seasonal effects, and $\gamma_{\mathcal{T}(t)}$ is a time fixed effect. $\boldsymbol{x}_{st}$ is a vector of other controls at the store-week level, including feature and display advertising in some of the model specifications.[9]

If, however, demand shocks are sufficiently local and predictable, then firms could differentially adapt advertising over time in different locations to these demand shocks. If such micro-targeting occurs, the fixed effects and controls in the baseline demand model are not sufficient to yield a causal advertising effect. To address this challenge, our second approach exploits quasi-random variation in local advertising across the borders of DMAs. This research design was first used in Shapiro (2018) to study the effects of television advertising on antidepressant

[8]We also use specifications with quarter fixed effects and with week fixed effects. Using week fixed effects decreases statistical power considerably for many brands, due to the reliance on national advertising. It also makes the week-of-year dummies redundant.

[9]Feature 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.

demand, and 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. The idea is to take advantage of the fact that consumers who live on different sides of DMA borders may face different levels of advertising due to market factors elsewhere in their DMAs. However, these individuals are otherwise similar, making the cross-border comparison a clean way to identify the effect of the differential advertising. In this way, at the borders, observed advertising is "out of equilibrium" relative to the level of advertising that firms would set if they could micro-target more locally. Intuitively, this approach simulates an experiment with two treatment groups.

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 specification (model (3)) to include time fixed effects, $\gamma_{\mathcal{B}(s,t)}$, that are border-specific:

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

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 these alternate specifications in our interactive online appendix (https://advertising-effects.chicagobooth.edu/).[10] As before, we use store fixed effects to control for persistent local factors related to demand.

The border strategy requires that two conditions hold. First, absent advertising, trends in demand across borders must be parallel net of the covariates in the model. This also means that omitted factors that affect demand must follow parallel trends across the border and be independent of changes in advertising.[11] Second, there must be sufficient variation in advertising net of the fixed effects included in the model. Said differently, there need to be significant cross-border differences in advertising, and the differences need to vary over time. We document the extent to which there is residual variation in our advertising data in Section 5.2.

In addition, both the baseline model and the border strategy require the assumption that

[10] This 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.

[11] Shapiro (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.

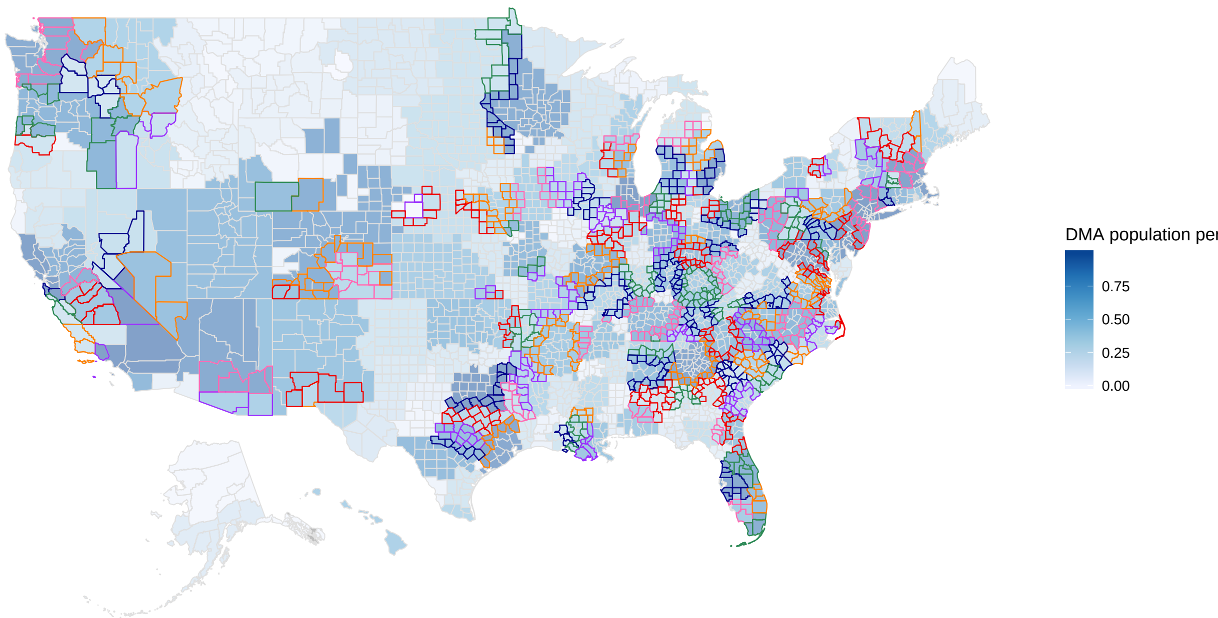

Figure 1: Border Counties in the United States

individuals view advertisements and make purchases in the same DMA. If households live and view advertisements in one DMA but shop across the border in another DMA, then our analyses may fail to capture some incremental sales due to advertising, biasing estimates downward. Tuchman (2019) analyzes purchases made by Nielsen Homescan panelists and finds that there is a very low incidence of cross-DMA shopping.

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.

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.[12] 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 effect of the border strategy on statistical power depends on the relative magnitude 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 (2019) proposes an instrumental variables approach that takes advantage of the fact that advertising decisions are made at a more aggregate level than the measurement of demand in the data. Li et al. (2019) discusses the theoretical conditions under which what they call Waldfogel instruments and the border strategy by Shapiro (2018) are valid, and compares the estimates from the different methods in a case study on the effect of political advertising on vote shares. 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.[13]

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

[12] For 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.

[13] See, for example, Moshary et al. (2019), which shows that the strength of the first stage varies tremendously across product categories when using political advertising as an instrument for product advertising. For most categories, the instrument is weak enough to cause concerns about weak instruments bias.

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. 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 move than 50% of all market-level spending in grocery and drug stores and one third of all spending at mass merchandisers.[14]

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.[15] 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, half-liter bottles, small glass 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

Some of our analyses, the policy experiments and ROI calculations in Section 8 in particular, 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

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

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

and prices paid. Data for more than 60,000 households are available in each of the years. Nielsen provides weights that are 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 the true total purchase volume of a product. 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, 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 and 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. Thus, we have to decide, for example, if an advertisement for "Coca-Cola" should be matched with sales of both regular Coca-Cola and Diet Coke.

We explore four different matching procedures and consider the sensitivity of our results to the match hierarchy. A tier 1 match indicates that the brand name in the sales data exactly matches the brand description in the ad data. A tier 2 match indicates that the brand name in the ad data is more general than the RMS brand name (Ad Intel: COCA-COLA SOFT DRINKS, RMS: COCA-COLA R). Tier 3 matches occur when the brand description in the ad data is more specific than the RMS brand name (Ad Intel: LAYS POTATO CHIPS CHICKEN AND WAFFLE, RMS: LAY'S). Finally, a tier 4 match indicates the situation when an Ad Intel brand is "associated" with but distinct from an RMS brand (Ad Intel: COCA-COLA ZERO DT, RMS: COCA-COLA R).

Because tier 1 and tier 3 matches advertise the RMS product in question and no substitute products, we expect that advertising in these tiers has a positive effect on sales of the focal product. The sign of the effect of tier 2 and 4 matches, however, is ambiguous. In particular, in tier 2 and 4 matches, the ad is relevant both to the focal product and other products that are potentially substitutes. If the partial ad effect on the substitutes is of equal or greater magnitude than the partial ad effect on the focal product, the net ad effect on the focal product could be negative. For example, a Coke Zero ad could reinforce the general Coca-Cola brand and lead to an increase in sales of regular Coca-Cola, which would reflect a positive ad effect. But Coke Zero ads could also lead some consumers to buy Coke Zero instead of regular Coca-Cola, which would appear as a negative ad effect. Correspondingly, in the empirical analysis we include two advertising measures for each brand. The first measure aggregates over the tier 1 and 3 matches, and thus captures advertising that is specific to the RMS brand. The second measure captures the less specific "umbrella" advertising, based on the tier 2 and 4 matches.

The discussion in the previous paragraph illustrates why it is important to match the advertising and RMS scanner data appropriately. We employed a manual process to evaluate the brand descriptions in each data set and to determine the relationships between the brands. The initial merge was carried out by two research assistants, and any disagreements were resolved by the authors.

5 Data Description

5.1 Brand-level summary statistics

Using the process described in Section 4.4, 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.[16]

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

[16]The 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.

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										
Cable	50.9	52.8	10	20.9	29.9	38.4	65.6	81	88.2	99.8
Network	34.5	34.1	0.9	4.2	8.6	19.6	47.5	56.6	66.6	82.4
Spot	3.7	8.7	0.1	0.4	0.7	1.5	8.6	20.9	37.8	82.5
Syndicated	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.

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 5.3 and a 90% percent range from 0.5 to 17.8.

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. Separately for each brand, we regress weekly DMA-level advertising, measured in GRPs, on a set of DMA, month, and week-of-year (season) fixed effects. Additional covariates included in this regression are own and competitor

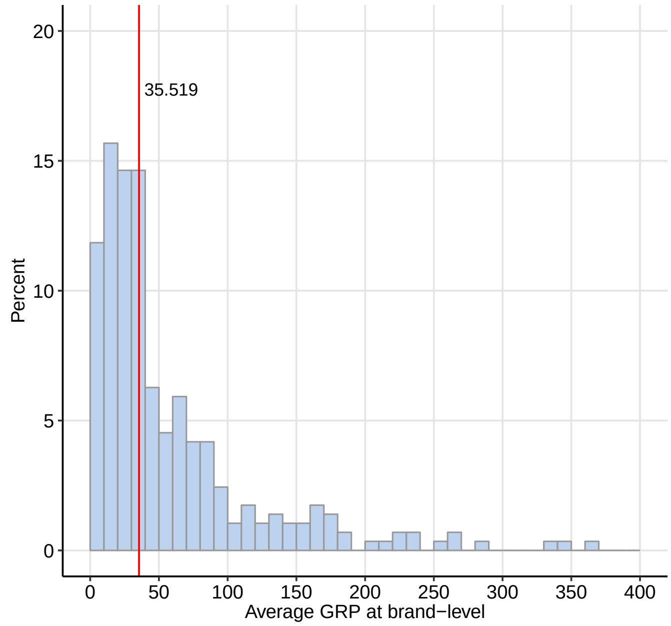

Figure 2: Variation in Advertising Intensity Across Brands

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

prices, and competitor advertising. We calculate the standard deviation of the residuals from this regression, which tells us the amount of residual variation in advertising that is not explained by the fixed effects and the other covariates. Finally, we calculate the ratio of the residual advertising variation relative to average DMA/week advertising. This measure is similar to a coefficient of variation and serves as a parsimonious way of quantifying the amount of variation in advertising net of the fixed effects and 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.[17] 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. In other words, the standard deviation of the residuals is 0.41 times the size of average weekly advertising for the median brand. 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 this averaging smooths out the variability in advertising levels. This makes the task of accurately measuring an advertising stock effect more challenging.

The border strategy discussed in Section 3.2 utilizes cross-border variation in advertising

[17] In the regression used to obtain the residuals, advertising flows are replaced with advertising stocks.

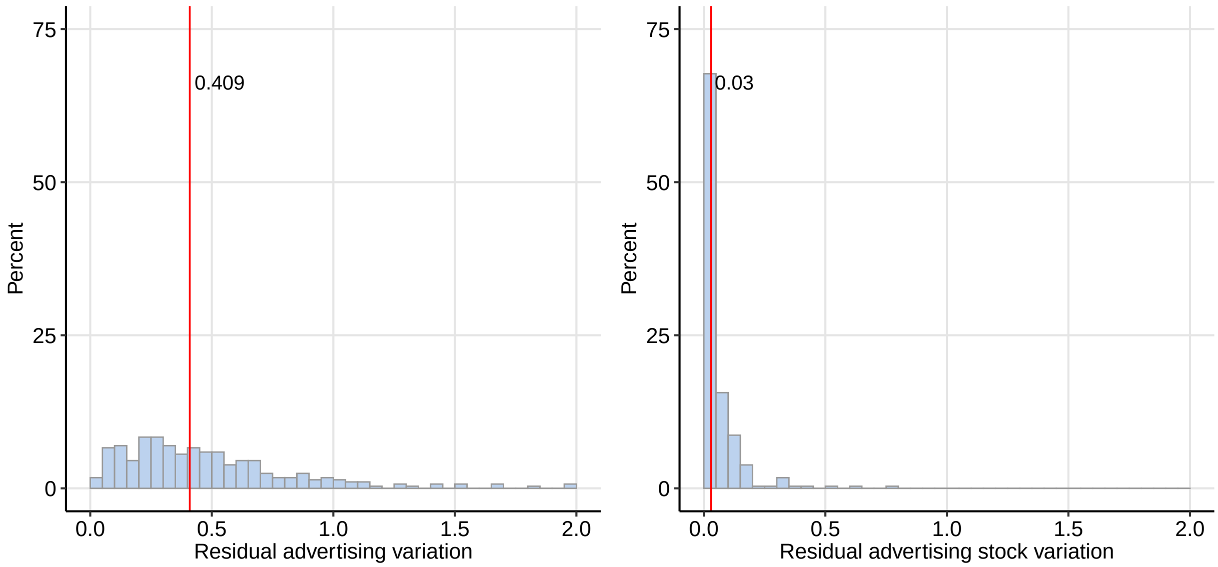

Figure 3: Residual Variation in Advertising

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.

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.[18] 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.

[18] $\overline{\Delta a_j} = \frac{1}{B \cdot T} \sum_{b=1}^{B} \sum_{t=1}^{T} \mid a_{jm_1t} - a_{jm_2t} \mid$

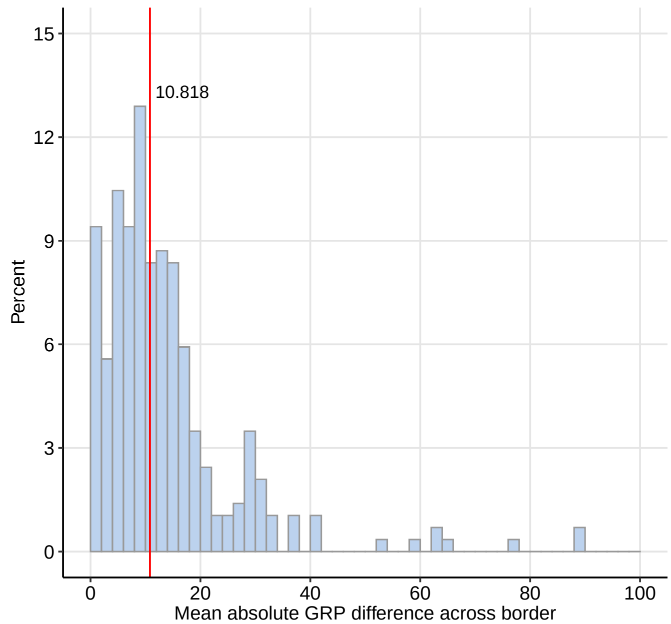

Figure 4: Weekly Absolute GRP Difference Across Borders

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

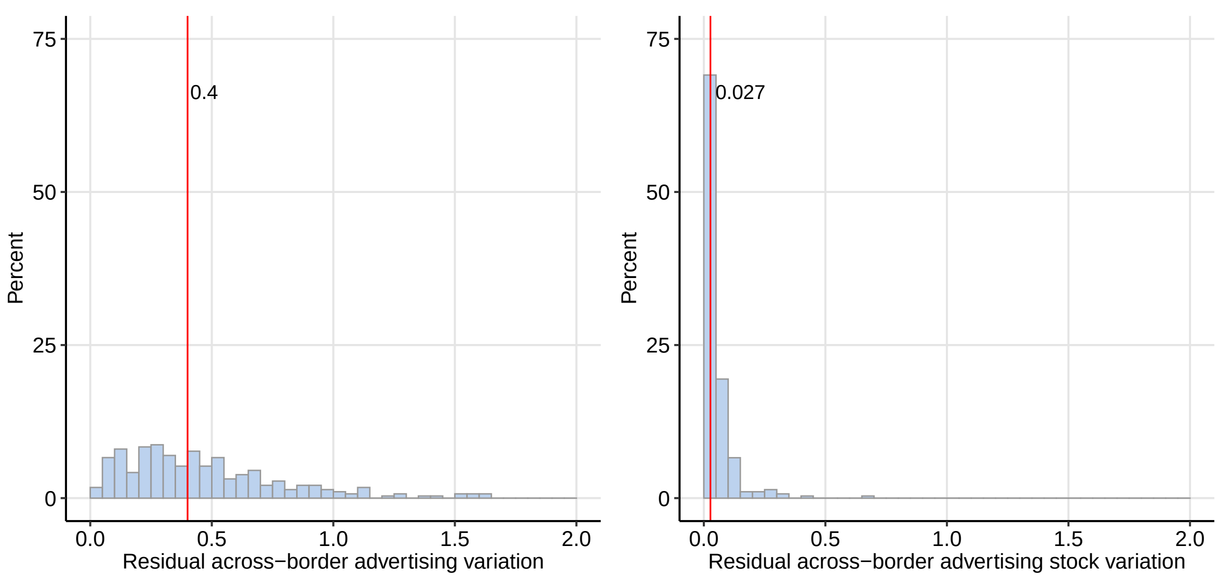

Figure 5: Residual Variation in Advertising: Border Strategy

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 (3)) and the border strategy (model (4)). 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.[19] We then examine the robustness of the results to the exact specification and identification strategy, and show when the estimates stabilize. 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.[20] 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 small subset of all the models we have estimated. Please see https://advertising-effects.chicagobooth.edu/ to explore the sensitivity of the results 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 $\boldsymbol{\beta}$.[21] 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%

[19] For 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.

[20] The naive model includes own price, competitor price, and advertising as covariates.

[21] Appendix C discusses the estimated cross-advertising elasticities.

Table 2: Main Estimation Results: Own-Advertising Stock Elasticities

	Median	**Mean**	**% p ≥ 0.05**	**% p < 0.05**		**Percentiles**			
Specification				**> 0**	**≤ 0**	**10%**	**25%**	**75%**	**90%**
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

Note: The table summarizes the estimated advertising stock elasticities for 288 brands. The time trend is linear, and the baseline specification replaces the time trend with month fixed effects. The estimates are obtained assuming a carryover parameter $\delta = 0.9$. In the naive specification, the specifications with additional fixed effects and a time trend, and in the baseline specification the standard errors are two-way clustered at the DMA level and the week level. In the border strategy the standard errors are two-way clustered at the border-side level and the week level.

of the estimates are negative and statistically significant, and 68.4% of the estimates are not statistically distinguishable from zero.

The results indicate small television advertising elasticity estimates compared to the estimates in the extant literature. In particular, the mean and median from the baseline specification and border strategy are notably smaller than the mean advertising elasticity of 0.23 and median advertising elasticity of 0.10 reported by Sethuraman et al. (2011). Furthermore, two-thirds of our estimates are not statistically distinguishable from zero in both the baseline and the border strategy specifications at a conventional 5% level. While the majority of brands in our data are estimated to have relatively small ad elasticities, there is considerable heterogeneity across brands. Approximately 10% of the brands in our data are estimated to have an ad elasticity larger than 0.10, the median reported by Sethuraman et al. (2011).

As we will discuss at greater length in Section 7, the failure to report and publish null effects provides one hypothesis for the discrepancy between our results and results based on meta-analyses of advertising effects in the literature. Note that even the results from the naive model indicate advertising effects that, on average, are substantially smaller than the estimates reported in the literature, and 58.3% of the estimates are negative or not statistically different from zero. Hence, it is unlikely that the discrepancy in the estimates is primarily the result of a failure to adequately adjust for confounding factors in the extant literature.

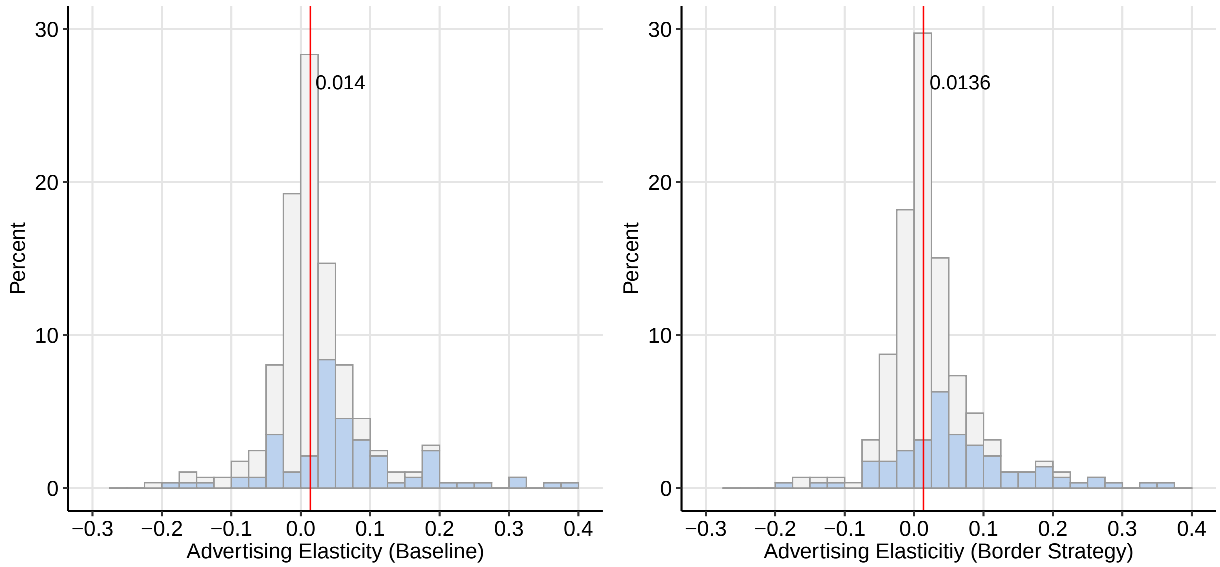

Figure 6: Main Estimation Results: Own-Advertising Stock Elasticities

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 Robustness to Identifying Assumptions

This section documents the robustness of our results to alternative identification strategies. The most basic specification that we estimate is the naive model that includes no fixed effects to adjust for confounding factors. The more sophisticated specifications that we will discuss incrementally add controls for potential confounders. The purpose of this analysis is to show how sensitive the results are to different identifying assumptions, and 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 first compare the naive model to a specification that includes store fixed effects. This specification acknowledges that advertisers are likely to observe persistent differences in demand across markets, and are thus able to adapt their advertising spending to target markets based on the observed demand. The store fixed effects adjust for the demand differences across markets as confounding factors. Adding store fixed effects reduces the median estimated advertising elasticity from 0.0299 to 0.0218, consistent with advertisers targeting markets with strong demand. The scatter plot shows that the majority of positive elasticity estimates from the naive model are shrunk to a smaller value. Next, we add week-of-year fixed effects to the model. This specification adjusts for seasonal confounds, as it is likely that firms can predict seasonality in product demand and adjust their advertising accordingly. Indeed, when the week-of-year fixed effects are added to the model, the median estimated elasticity decreases from 0.0218 to 0.0152. The next two specifications allow for time trends or demand shocks at the national level that the advertisers might be able to predict and use to adjust their advertising accordingly. The first of these two specifications adds a parametric (linear) time trend to the model, whereas the second specification allows for a flexible trend in the form of time (month) fixed effects. The median estimated elasticity is 0.011 for the parametric time trend model and 0.014 for the model with time fixed effects. Overall, these medians are similar to the median of 0.0152 for the more parsimonious model that includes store and week-of-year fixed effects. However, the distributions of elasticity estimates for both time trend models exhibit less variance and are as such tighter compared to the distribution of the the store and week-of-year fixed effects model. Finally, the border-strategy uses border-month fixed effects instead of month fixed effects, and also restricts the sample to stores in border counties. Using the border strategy, the median advertising elasticity, 0.0136, is nearly unchanged compared to the baseline model with store, seasonal, and month fixed effects. Furthermore, the corresponding scatter plot shows that there are very few large brand-level differences between the baseline and the border strategy estimates.

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. Second, the baseline specification and border strategy yield very similar estimates of the brand advertising elasticities. Hence, there is no evidence of micro-targeting of advertising to idiosyncratic, local demand shocks. Also, the evidence alleviates the concern that individuals at the border differ from the overall population in their response to advertising.

Overall, the analysis in this section indicates that both the baseline and border strategy advertising elasticity estimates are robust. The evidence does not point to any 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.

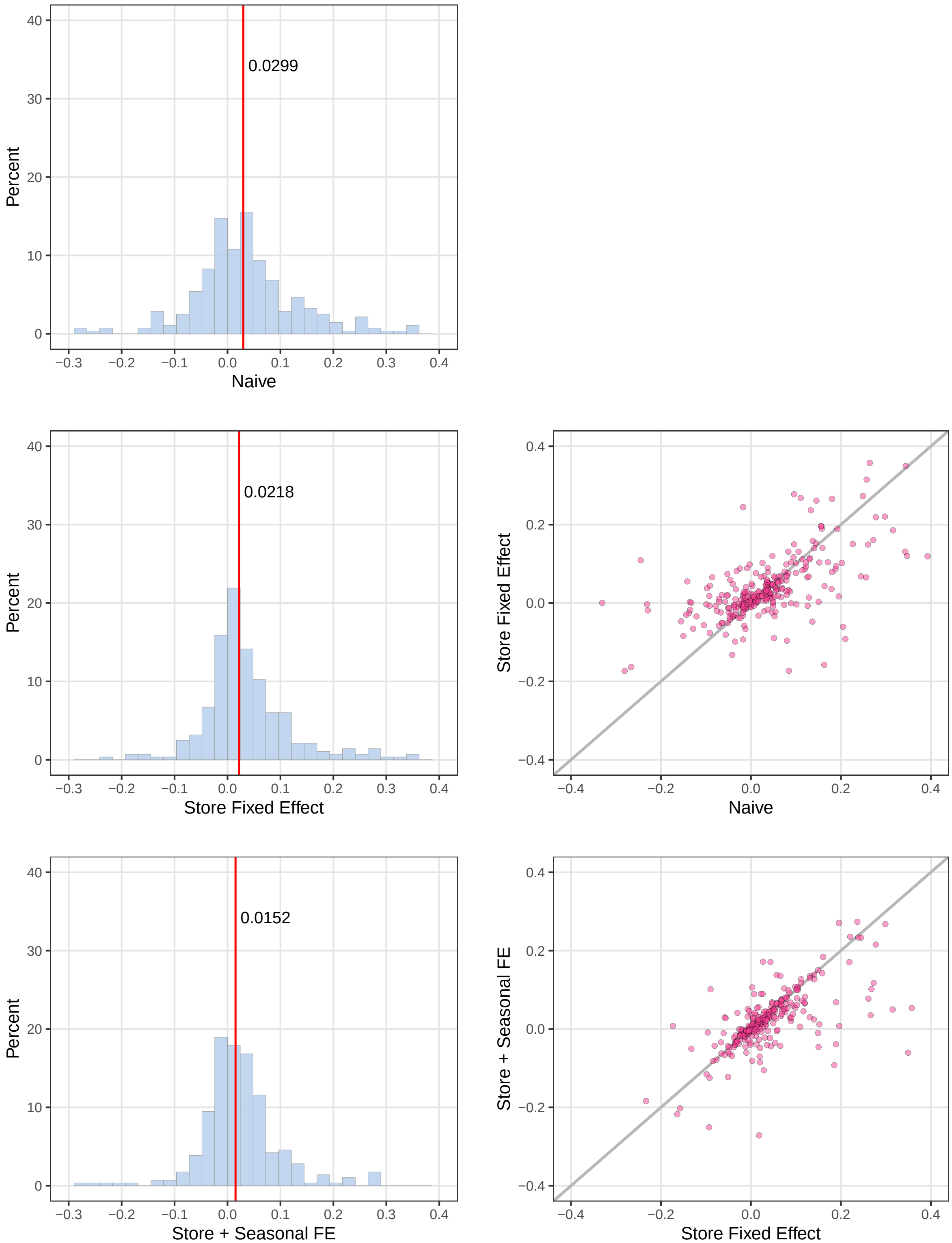

Figure 7: Advertising Stock Elasticities by Specification

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: $\delta = 0.9$.

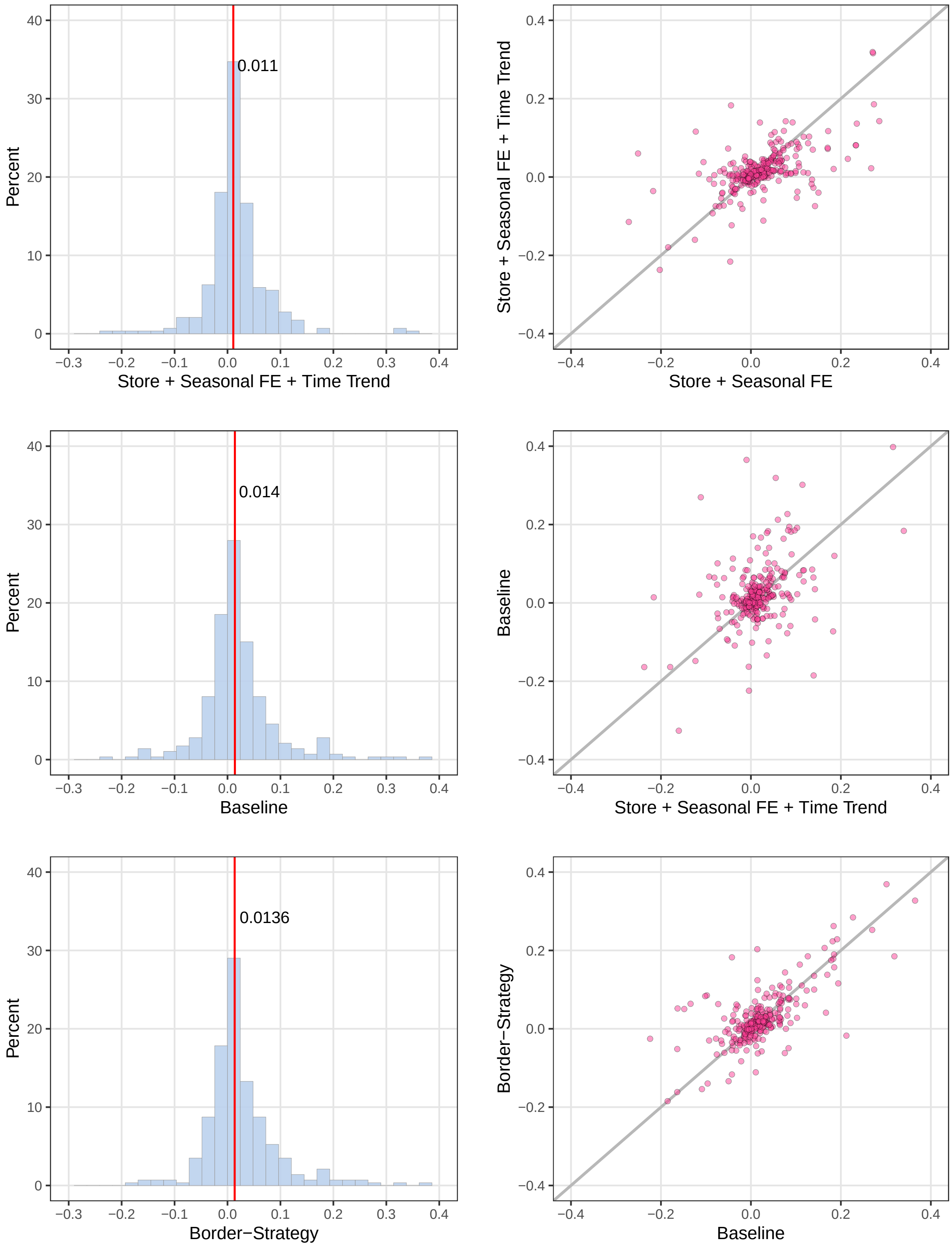

Figure 8: Advertising Stock Elasticities by Specification

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: $\delta = 0.9$.

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%
Baseline Specification										
Assumed δ	0.00	0.0023	0.0030	69.10	23.96	6.94	-0.0079	-0.0031	0.0070	0.0135
	0.25	0.0042	0.0040	69.10	24.31	6.60	-0.0087	-0.0024	0.0092	0.0171
	0.50	0.0050	0.0062	66.32	26.39	7.29	-0.0136	-0.0036	0.0123	0.0254
	0.75	0.0064	0.0115	63.54	29.17	7.29	-0.0173	-0.0052	0.0208	0.0439
	0.90	0.0140	0.0233	66.32	26.39	7.29	-0.0406	-0.0082	0.0426	0.0919
	0.95	0.0144	0.0283	67.36	22.92	9.72	-0.0667	-0.0182	0.0661	0.1483
	1.00	0.0067	0.0137	69.79	18.40	11.81	-0.1498	-0.0292	0.0797	0.2000
Estimated δ		0.0090	0.0116	51.04	35.42	13.54	-0.1102	-0.0149	0.0530	0.1733
Border Strategy										
Assumed 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: The table summarizes the estimated advertising stock elasticities for 288 brands. In the baseline specification the standard errors are two-way clustered at the DMA level and the week level. In the border strategy the standard errors are two-way clustered at the border-side level and the week level.

6.2.2 Robustness to Choice of Carryover Parameter

Thus far, we have assumed an advertising carryover rate of $\delta = 0.9$. To assess the sensitivity of the results to this assumption, we replicate our analysis using alternative values for δ. We only present the results using the border strategy. The results for alternative specifications are available online at https://advertising-effects.chicagobooth.edu/ .

Table 3 presents the estimation results for various values of the carryover 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 carryover parameter. However, the share of statistically insignificant coefficients, the share of positive and significant coefficients, and the share of negative coefficients is robust to any of the assumed δs.

6.2.3 Results Using Estimated Advertising Carryover Factor

So far we have assumed a constant carryover parameter, δ, across brands. We now allow for an additional degree of freedom in the model and estimate the carryover parameter, δ, using a grid search. We use 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.

Estimating 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.

Figure 9 shows the distribution of the advertising effect estimates, separately for the baseline specification and the border strategy, and Table 3 summarizes the results. Using the baseline specification, the median advertising elasticity is 0.0090 and the mean is 0.0116. 13.5% of the estimates are negative and statistically significant, and 51.0% of the estimates are not statistically distinguishable from zero. Using the border strategy, the median advertising elasticity is 0.0111 and the mean is 0.0263. 12.5% of the estimates are negative and significant and 49.0% of the estimates are not statistically distinguishable from zero. When we estimate δ, the distribution of the estimated advertising elasticities exhibits a larger spread compared to the case when we set $\delta = 0.9$. This is evidenced by the fact that the 10^{th} and 90^{th} percentiles take on more extreme values when we estimate δ, relative to when we set $\delta = 0.9$.

Figure 10 shows the histogram of the estimated carryover parameters, δ. The estimates of δ for the brands that, when $\delta = 0.9$ is assumed, have statistically significant advertising elasticity estimates larger than 0.01 are highlighted in blue. For the brands that do not exhibit relatively large and precise estimates under the assumed δ, we find considerable bunching at the bounds, 0 and 1. This is consistent with the discussion above.

Overall, these results indicate that the main conclusions from Section 6.1 are robust. The median and mean of the estimated elasticities are similar, although generally somewhat smaller, than the results obtained for a carryover parameter value of $\delta = 0.9$. The percentage of both positive and negative statistically significant estimates is larger.

6.2.4 Other Robustness Checks

We conducted extensive robustness analyses that, for the sake of brevity, are not included in the paper. In particular, we used more granular time fixed effects at the week and border-week level, and we estimated specifications with and without controls for own and competitor prices, as well as specifications including feature and display advertising. All corresponding results are

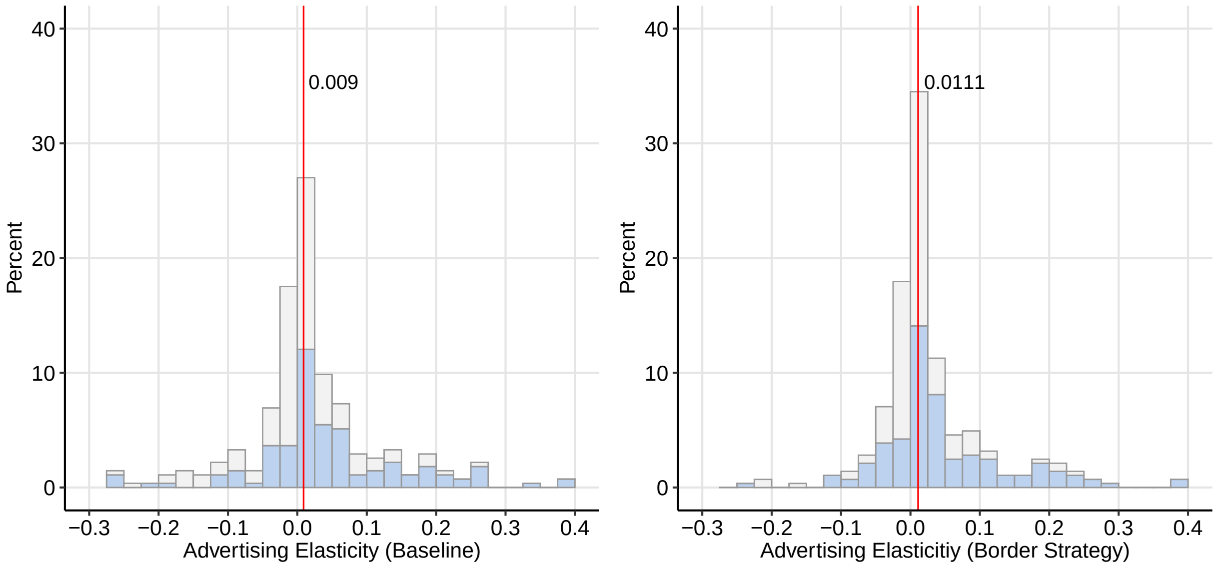

Figure 9: Advertising Stock Elasticities With Estimated Carryover, δ

Note: The left panel shows the distribution of the advertising stock elasticity estimates from the baseline specification, and the right panels shows the estimates from the border strategy. Bars highlighted in blue indicate statistically significant estimates. The vertical red line denotes the median of the distribution.

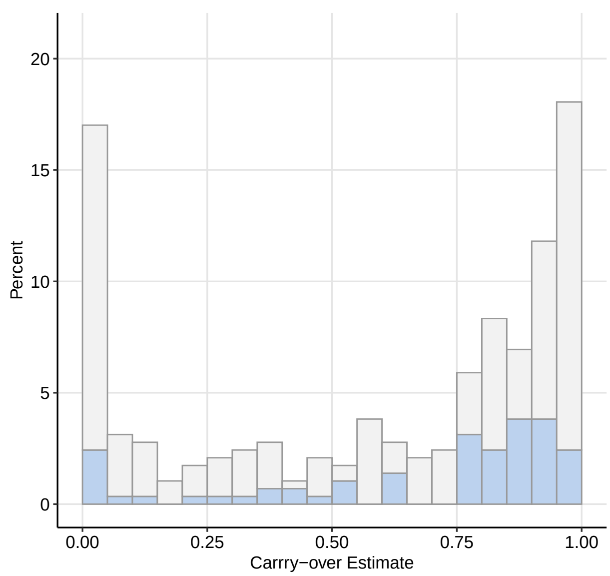

Figure 10: Distribution of Estimated Carryover, δ

Note: The δ estimates for brands that, when $\delta = 0.9$ is assumed, have statistically significant advertising elasticity estimates larger than 0.01 are highlighted in blue.

available online at https://advertising-effects.chicagobooth.edu/.

In all of these specifications, the main conclusions remain robust. In particular, the median and mean of the estimated advertising elasticity distributions are at or near 0.01 and 0.025, respectively, and there is generally a large percentage of non-positive elasticities.

We also considered the robustness of our results to the functional form used to approximate the relationship between advertising and sales. Specifically, we estimate a cubic B-spline, with knots placed at the 10th, 25th, 50th, 75th and 90th percentiles of observed advertising stocks for each brand, and compare the predicted quantities using this flexible function to the predicted quantities using the assumed $\log(1 + A)$ functional form. Appendix D compares the results for two example brands. In both cases, the $\log(1 + A)$ function appears to provide a reasonable approximation.

7 Generalizability

One main goal of this paper is to provide generalizable results on the effectiveness of television advertising. To accomplish this goal, we provided estimates for a large number of brands that were chosen using a clear and replicable research protocol, and we did not selectively drop any brands from the sample based on the estimated advertising elasticities.

Our results indicate relatively small elasticities compared to the results from extant case studies or meta-analyses, and a large percentage of null effects or negative elasticity estimates. In this section, we explore possible sources of the discrepancy between our results and the results in the literature. One explanation is that the case-study approach to conducting empirical research selects results into publication based on the magnitude or statistical significance of the estimates. This selection may be due to a research publication process that systematically rejects findings that have small or unintuitive effect signs and findings that are not statistically significant. Such a process may also lead to the file-drawer problem, whereby researchers do not finish or submit a study in anticipation of rejection. Given such a publication process, the published results are not randomly selected from the true population distribution of effects. Hence, even a meta-analysis will not yield generalizable results. We provide some suggestive evidence on the presence of publication bias by comparing our results to a distribution obtained using some plausible assumptions on how research might get selected into publication. We also examine to what extent the generalizability of our results is affected by statistical power, and, in particular, if the large incidence of small and statistically insignificant estimates is due to a lack of ex ante statistical power to detect an advertising effect.

7.1 Publication Bias

In Figure 11 we overlay the distribution of advertising effect estimates in the full sample (shaded blue) with the distribution of brand estimates that are both positive and statistically different from zero at the 5% level (shaded pink). The latter distribution includes estimates that are

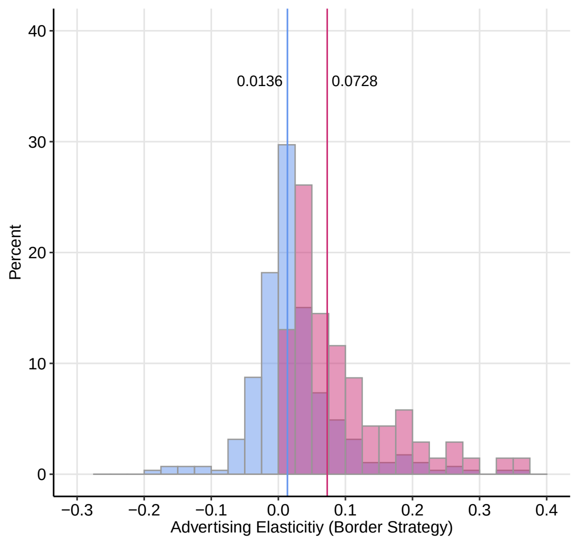

Figure 11: Long Run Ad Elasticities Using Border Model w/ $\delta = 0.9$

Note: Results from estimation of the border strategy with $\delta = 0.9$ (equation 4). Blue bars indicate the full, non-selected distribution of estimates. Highlighted red bars represent the distribution of positive and statistically significant estimates. The selected distribution retains 24.3% of the brands in the full distribution. The mass of the red distribution has been normalized to the mass of the full distribution. The vertical lines indicate the median of each distribution.

likely to survive a publication process that systematically rejects statistically insignificant and "implausible" results. 24.3% of all brands are selected into publication based on this assumed process. Table 4 provides summary statistics on this conditional distribution. In the selected sample, the median ad elasticity is 0.0728 and the mean is 0.1015, compared to a median and mean of 0.0136 and 0.0258, respectively, in the original population of brands. The estimates are now significantly more similar to the median and mean of 0.10 and 0.23, respectively, that are reported in Sethuraman et al. (2011), although the assumed selection process does not fully reconcile the differences. Table 4 also provides results for an even more stringent selection process with a publication threshold of $p < 0.005$. The mean and median in the resulting distribution, which includes 11% of all brands in the original sample, are somewhat larger than the mean and median under the 5% publication threshold.

While not a definite proof, the comparison of the full and selected distributions of estimated advertising elasticities is strongly consistent with publication bias in the extant literature. Quantitatively, the assumed selection process can explain some but not all of the differences with respect to the published results.

7.2 Statistical Power

As another check on the generalizability of our results, we investigate how the advertising effects distribution is affected by noisy (imprecise) estimates. In particular, we examine if the statistically insignificant estimates are true null effects or simply noisy, and whether large estimates are

Table 4: Advertising Effects: Selection on Results and Statistical Power

	Median	Mean	% Brands	% p $\geq$ 0.05	% p < 0.05		Percentiles			
Border Strategy					> 0	≤ 0	10%	25%	75%	90%
Positive & $p < 0.05$	0.0728	0.1015	24.31	0.00	100.00	0.00	0.0202	0.0365	0.1372	0.2238
Positive & $p < 0.005$	0.0832	0.1131	11.46	0.00	100.00	0.00	0.0266	0.0362	0.1639	0.2555
50% power to detect elasticity = 0.05	0.0073	0.0083	54.51	69.43	21.02	9.55	-0.0215	-0.0047	0.0214	0.0367

Note: The first two rows show the distribution of advertising elasticity estimates conditional on a positive estimate with p-value below a threshold of 0.05 or 0.005. The third row shows the results conditional on at least 50% ex ante power to detect an advertising elasticity of 0.05 at the 5% level.

truly large or simply noisy. If the insignificant estimates were simply noisy, the median and mean could still be informative, but the incidence of null effects would not necessarily be generalizable. Similarly, if the large estimates were noisy rather than truly large, they would not be informative about the true distribution of advertising effects.

To explore how imprecise estimates affect the results, we focus on brands with at least 50% ex ante power to detect an advertising elasticity of 0.05 at the 5% level.[22] This set includes 157 of the 288 brands. The results are presented in Table 4. The median advertising elasticity is 0.0073, and the mean is 0.0083. 69.4% of the elasticities are not statistically significant. This percentage is slightly larger than the percentage of insignificant results in the full sample of brands. Hence, the large incidence of estimates that are not statistically significant is not simply due to noise. Rather, the evidence suggests that more than two-thirds of all brands have small ad effects that are indistinguishable from zero.

Notably, the distribution of advertising effects among the brands with at least 50% ex ante power is compressed, and there are no elasticity estimates that are larger than 0.1 in absolute value. The particularly large advertising elasticity estimates that we found in the full sample are eliminated; the 90th percentile of the distribution is 0.0367, compared to 0.1015 in the full sample. This evidence suggests that the large estimated effects indicate a significant degree of noise rather than a truly large advertising effect.

These points are illustrated in Figure 12. We arrange the brands on the x-axis, in order of the point estimates. On the y-axis we plot the estimated advertising elasticities with the corresponding 95% confidence interval. The estimates near zero have the smallest confidence intervals, on average. The estimates in the right and left tails tend to be imprecisely estimated. That is, the large point estimates cannot be distinguished statistically from small values. In a situation with low statistical power, any false positives are predicted to come with large point

[22] Specifically, 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).

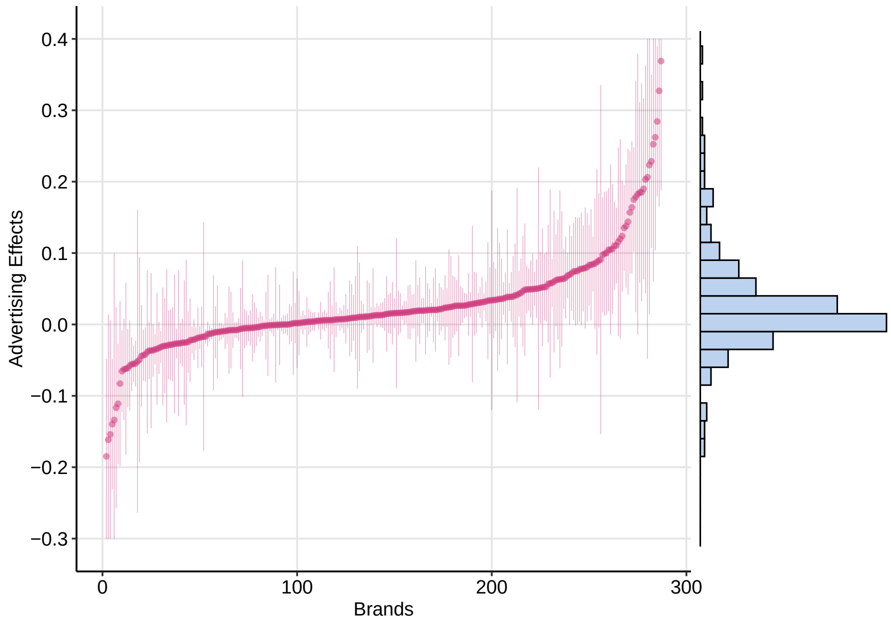

Figure 12: Advertising Effects and Confidence Intervals using Border Strategy

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 4).

estimates.

7.3 Discussion

We draw a few key take-aways from the analysis in this and the previous section. First, the mean of the estimated advertising elasticities is substantially smaller than the mean elasticity of 0.12 reported in Sethuraman et al. (2011). Second, the estimated advertising effects are not statistically different from zero for two thirds of the brands in our sample, and these null effects do not seem to be due to a lack of statistical power. Third, our analysis shows that a process that only selects positive and statistically significant results into publication could substantially bias our understanding of the distribution of advertising effects across brands. Finally, note that the mean of the estimated elasticities is also smaller than the mean elasticity of 0.05 for established brands reported in Lodish et al.'s analysis of split-cable experiments conducted between 1982 and 1988. Because the products in Lodish et al. (1995) are likely similar to our sample of brands, and because it is highly unlikely that the Lodish et al. results are affected by publication bias, this comparison suggests that the effectiveness of TV advertising may be decreasing over time. Such a decline is plausible, given the vast increase in channels and media for advertising compared to the 1980s.

8 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 full dynamically optimal advertising schedule, such as in Dubé et al. (2005), which is beyond the scope of this paper.

8.1 ROI Calculations

Consider the impact of changing brand j's advertising by the amount Δa_d in period t. The baseline advertising stock in DMA d in period t is A_{dt}, and the advertising stock resulting from the change in advertising is $A'_{dt} = A_{dt} + \Delta a_d$. Q_{st} denotes the quantity of brand j sold at store s under the baseline advertising stock, A_{dt}. Consistent with our demand specification, Q_{st} can be written as:

$$\log(Q_{st}) = z_{st} + \beta \log(1 + A_{dt}),$$
$$Q_{st} = e^{z_{st}}(1 + A_{dt})^{\beta}.$$

Here, z_{st} contains all other factors besides advertising that affect quantity sales, including prices, competitor advertising, store, season, and time intercepts, etc. For any period $\tau \in \{t, ..., t + L\}$, the relative change in sales or sales *lift* that results from the change in advertising in period t is:

$$\lambda_{s\tau} \equiv \frac{Q'_{s\tau}}{Q_{s\tau}} = \frac{(1 + A'_{d\tau})^{\beta}}{(1 + A_{d\tau})^{\beta}} = \left(\frac{1 + A_{d\tau} + \delta^{\tau - t} \Delta a_d}{1 + A_{d\tau}} \right)^{\beta}. \tag{5}$$

Notably, all store, season, and time-specific components cancel out, and thus (5) provides the relative increase in overall sales in DMA d that results from the change in advertising. That is, $\lambda_{s\tau} = \lambda_{d\tau}$ for all stores s in DMA d. Hence, the DMA-level change in profits in period τ that results from the increase in advertising in period t is:

$$\Delta \pi_{d\tau} = \sum_{s \in \mathcal{S}_d} (\lambda_{d\tau} - 1) Q_{s\tau} \cdot m \cdot p_{s\tau}, \tag{6}$$

where $\mathcal{S}_d$ includes all stores in DMA d, $Q_{s\tau}$ is the baseline sales quantity in store s, $p_{s\tau}$ is the retail price in the store, and m represents the manufacturer's dollar margin as a percentage of the retail price.[23] Summing across all DMAs and all periods $\tau \in \{t, ..., t + L\}$ yields the total

[23] $m = p^{-1}(w - mc)$, where w is the wholesale price and mc is the marginal cost of production.

increase in profits that results from the advertising increase Δa_d in period t:

$$\Delta\pi = \sum_{\tau=t}^{t+L} \sum_{d=1}^{D} \Delta\pi_{d\tau}.$$

We denote the cost of buying Δa_d GRPs in DMA d by c_{dt}, such that the total cost of the additional advertising is:

$$C = \sum_{d=1}^{D} c_{dt} \Delta a_d.$$

Finally, the ROI resulting from the change in advertising is:

$$ROI = \frac{\Delta\pi - C}{C}.$$

8.2 Data Sources for ROI Calculations

We calculate $\lambda_{d\tau}$, the sales lift that results from changing advertising by Δa_d, using the estimated advertising elasticities from the border strategy with the carryover parameter $\delta = 0.9$.[24] In order to calculate incremental profits using equation (6), we need an estimate of the sales quantities in DMA d in week t (at the observed advertising level, A_{dt}).

The total sales volume from the RMS data under-estimates total market-level sales, because the data available to us do not contain information on all retailers in the market. We correct for this problem as follows. Using the Homescan household panel data and the projection factors provided by Nielsen, we predict market-level quantities, Q_{dt}^H (see Section 4.2). We then calculate the weekly average of the Homescan quantities in market d, $\bar{Q}_d^H$. Similarly, we calculate the weekly average of the market-level sales quantities observed in the RMS data, $\bar{Q}_d^R$.[25] We use the ratio $\bar{Q}_d^H / \bar{Q}_d^R$ to scale the weekly store-level RMS sales quantities such that the aggregate quantity across stores predicts the total sales volume at the market level:

$$Q_{st} = \frac{\bar{Q}_d^H}{\bar{Q}_d^R} Q_{st}^R.$$

We use this hybrid of the RMS and Homescan data because the RMS data are likely to provide more accurate information on sales quantity differences across weeks than the Homescan data, whereas the average Homescan volume provides more accurate information on total market-level sales quantities.

To estimate the dollar margin that a manufacturer earns from an incremental sales unit, we use the observed retail prices in the RMS data and multiply by a margin-factor m that represents the manufacturer's dollar margin as a percentage of the retail price. Because we do

[24]We 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 choose to focus on a single specification here.

[25]The weekly averages are re-calculated for each year in the data.

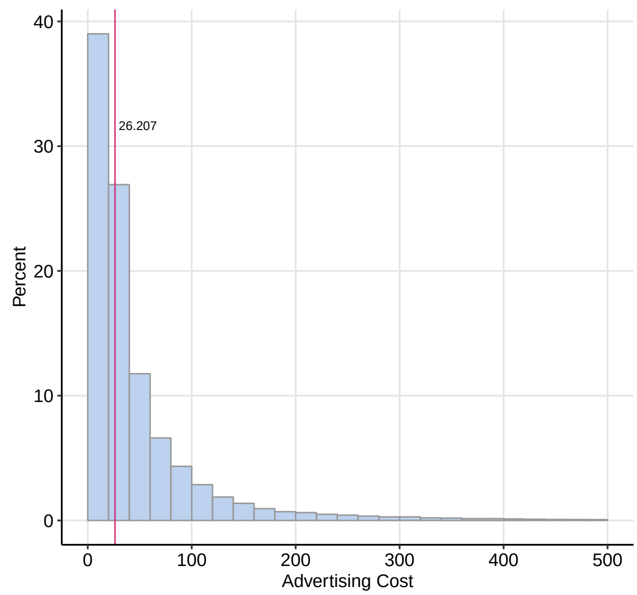

Figure 13: Distribution of Cost per GRP across DMA-Brand-Years

not observe wholesale prices and manufacturing costs, we need to make assumptions on what margins the manufacturers earn. We consider a range of likely values for the manufacturer margin, $m = 0.15, 0.2, 0.25, 0.3, 0.35, 0.4$. This range corresponds to a range of retail gross margins between 20% and 30% and manufacturer gross margins between 15% and 45%.[26] In the results section below, we consider how the distribution of estimated ROIs changes under different assumptions about margins.

Finally, we need data on c_{dt}, the cost of buying an incremental advertising GRP in DMA d in week t. The exact marginal advertising cost is not observed to us. Hence, we use data on advertising expenditures in the Nielsen Ad Intel data set and proxy for c_{dt} using the average cost of a GRP in each DMA-year. We calculate the advertising cost separately for each brand and thus capture differences in the campaign costs across brands.[27] We assess the sensitivity of the ROI predictions to this specific advertising cost calculation to ensure that measurement error in the advertising costs does not substantially change the conclusions.

Figure 13 summarizes the distribution of advertising costs. Each observation in the histogram is the average cost of a GRP calculated for a brand, DMA, and year combination. The median cost of buying one additional GRP in a DMA is \$26.21, although there is significant variation in the cost of advertising across brands, media markets, and years.

[26] To see this, note that m can be expressed as the product of the manufacturer margin and one minus the retail margin

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

[27] Appendix E provides more detail about the advertising expenditure data.

8.3 Results

We consider two policy experiments. First, we compute the average ROI of advertising in a given week. Second, we compute the overall ROI of the observed advertising investment.

8.3.1 Average ROI of Weekly Advertising

The goal of the first policy experiment is to estimate the average ROI of each brand's total observed level of advertising in a given week. For a given brand and week t, we change all positive advertising levels according to $\Delta a_d = -a_{dt}$ in each DMA, and thus stop all advertising for the brand across all markets. We predict the total ROI from this change according to the approach in Section 8.1. We then average the ROIs across all weeks to compute the average ROI of weekly advertising, and report the negative, i.e. -1 times the predicted ROI. This reported ROI predicts the return of the observed advertising investment relative to a counterfactual with no advertising in a given week, holding the rest of the advertising schedule fixed.[28] We focus on the average ROI of all advertising in a given week because it corresponds to a more realistic counterfactual—advertising versus not advertising in a week—than the alternative of predicting the marginal ROI of one incremental GRP. The results for these two alternatives, however, are very similar.

Figure 14 shows the distribution of the predicted ROIs. The six panels show the results separately over the range of the assumed manufacturer margins, m, between 15% and 40%. 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 -79.0%, and only 29.8% of brands have a positive ROI. 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 29.5% of brands have a positive ROI. 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 ad effect, 55.5% of the brands have a negative ROI. The median ROI for these brands with positive elasticities is -24.7%.

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 -79.0% to -73.8%, and the incidence of negative ROI predictions is reduced from 70.2% to 67.4%. Hence, measurement error in the 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. We now assess how much larger the TV advertising effects would need to be for the observed level of weekly advertising to be profitable. Thus, for each brand we compute the break-even advertising elasticity, i.e. the elasticity at which the observed level of

[28] In all ROI calculations, we hold constant all observed prices as well as advertising for competitor brands.

Table 5: Average ROI of Weekly Advertising

	Median	% negative	Percentiles			
			10%	25%	75%	90%
Main results						
15% Margin	-89.512	77.19	-188.421	-105.883	-21.226	227.575
20% Margin	-86.015	74.39	-217.894	-107.844	5.032	336.767
25% Margin	-82.519	72.63	-247.368	-109.805	31.290	445.959
30% Margin	-79.023	70.18	-276.841	-111.765	57.548	555.151
35% Margin	-75.527	68.42	-306.315	-113.726	83.806	664.343
40% Margin	-72.031	66.32	-335.788	-115.687	110.064	773.534
30% Margin						
Brands with 50% power to detect elasticity = 0.05	-77.375	70.51	-257.531	-114.340	38.884	553.673
Brands with positive ad elasticity	-24.659	55.50	-89.763	-79.074	204.031	1233.393
Main results with 80% of the estimated advertising cost	-73.779	67.37	-321.051	-114.707	96.935	718.938

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

weekly advertising has an ROI of 0. For each brand, we then compare the break-even elasticity to the estimated advertising elasticity and determine whether the break-even ad elasticity is within the 95% confidence interval for the estimate. Appendix F discusses this analysis in more detail. We find that for the median brand, the estimated ad effect would need to be 2.5 times larger in order for the observed level of weekly advertising to be profitable (assuming $m = 0.3$). Further, for 61.4% of the brands the break-even advertising elasticity is larger than the upper bound of the confidence interval for the estimated elasticity.

8.3.2 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 15 shows the distribution of the ROIs of all observed advertising, again separately for

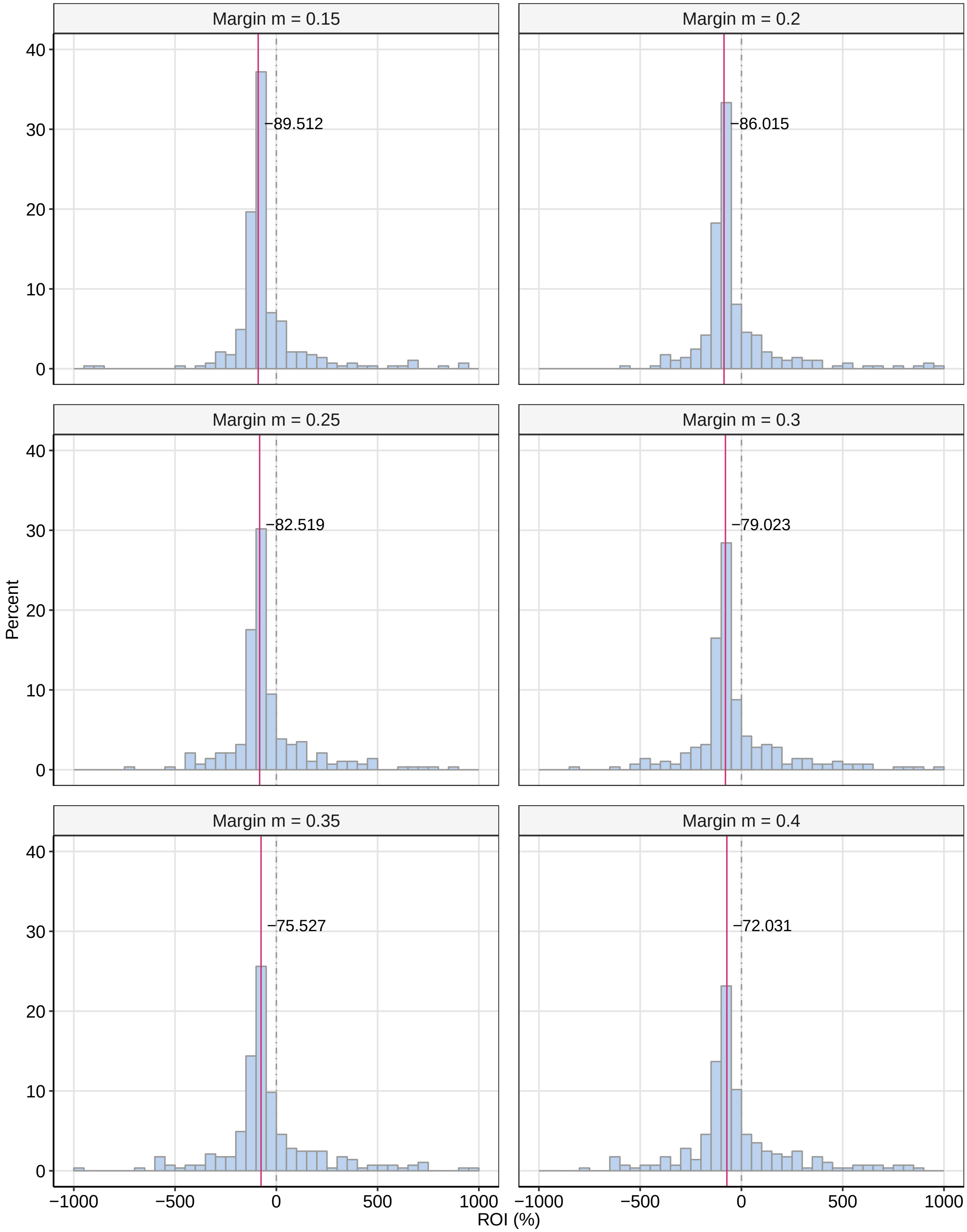

Figure 14: Distribution of Average ROI of Weekly Advertising

Table 6: ROI of All Observed Advertising

	Median	**% negative**	**Percentiles**			
			10%	**25%**	**75%**	**90%**
Main results						
15% Margin	-58.769	64.21	-454.175	-131.354	180.434	928.743
20% Margin	-45.026	59.65	-572.233	-141.806	273.911	1271.657
25% Margin	-31.282	56.14	-690.291	-152.257	367.389	1614.571
30% Margin	-17.539	51.58	-808.349	-162.709	460.867	1957.485
35% Margin	-3.795	51.23	-926.407	-173.160	554.345	2300.399
40% Margin	9.948	47.37	-1044.466	-183.612	647.823	2643.314
30% Margin						
Brands with 50% power to detect elasticity = 0.05	-23.949	53.85	-422.661	-151.467	404.739	1290.284
Brands with positive ad elasticity	133.889	27.75	-63.460	-17.852	965.847	3730.697
Main results with 80% of the estimated advertising cost	3.076	48.77	-985.437	-178.386	601.084	2471.856

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

the different assumed manufacturer margins in the 15% to 40% range. Table 6 provides further details and summary statistics. The results indicate substantially larger ROIs compared to the the weekly advertising returns. Assuming that the manufacturer's margin is 30% ($m = 0.3$), the median ROI is -17.5%, and slightly less than half of all brands have a positive ROI. Furthermore, among the brands with positive advertising elasticities, the median ROI is 133.9%, and 72.2% of the brands have a positive predicted return on advertising.

These results show that despite the small advertising elasticities that we documented, advertising works for almost half of all brands, i.e. the observed level of advertising increases profits over the counterfactual baseline of no advertising. However, the vast majority of brands over-invest in advertising, and could increase profits by reducing advertising. Further, in this case we find that the break-even ad elasticity is significantly larger than the estimated ad elasticities for only 15.8% of the brands in our data. Thus, despite the small ad effects reported in Section 6, our ROI analysis indicates that the observed level of advertising could still be profitable for the majority of brands in the sample.

8.4 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

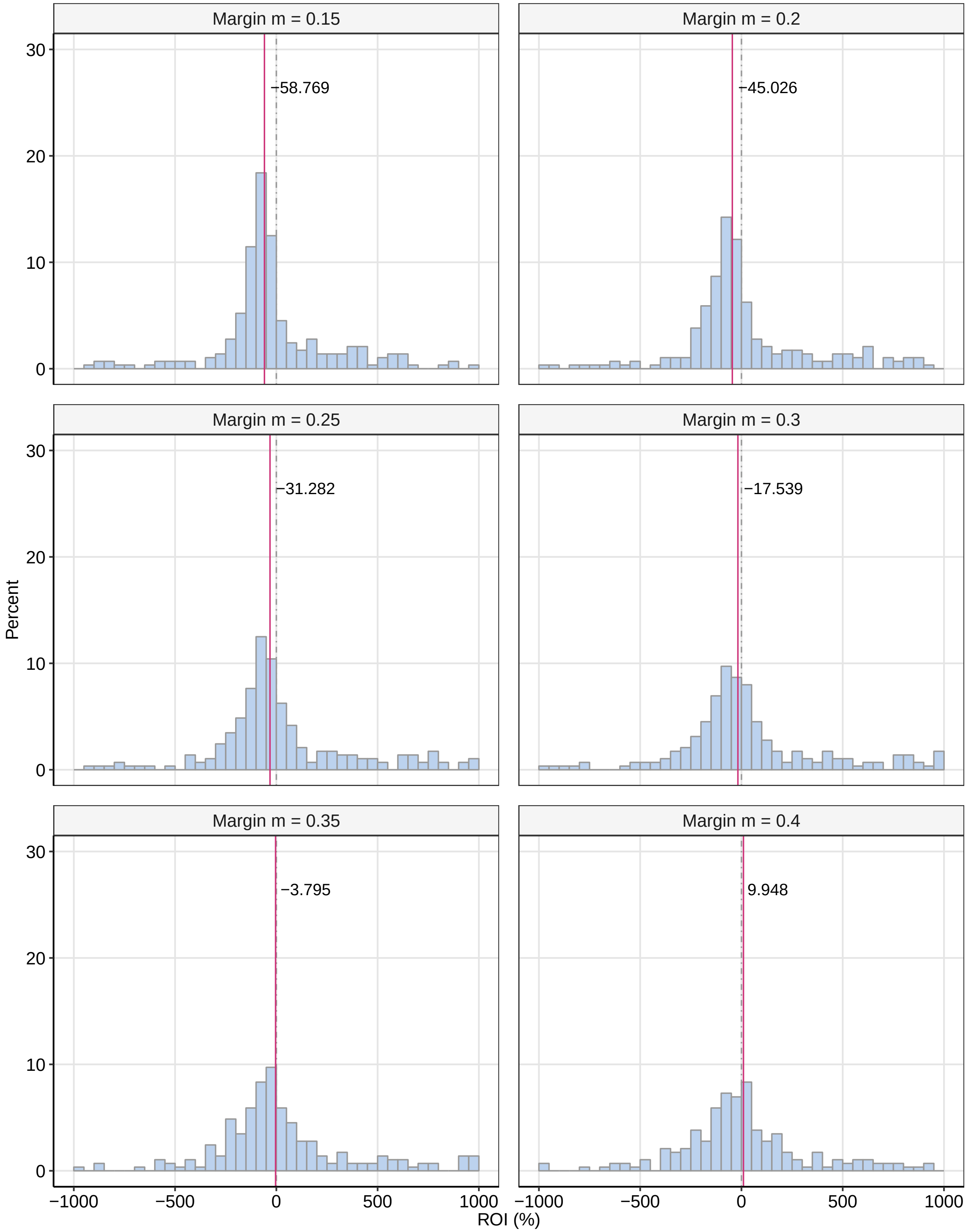

Figure 15: Distribution of ROI of All Observed Advertising

on the estimated advertising effects reported in Section 6.1, we predict that slightly less than half of all brands in our data earn a positive ROI on their total advertising spending during the sample period. The vast majority of brands over-invest in advertising and could increase profits by reducing their advertising spending.

9 Discussion of Our Research Approach

Providing generalizable estimates of TV advertising effects necessitates transparent and replicable estimation methods and an a priori relevant population of products, including the corresponding measures of advertising, quantities, prices, and promotions. We discuss both of these requirements in light of commonly held views on how to obtain valid advertising effect estimates, in particular views communicated to us when we presented drafts of this paper.

9.1 Transparent and replicable estimation methods

We encountered the belief among some researchers that the estimation method should be modified based on the initially obtained results. In particular, some expressed concerns about the negative advertising effects and suggested that these estimates were indicative of potential flaws in the estimation approach. The recommendation was to modify the estimation method and include covariates to avoid such "implausible" advertising effect estimates.

Such views express that inferences about the parameters of interest should incorporate the prior belief on the magnitude of these parameters. Two possible approaches to incorporate prior information are as follows:

(1) Explicitly state a prior distribution on the advertising effects, for example a distribution that only puts positive mass on positive effects, and obtain the final results, the posterior distribution of advertising effects, using Bayesian inference.

(2) If Bayesian inference is computationally too challenging, an alternative approach may be used:

   (i) Using the originally proposed estimation approach, identify the sub-population of products characterized by "implausible" advertising effect estimates.

   (ii) Propose a modified estimation approach for the sub-population identified in step (i) using a clearly documented research protocol. As a more drastic measure, possibly remove products with persistently "implausible" estimates from the sample.

   (iii) Report the final distribution of the estimated advertising effects based on the modified estimation approach in step (ii).

Either approach may yield "better" estimates, in the sense that the estimates improve the decisions that are made based on the results, such as an improvement in the advertising tactics used by a firm or the conclusions from a merger analysis. However, the dependence of the results on prior beliefs needs to be transparently communicated as part of the research.

If approach (1) is used, the researcher needs to explicitly state the prior and thus the dependence of the posterior distribution of advertising effects on the prior belief. It would also be natural to include a sensitivity analysis with a flat (uninformative) prior to evaluate by how much the prior influences the sign and size of the estimated advertising effects.

If approach (2) is used, the researcher needs to explain how the reported distribution of advertising effects in step (iii) depends on step (i), which identifies the "implausible" estimates, and step (ii), which proposes a modification to the originally proposed estimation approach and possibly drops products from the sample. In particular, only reporting the results from step (iii) without a clear explanation of how the results depend on (i) and (ii) is a flawed and misleading research approach. Indeed, most researchers would likely agree that it would be fraudulent for a single team of investigators to use approach (2) but *intentionally* only report the results from the final step (iii). However, the collective publication process may yield the same outcome, even if none of the participants in the process—the authors, reviewers, and editors—are ill-intentioned. In particular, estimates that appear "implausible" after step (ii) may not be selected into publication, either because they get rejected or are never submitted to a journal in anticipation of a rejection (the file drawer problem). This collective process leads to publication bias.

This paper focuses on (i) and leaves an exploration of (ii) and (iii) for future research. In particular, given the high likelihood of and evidence for publication bias in the extant literature, it is important to analyze the population distribution of advertising effects that is based on a priori reasonable estimation methods and free of selection based on estimation results.

9.2 Relevance of the population

The analysis in this paper is based on a large number of CPG products and the Nielsen Ad Intel and RMS scanner data. This data source is widely used by advertising agencies, marketing researchers, and economic consulting companies, and as such, it is an important population to study. In particular, it is important to document the estimates—negative advertising effects in particular—that are a priori unexpected or "implausible." These results reveal that even using one of the best and most widely used data sources, advertising effects are either hard to measure or the direction of the effects is not always as expected. One conclusion that can be drawn is that alternative data or data collection methods may yield more accurate results.

10 Conclusions

In this paper, we provide a generalizable distribution of television advertising elasticities that can serve as a prior distribution for firms and researchers. Our analysis is based on a sample of 288 large, national brands that are selected using a clear research protocol. 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 three 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 estimates are roughly an order of magnitude smaller than benchmarks established by meta-analyses of published case studies (Sethuraman et al. (2011)). However, when we restrict our sample to estimates that are most likely to get published, the magnitude of estimates is closer to the results from the established literature. Our results are suggestive of the existence of publication bias and they help quantify the potential magnitude of the effect of publication bias on the distribution of estimated advertising elasticities.

Third, our results are robust. The results "stabilize" once confounding due to market (store) and season specific factors is adjusted for, 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. Our results are not driven by a lack of precision. When focusing only on brands with high ex ante statistical power, the frequency of small and statistically insignificant advertising elasticities is unchanged.

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.

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.*

Andrews, I. and M. Kasy (2019): "Identification of and Correction for Publication Bias," *American Economic Review*, 109, 2766–2794.

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.

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.

Deng, Y. and C. F. Mela (2018): "TV viewing and advertising targeting," *Journal of Marketing Research*, 55, 99–118.

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.

Eastlack, J. O. and A. G. Rao (1989): "Advertising Experiments at the Campbell Soup Company," *Marketing Science*, 8, 57–71.

Franco, A., N. Malhotra, and G. Simonovits (2014): "Publication bias in the social sciences: Unlocking the file drawer," *Science*, 345, 1502–1505.

Frankel, A. and M. Kasy (2018): "Which findings should be published?" *University of Chicago Working Paper*.

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 Obrusiveness," *Marketing Science*, 30, 389–404.

Gordon, B. and W. Hartmann (2013): "Advertising Effects in Presidential Elections," *Marketing Science*, 32, 19–35.

Hitsch, G. J., A. Hortaçsu, and X. Lin (2019): "Prices and Promotions in U.S. Retail Markets: Evidence from Big Data," *manuscript*.

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.

Ioannidis, J. P. A. (2005): "Why Most Published Research Findings Are False," *PLoS Medicine*, 2, 696–701.

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.

LEWIS, R. A. AND J. M. RAO (2015): "The Unfavorable Economics of Measuring the Returns to Advertising," *Quarterly Journal of Economics*, 130, 1941–1973.

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.

MANIADIS, Z., F. TUFANO, AND J. A. LIST (2014): "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects," *American Economic Review*, 104, 277–90.

MEAGER, R. (2018): "Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature," *manuscript*.

MOSHARY, S., B. SHAPIRO, AND J. SONG (2019): "How and When to Use the Political Cycle to Identify Advertising Effects," *Available at SSRN*.

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*.

THOMAS, M. (2019): "Spillovers from Mass Advertising: An Identification Strategy," *manuscript*.

TUCHMAN, A. E. (2019): "Advertising and Demand for Addictive Goods: The Effects of E-Cigarette Advertising," *Marketing Science (forthcoming)*.

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.

Appendix (Supplementary Online Material)

A Elasticities

To illustrate the possible interpretations of β, we drop the store index s and focus on one specific advertising component, a_t, with corresponding coefficient β. The elasticity of demand in period t with respect to advertising in period $\tau \in \{t-L, t-L+1, \dots t\}$ is given by

$$\frac{\partial Q_t}{\partial a_\tau}\frac{a_\tau}{Q_t} = \beta\delta^{t-\tau}\frac{a_\tau}{1+A_t}.$$

Furthermore, the advertising stock elasticity is equivalent to the total sum of the advertising elasticities:

$$\frac{\partial Q_t}{\partial A_t}\frac{A_t}{Q_t} = \beta\frac{A_t}{1+A_t} = \sum_{\tau=t-L}^{t}\frac{\partial Q_t}{\partial a_\tau}\frac{a_\tau}{Q_t}.$$

To further clarify the difference between the short-run and long-run effect of advertising, suppose that advertising is constant at the level $a_t \equiv a$, such that $A_t = \varrho a$ in all periods t, where $\rho = (1-\delta)^{-1}(1-\delta^{L+1})$. Then the elasticity of per-period demand with respect to the constant advertising flow a is

$$\frac{dQ_t}{da}\frac{a}{Q_t} = \beta\frac{\rho a}{1+\rho a}. \tag{7}$$

This elasticity measures the effect of a permanent percentage increase in advertising, which is bounded above by β. Similarly, assuming again that $a_t = a$ in all periods t, and also that all other factors affecting demand (prices, etc.) are constant, we can derive the effect of a current increase in advertising at time t on total or long-run demand in periods $t, \dots, t+L$:

$$\left(\frac{\partial}{\partial a_t}\sum_{\tau=t}^{t+L} Q_\tau\right)\frac{a_t}{Q_t} = \beta\frac{\rho a}{1+\rho a}. \tag{8}$$

The effect of permanent percentage increase in advertising (7) is equivalent to the total, long-run increase in demand (8). Both effects are bounded above by β and will be approximately equal to β if the advertising stock, ρa, is large. For example, if $\delta = 0.9$, $L = 52$, and advertising $a = 20$ GRPs, then $\rho a/(1+\rho a) = 0.995$, and the long-run demand effect is well approximated by β.

The short-run advertising elasticity is

$$\frac{\partial Q_t}{\partial a_t}\frac{a_t}{Q_t} = \beta\frac{a_t}{1+A_t}.$$

If $a_t = a$ in all periods t and if the advertising stock is large, then

$$\frac{\partial Q_t}{\partial a_t}\frac{a_t}{Q_t} = \beta\frac{a}{1+\rho a} \approx \beta\frac{a}{\rho a}.$$

Hence, the ratio of the long-run effect to the short-run effect of advertising is ρ, which is approximately equal to $1/(1-\delta)$ if δ^L is small.

B Data Construction

The objective of this project is to estimate the effect of TV advertising on retail sales for a wide range of brands. To do that, we need the following data for each brand:

- Weekly volume, price, promotion, and feature/display at store or market level.
- Weekly advertising (GRP, duration, or spending) at television market (DMA) level.

We create the data we want in the following steps:

1. Build Ad Intel Data
   (a) The ad occurrences and viewerships are separate in the raw Ad Intel data. We need to merge them in order to find the GRP for each advertisement.
   (b) There are some discrepancies between the national and local records of Network TV ads. We need to resolve those discrepancies.
2. Create brand map between Ad Intel and RMS data sets.
   (a) Ad Intel and RMS use different brand definitions, so for each RMS brand, we need to find all the corresponding Ad Intel brands.
3. Aggregate Data
   (a) RMS data comes in UPC-Store-Week level, so we need to aggregate it to Brand-Store-Week level.
   (b) Ad Intel data comes in {Ad Intel Brand}-Market-Channel-Second level, so we need to aggregate it to {RMS Brand}-Market-Week level.
4. Identify RMS Stores to be Used in Estimation
5. Identify Products to be Used in Estimation

Each of these steps is described in more detail below.

B.1 Build Ad Intel Data

B.1.1 General Concepts

Media Types Ad Intel covers 4 TV media types: Cable, Network, Syndicated, and Spot.

- For Cable TV, ads are purchased at a national level.

- For Network and Syndicated TV, ads are purchased at a national level. The programs are broadcast at local TV stations.
  - The local TV stations are typically affiliated to a national network. For example, WBZ is the Boston affiliate of CBS.
- For Spot TV, ads are purchased at the DMA level. The programs are also broadcast at local TV stations.

Since Network and Syndicated TV ads are purchased nationally but broadcast locally, the Ad Intel record them in two ways:

- The Network TV and Syndicated TV occurrence files record them at national level.
  - i.e. the date and time each ad is supposed to be broadcast at every local station
- The Network Clearance Spot TV and Syndicated Clearance Spot TV occurrence files record them at local channel level.
  - i.e. the date and time each ad is actually broadcast at every local station
- The local channels have some authority to replace or move nationally scheduled ads, and the Nielsen data is also not perfect. Hence there are discrepancies between those national and local files.

Occurrence Data The occurrence data provides detailed information for each advertisement, including:

- Date [AdDate]
- Time [AdTime]
  - Note that Ad Intel does not capture any local ads between 2AM and 5AM.
- Media Type [MediaTypeID]
- Channel [DistributorCode, DistributorID]
- Market (can be national) [MarketCode]
- Primary, Secondary, and Tertiary Brands [PrimBrandCode, ScndBrandCode, TerBrandCode]
- Duration [Duration]
- The associated TV program [NielsenProgramCode, TelecastNumber]
- Other time-related info [TVDayPartCode, DayOfWeek, TimeIntervalNumber]

Impression (Viewership) Data For the national media types (Cable, Network, and Syndicated), Ad Intel provides the estimated impression for each TV program--defined as a pair of NielsenProgramCode and TelecastNumber.

For the local media types (Network Clearance, Syndicated Clearance, and Spot), Ad Intel provides the estimated impression at {Local Station}-Month-{Day of Week}-{5 Minute Time Interval} level.

Note: There are only 25 markets (the "Local People Meter" markets) for which the local impressions are available in all months. For the rest of markets, local impressions data are only available in four "sweeps months": February, May, July, and November. Therefore, we need to impute the impressions for the non-sweeps months in non-LPM markets. Now we use an average between the two closest available months, weighted by the time difference. For example, for June we use 1/2 May + 1/2 July, and for March we use 2/3 February + 1/3 May.

Universe Estimates Ad Intel also provides the estimated total number of TV audience at national and market level. Those universe estimates are updated yearly.

B.1.2 Build the Regular Parts

The logic of the regular build is very simple. For each media type in each month, we need to do the following:

1. Merge occurrences with impressions
   (a) For national data, merge on NielsenProgramCode and TelecastNumber
   (b) For local data, merge on DistributorID, DayOfWeek, and TimeIntervalNumber
   (c) Remember to do the imputation for non-LPM markets in non-sweep months.
2. Merge the result with universe estimates
3. Calculate the GRP as 100 * Impression / Universe for each ad occurrence

B.1.3 Resolve the "Missing Network" Discrepancy

The objective of this part is also simple: we need to find the national Network TV ads that are not recorded in the Network Clearance data, and if the missing cannot be reasonably explained, we believe that the local data is wrong, and we add those "unexpectedly missing" occurrences into the local records. We say a national ad is "expectedly missing" if it's replaced by another local ad, or if it's scheduled air-time is between 2AM and 5AM. In practice, this procedure is quite complicated to implement. We take the following steps:

1. Find the information for each local station, including:
   (a) The market (MarketCode) and network (Affiliation) for each local station (DistributorCode).

(b) The DistributorID for each DistributorCode.

i. This is in fact a one-to-one relationship, but we have to record that because the "Station Affiliation" data only has DistributorCode, while the impressions data only has DistributorID.

2. For each network and each local station, stack all the monthly data.

(a) We cannot use the raw monthly data because the national and local files have different dates.

(b) Stacking also prevents errors at month boundaries. For example, a national ad at 2012/05/31 23:30:00 may be distributed locally at 2012/06/01 00:30:00. This will not be captured if we process the data month-by-month.

3. For each local station, find the "unexpectedly missing" occurrences. In short, we categorize all the national ads as following:

(a) A national ad is directly matched to the local data if its closest local occurrence has the same primary brand code.

(b) A national ad is indirectly matched to the local data if there's a local occurrence that is aired within some time limit before or after the scheduled air-time. This step accounts for the ads that are moved around. The time limit is 3 hours for ETZ/CTZ, 6 hours for MTZ, and 7 hours for PTZ.

(c) A national ad is replaced by another ad if another spot / network clearance / syndicated clearance ad runs into its scheduled time slot.

(d) A national ad is not captured locally if its scheduled air-time is between 2AM and 5AM.

(e) We mark all remaining national ads as unexpectedly missing at this local station.

4. We get all the "unexpectedly missing" occurrences at each station, and we re-organize them into monthly files. We then merge those monthly files with the monthly local impressions data.

Note: The "broadcast delay" for mountain and pacific time zones causes trouble.

- A nationally scheduled program or ad can be broadcast with a delay of 0/1/2/3 hours in pacific-time markets or 0/1 hours in mountain-time markets. This delay can be pretty arbitrary.

- In step 3, we say a national ad is "unexpectedly missing" only if it's "unexpectedly missing" under all the possible delays, i.e. 0/1 hour in MTZ and 0/1/2/3 hours in PTZ.

- In step 4, for PTZ/MTZ markets we average the impressions at the airtime and 3/1 hours after the airtime.

B.2 Create Brand Map between RMS and Ad Intel

We create a map between the brands in the RMS and Ad Intel data sets using string matching. We classify the matches in 4 "tiers," which are described below. In theory, tier-1 and tier-3 advertising should have a positive effect on sales, while the effect of tier-2 and tier-4 ads can be either positive or negative.

1. RMS and Ad Intel brand names are identical.
2. Ad Intel brand is more general than the RMS brand.
   - Example: Ad Intel brand COCA-COLA SOFT DRINKS is a tier-2 match to RMS brand COCA-COLA R.
3. Ad Intel brand is more specific than the RMS brand.
   - Example: Ad Intel brand LAYS POTATO CHIPS CHICKEN AND WAFFLE is a tier-3 match to RMS brand LAY'S.
4. Ad Intel brand is an "associate" to the RMS brand.
   - Example: Ad Intel brand COCA-COLA ZERO DT is a tier-4 match to RMS brand COCA-COLA R.

We also carry out some module aggregation, which amounts to aggregating some very specific RMS modules together. For example, the RMS modules NUTS-BAGS, NUTS-CANS, NUTS-JARS, and NUTS-UNSHELLED are essentially the same thing, and advertisements never distinguish between them.

Finally, we do some aggregation across flavors and sub-brands. For example, the brand "Lean Cuisine Frozen Entree" has 50 sub-brands in RMS (e.g. LEAN CUISINE ONE DISH FAVORITE or LEAN CUISINE SPA COLLECTION). Aggregating them together makes the matching easier, and it creates more tier-3 matches and fewer tiers-2/4 matches.

B.3 Aggregate Data

Ad Intel The Ad Intel data build comes at the {Ad Intel Brand}-Channel-Time level, and in the end we want to aggregate it to the {RMS Brand}-Market-Week level.

First, we aggregate the ad data to the {Ad Intel Brand}-{Media Type}-Market-Week level. The aggregation here only involves adding up Duration and GRP.

- Some ad occurrences come with 2/3 brands, but those brands are mostly the same product (e.g. Snapple Black Tea and Snapple Green Tea). To avoid double-counting the ads, we use the following trick: if an occurrence has two/three brands, treat it as two/three occurrences with half/one-third of the Duration and GRP.

RMS The RMS data build comes at UPC-Store-Week level, and we want to aggregate it to Brand-Store-Week level.

- One RMS brand may contain hundreds of UPCs with different sizes (size1_amount, say 12 OZ or 24 OZ) and different multi-pack status (multi, say 6-pack or 12-pack).
  - Therefore, instead of using the units field in the RMS data, we need to calculate the volume in equivalency units: volume := units*multi*size1_amount. We adjust price accordingly.
- For each store-week, the brand-level variables are calculated as follows:
  - Volume: sum of UPC-level volumes
  - Price: weighted average of UPC-level prices. The weight for a UPC is its average weekly revenue in this store.
  - Promotion: weighted average of UPC-level promotion indicators (price / base_price < 0.95).
  - Feature/Display: weighted average of UPC-level feature/display indicators (remove missing values).

B.4 Store and Border Selection

We removed the stores that switch between different counties and stores that are not continuously tracked by Nielsen between 2010-2014. We then rank the stores by the total 2010-2014 revenue (across all products), and find the stores that constitute 90% of total revenue. We use those stores for all of our analyses.

Nielsen provides a mapping between counties and DMAs. From this, we constructed a data set that flags the counties that lie on a border between DMAs. However, some counties change DMAs over time, since the borders are re-drawn periodically. Therefore, we removed all the counties that did not stay in a single DMA, and we removed the borders that were re-drawn.

B.5 Product Selection

We began our analysis with the top 500 national brands in the RMS data based on sales revenue between 2010-2014. The above flavor and module aggregation steps reduce the count of unique brands somewhat. We are able to match 358 of these aggregated RMS brands to brands in the Ad Intel data.

Screening Based on Tiers 1+3 Advertising For each of the 358 RMS brands in our universe, we calculate the fraction of market-weeks with positive tiers 1+3 GRPs, and the mean tiers 1+3 GRPs conditional on it being positive. We drop 70 brands who have positive GRPs in less than 5% of observations, or whose "positive mean" is below 10 GRPs.

Table 7: Frequency of Departments and Revenue Share

Department	No. of brands	Homescan revenue share
DRY GROCERY	127	52.19
NON-FOOD GROCERY	50	13.47
HEALTH & BEAUTY CARE	33	4.39
FROZEN FOODS	23	10.75
DAIRY	21	9.90
ALCOHOLIC BEVERAGES	19	3.49
PACKAGED MEAT	11	3.83
DELI	5	2.24
FRESH PRODUCE	1	0.14
GENERAL MERCHANDISE	1	0.20

Note: Three brands in our sample have products in two departments.

C Cross Advertising Elasticities

In the main text we reported own-advertising elasticity estimates. All model specifications control for competitor advertising in the product category, and we now discuss the corresponding competitor advertising effect estimates. While theory predicts that own-advertising effects should typically be positive, the direction of the competitive advertising effect is ambiguous. In the previous literature that has explicitly considered a competitor's advertising effect, some papers have shown positive spillovers of advertising (e.g. Sahni 2016, Shapiro 2018, and Lewis and Nguyen 2015), while others have shown negative, business stealing effects (Sinkinson and Starc 2019). Advertising for a direct substitute may steal sales from the focal brand. However, a competitor brand's ads may also bring new customers into the category and could therefore lead to an increase in sales for the focal brand. The net effect of these different forces depends on the relative strength of these two advertising effects.

Table 9 shows summary statistics for the estimated cross-advertising elasticities corresponding to the baseline and borders model specifications in equations (3) and (4), and Figure 16 shows histograms of the corresponding distributions of advertising effects. Recall that the number of competitor brands included in the model varies across brands and ranges between one and three competitors (see footnote 7). In Table 9 and Figure 16 we only show the cross-elasticities with respect to the top competitor brand, i.e. the competitor brand with the largest market share in the product category.

The distribution of cross-advertising elasticities is centered at zero and very disperse. That is, the particulars of what causes competitor advertising to help or hurt own demand is likely case dependent. Results from past case studies are unlikely to be a good guide for predicting whether any particular cross advertising elasticity will be positive or negative. The location and

Table 8: Frequency of Categories

Category	No. of brands	Category	No. of brands
PAPER PRODUCTS	16	VEGETABLES-FROZEN	3
SNACKS	13	CHEESE	3
CARBONATED BEVERAGES	13	LAUNDRY SUPPLIES	3
BEER	11	SANITARY PROTECTION	3
DETERGENTS	11	WRAPPING MATERIALS AND BAGS	3
CANDY	11	DEODORANT	3
JUICE, DRINKS - CANNED, BOTTLED	10	NUTS	3
PACKAGED MEATS-DELI	10	BABY FOOD	2
SOFT DRINKS-NON-CARBONATED	9	PREPARED FOOD-DRY MIXES	2
CEREAL	9	COOKIES	2
PREPARED FOODS-FROZEN	7	UNPREP MEAT/POULTRY/SEAFOOD-FRZN	2
SALAD DRESSINGS, MAYO, TOPPINGS	6	COT CHEESE, SOUR CREAM, TOPPINGS	2
PET FOOD	6	PACKAGED MILK AND MODIFIERS	2
BREAKFAST FOOD	6	WINE	2
LIQUOR	6	HOUSEHOLD SUPPLIES	2
VITAMINS	6	PET CARE	2
MEDICATIONS/REMEDIES/HEALTH AIDS	6	SKIN CARE PREPARATIONS	2
DISPOSABLE DIAPERS	6	SEAFOOD - CANNED	1
CONDIMENTS, GRAVIES, AND SAUCES	5	PREPARED FOOD-READY-TO-SERVE	1
CRACKERS	5	JAMS, JELLIES, SPREADS	1
COFFEE	5	DESSERTS, GELATINS, SYRUP	1
PIZZA/SNACKS/HORS DOEURVES-FRZN	5	TEA	1
DRESSINGS/SALADS/PREP FOODS-DELI	5	SPICES, SEASONING, EXTRACTS	1
YOGURT	5	FRESH MEAT	1
COUGH AND COLD REMEDIES	4	PUDDING, DESSERTS-DAIRY	1
ICE CREAM, NOVELTIES	4	EGGS	1
BUTTER AND MARGARINE	4	FRESH PRODUCE	1
MILK	4	PERSONAL SOAP AND BATH ADDITIVES	1
ORAL HYGIENE	4	CHARCOAL, LOGS, ACCESSORIES	1
HAIR CARE	4	STATIONERY, SCHOOL SUPPLIES	1
FRESHENERS AND DEODORIZERS	4	TOBACCO & ACCESSORIES	1
BREAD AND BAKED GOODS	4	FIRST AID	1
SOUP	3	PASTA	1
GUM	3	VEGETABLES - CANNED	1
BREAKFAST FOODS-FROZEN	3	DOUGH PRODUCTS	1

Note: Four brands in our sample have products in two categories.

Table 9: Top Competitor Advertising Stock Elasticities

	Median	**Mean**	**% Brands % p $\geq$ 0.05**		**% p < 0.05**		**Percentiles**			
Specification					**>0**	**≤ 0**	**10%**	**25%**	**75%**	**90%**
Baseline Specification	0.0028	-0.0025	66.32	79.17	10.42	10.42	-0.1246	-0.0362	0.0348	0.1087
Border Strategy	0.0082	-0.0087	66.67	78.01	10.47	11.52	-0.1055	-0.0208	0.0437	0.0937

Note: The estimates are obtained assuming a carryover parameter $\delta = 0.9$. Standard errors are two-way clustered at the border-side level and the week level in the border strategy specification.

shape of the distributions is similar between the baseline and the border strategy approaches. A notable difference is that in the borders approach, a larger percentage of estimates is statistically different from zero. This does not appear to be due to a difference in the magnitudes of the estimated effects, but may be attributable to an increase in statistical power due to the border strategy's ability to explain more of the variation in the dependent variable.

D Functional Form of Ad Effects - Splines

In this section, we explore how sensitive our results are to our chosen functional form. In particular, the estimation of ROI near zero is heavily reliant on a steep curve generated by the $\log(1+A)$ function. We choose to estimate a cubic B-spline with knots placed at the 10th, 25th, 50th, 75th and 90th percentile of observed advertising stock for each brand, using the border-strategy with border-month and store fixed-effects. Below in Figure 17 we plot the predicted quantity as a function of ad stock for two brands, overlaying the spline predicted quantities and the $\log(1+A)$ function predicted quantities. In the left panel, we show this for Chobani and in the right panel, we show it for Gatorade. We choose to display these brands as they both had positive and reasonably precise estimates of the advertising effect, making it plausible that we could get some informative estimate of the shape of the curve. While only two brands are presented here, we provide spline estimation for all brands in our interactive online appendix. For both Chobani and Gatorade, the splines show a steep positive curve near to zero, which qualitatively matches the $\log(1+A)$ function. For Gatorade, despite some potential overfitting, the curve follows pretty closely with the $\log(1+A)$ function. For Chobani, it appears that after an initial effect of advertising near zero advertising stock, the effect is essentially zero. In both cases, the $\log(1+A)$ function appears to be a somewhat reasonable approximation. As a result, we do not believe our overall results are driven by the assumption of a $\log(1+A)$ function of advertising effectiveness.

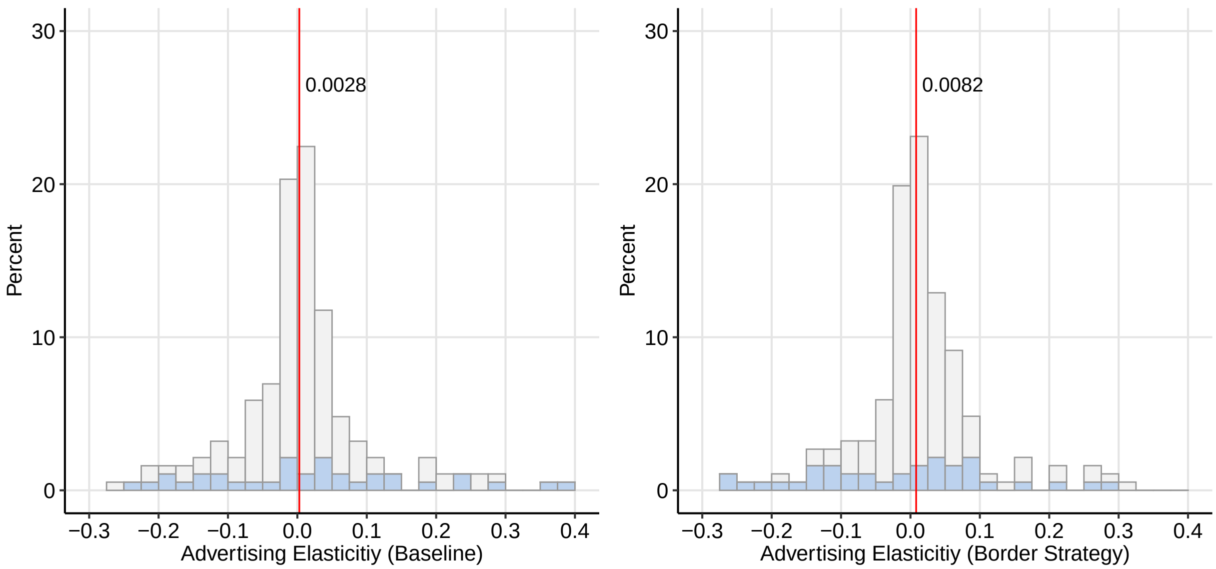

Figure 16: Competitor Advertising Stock Elasticities

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.

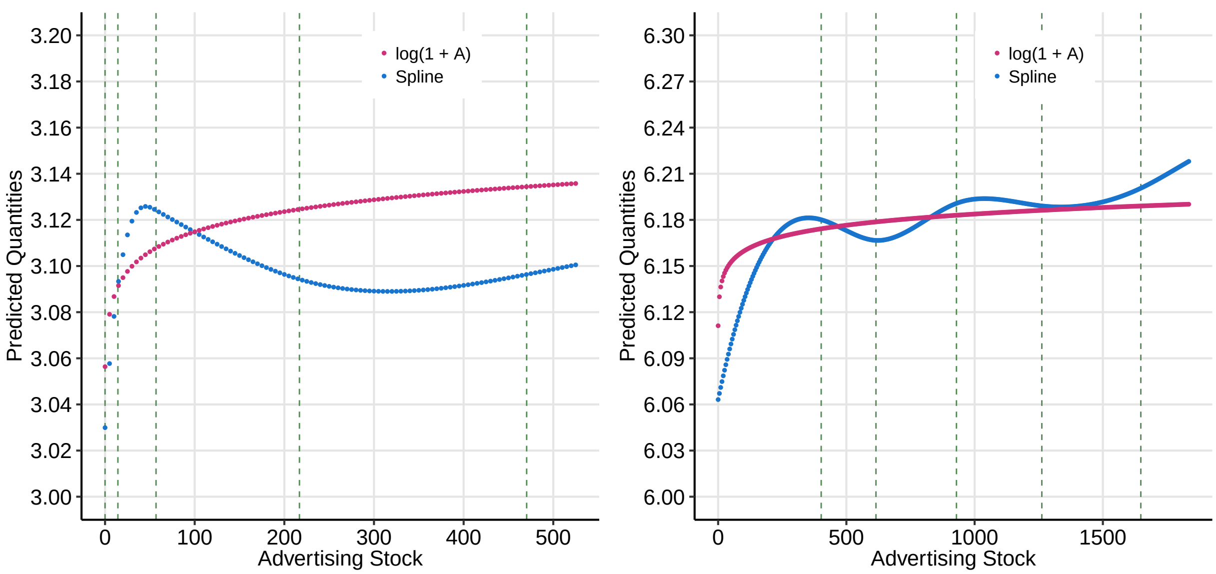

Figure 17: Predicted Quantity using Border-Strategy and Splines Estimation

Note: The left panel is for Chobani, while the right panel is for Gatorade. In both panels, we use the border-strategy model (red dots) with border-month, seasonal, and store fixed effects, and $\delta = 0.9$. For splines estimation (blue dots), we use cubic B-spline with knots placed at the 10th, 25th, 50th, 75th and 90th percentile of observed advertising stock (green dashed lines).

E Advertising Cost

In order to calculate a manufacturer's return-on-investment (ROI) from advertising, we need to estimate the cost of buying an ad GRP in DMA d in week t for each manufacturer. We use data on advertising expenditure, impressions, and audience size contained in the Nielsen Ad Intel data set for this purpose.

Expenditure Data

- For Cable, Network, and Syndicated TV, ads are purchased at the national level.
  - For network ads, Nielsen obtains expenditure data from the networks. If expenditure data is unavailable, Nielsen derives estimates of expenditures using supplementary industry data and proprietary models.
  - For cable ads, Nielsen's source for expenditure data is SQAD's NetCosts database. SQAD compiles occurrence-level data on actual purchases reported by contributing ad agencies. The measures SQAD shares with Nielsen are averages at the monthly-network-daypart level. The reported figures are believed to reflect the true weighting of upfront and scatter buys.
  - Expenditure data are originally at the {Month}-{Network}-{Daypart} level for national and cable ads. Ad Intel further prorates expenditure and records the data at the {AdTime}-{Network}-{Daypart}-{Program}-{Duration} level.
- For Spot TV, ads are purchased at the DMA level.
  - Nielsen estimates spot TV expenditures by blending cost-per-point data supplied by SQAD with Nielsen's local market ratings data. SQAD's cost-per-point data is based on actual spot television buys reported by contributing ad agencies.

Impression (Viewership) Data

- For Cable TV, impressions are recorded at the national level.
- For Network and Syndicated TV, while ads are purchased at the national level, the programs are broadcast at local TV stations and viewership is recorded at the DMA-level.
- For Spot TV, ads are broadcast at local TV stations and viewership is recorded at the DMA-level.

Universe Estimates

- Ad Intel provides the estimated total number of TV households at the national and market level. These universe estimates are updated yearly.

Build Advertising Cost

For each manufacturer, we do the following:

1. Merge expenditure with impressions for each ad occurrence;

2. Aggregate expenditure and impressions to the {National}-{Year} level. This involves adding up expenditure and impressions across media type, date, and markets;

   - We calculate advertising cost at the annual level since expenditure fluctuates across weeks. Hence, advertising cost for all weeks in the same year y remains the same.

3. Calculate national advertising cost per GRP in year y as:

$$\text{adcost per GRP}_{\text{national},y} = \frac{\sum_d \sum_{t \in y} \text{Expenditure}_{dt}}{100 \times \sum_d \sum_{t \in y} \text{Impression}_{dt} / \text{Universe}_{\text{national},y}}$$

4. Calculate DMA-level factor for national advertising cost using:

$$\text{Factor}_{dt} = \frac{\text{Universe}_{dt}}{\text{Universe}_{\text{national},t}}$$

5. Estimate advertising cost per GRP in DMA d in week t:

$$\text{adcost per GRP}_{dt} = \text{adcost per GRP}_{\text{national},y} \times \text{Factor}_{dt}$$

F Break-Even Ad Effects

In this section, we analyze how much larger TV ad effects would need to be in order for the observed level of advertising to be profitable. To this end, for different assumed values of margin factors and advertising costs, we compute the "break-even" ad elasticity for each brand. That is, we solve for the elasticity at which the observed level of weekly advertising would yield an ROI of 0. Elasticities larger than the break-even will produce positive profits, and elasticities smaller than the break-even would imply negative profits. In Section 8 we showed that the ROI of an average week's advertising can be quite different from the ROI of all observed advertising. Thus, we calculate the break-even ad effect separately for the average weekly ROI and the overall ROI. Using Chobani as an example, Figure 18 shows how the break-even ad effect varies as a function of the assumed margin factor m and the chosen ROI metric.

For each brand, we compare the break-even ad elasticity to the estimated ad elasticity and determine whether the break-even ad elasticity is within the confidence interval of our estimate. To summarize the results across brands, we calculate the ratio of the break-even ad effect to the estimated ad effect. The resulting multiplier tells us how much larger advertising effects would need to be in order for the observed level of advertising to be profitable. Figure 19 shows the distribution of this multiplier across brands for both the weekly break-even ROI and the overall

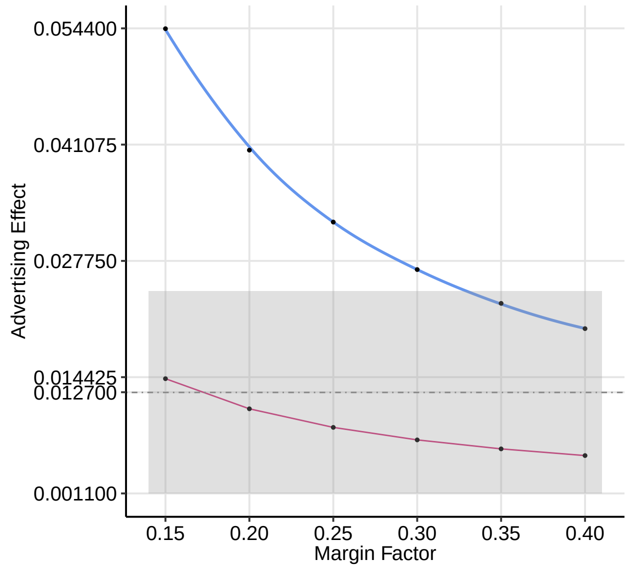

Figure 18: Break-even Advertising Effect (Chobani)

Note: The blue line is the break-even ad effect for the average weekly ROI, while the red line is for the overall ROI. For Chobani, our estimated advertising effect is about 0.0127 (gray dashed line) and the shaded area marks the 95% confidence interval.

break-even ROI. The left panel shows that for the median brand in our data, the estimated ad effect would need to be 2.5 times larger in order for the observed level of weekly advertising to be profitable (assuming a margin factor of $m = 0.3$). Further, for 61.4% of the brands in our data the weekly break-even ad elasticity is significantly larger than the estimated ad elasticity. For these brands, we reject the null hypothesis that the ROI of an average week's advertising is greater than or equal to 0. In contrast, the right panel of Figure 19 shows the results when considering the ROI of all observed advertising. For the median brand in the data, the break-even multiplier is less than 1, which is consistent with the fact that the ROI of all observed advertising is positive for the median brand in the data. Notably, amongst the 138 brands with a negative ROI of all observed ads, 67.4% have a confidence interval which covers the break-even ad effect, so we can't definitively conclude that the advertising for those brands is not profitable. Across the whole sample, we reject the null hypothesis that the ROI of all observed advertising is greater than or equal to 0 for only 15.8% of brands. Thus, despite the small ad effects reported in Section 6, the observed level of advertising could still be profitable for the majority of brands in the sample.

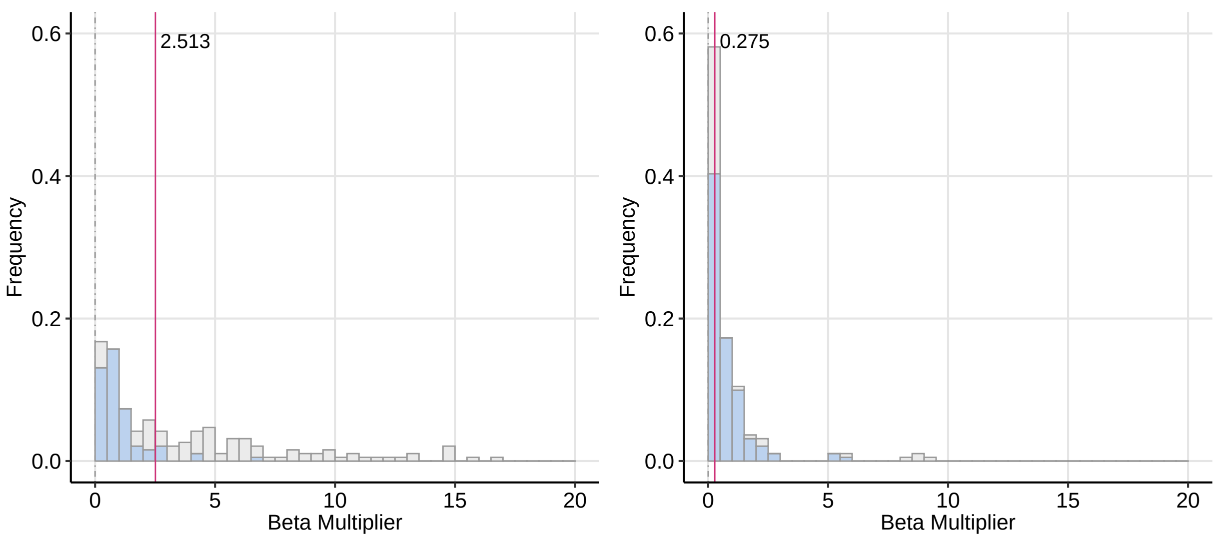

Figure 19: Ratio of Break-even Ad Effect to the Estimated Ad Effect

Note: The left panel shows the distribution of the ratio of the break-even ad effect to the estimated ad effect (multiplier) for weekly break-even ROIs. The right panel shows the multiplier for overall break-even ROI's. The histograms only include the 191 brands with a positive estimated ad effect.