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

February 19, 2019

Preliminary draft—comments are welcome

Abstract

We provide generalizable and robust results on the causal sales effect of TV advertising based on the distribution of advertising elasticities for a large number of products (brands) 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 anti-trust and public policy makers. A single case study cannot provide generalizable results, and hence the marketing 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 approach and assumptions made when constructing the data used in estimation from the raw sources. Second, as we aim to provide causal estimates, we document if the estimated effects are sensitive to the identification strategies that we use to claim causality based on observational data. Our results reveal substantially smaller effects of own-advertising 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 or median of the advertising effect distribution increases 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 Paul Ellickson, Wes Hartmann, and Carl Mela for helpful comments and suggestions. We also benefitted from the comments of seminar participants at Columbia University, Johns Hopkins University, and the 2018 Marketing Science Conference. 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 and the robustness of the results to assumptions on how to construct the data and to empirical strategies to obtain causal estimates of the advertising effect. 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. 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 ROI's even without conducting its own analysis for the products sold. Once specific advertising elasticity estimates for the firm's products are obtained, for example using an internal analysis conducted by the firm's data science team or using an external analysis by a marketing consulting firm, the prior serves as a benchmark to assess the credibility of these estimates. An incorrect, biased prior distribution will result in sub-optimal advertising decisions even when product-specific estimates are available. For example, if the prior overstates the true range of advertising elasticities, a specific elasticity estimate that is small compared to the prior will be inflated under rational Bayesian decision-making (i.e., the difference of the specific estimate relative to the prior mean will be shrunk). Even more detrimentally, both internal and external analysts may discard or misrepresent such estimates out of their own self-interest if the managers who commissioned the analysis judge small advertising elasticity estimates as a sign of incompetence or dislike estimates that deviate from the accepted view of advertising effectiveness for other, possibly career-driven reasons.

Hence, to be useful for decision-making, generalizations of effects need to be based on unbiased estimates that represent the population of interest. Most of the advertising research in empirical industrial organization and marketing during the last decades has used a case study approach that carefully examines specific industries. On their own, any single case study provides only limited information on the population distribution of advertising effects. The marketing literature in particular has recognized this limitation and provides meta-analyses that generalize previously published advertising elasticity estimates. Even a meta-analysis, however, will not provide the true distribution of advertising effects if the sample of case studies is not representative of the true population distribution. *Publication bias* is one key factor that may yield published results that are not representative. 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. In particular, editors or reviewers may reject advertising effect estimates that are not statistically significant or judged as small or "implausible," i.e. negative. Thus, false positives get published while true negatives get discarded. In anticipation of a rejection, a researcher may not complete or submit research with such results. This selection on the studies submitted to journals is frequently referred to as the *file drawer problem* (Franco et al. 2014).[1] Publication bias can arise without any ill intent among

[1] Andrews and Kasy (2017) show evidence that there is censoring of results in published studies, and they provide

the authors, reviewers, or editor. However, the anticipation of publication bias may also encourage the harmful practice of p-hacking, whereby researchers adjust their model specifications, covariates, or identifying assumptions until the results are acceptable for publication. Some prominent scholars have suggested that the standard for statistical significance be adjusted to $p < 0.005$ from $p < 0.05$ to address concerns about replicability and p-hacking (Benjamin et al. 2018). However, in the presence of publication bias, tightening the threshold for statistical significance may further weaken the generalizability of the published advertising effects because small effects are less likely to be published. Due to publication bias, a biased prior distribution of the estimated effects can be self-perpetuating. For example, if reviewers and editors mistakenly believe that advertising elasticities are almost always large (i.e. above a specific positive value) and statistically different from zero, they will likely reject null effects. Researchers will abandon projects that yield unexpected results, a problem that is exacerbated by the case study model which makes it relatively easy to abandon one for another case study. Publication bias will also be self-perpetuating in the industry if companies selectively publish white papers with results that are conducive to their business interests, or try to put restrictions on the results that can be published by academics. Thus, like evil begets evil, publication bias begets publication bias.

This paper provides generalizable results on the effect of television advertising on sales that do not suffer from publication bias. Advertising is likely to be particularly susceptible to publication bias because advertising effects tend to be small (e.g. Lodish et al. or Lewis and Rao 2015 in the context of digital advertising), and many well-known studies of advertising effects are now known to be under-powered. Marketing managers, however, tend to have strong prior beliefs that advertising is effective, and similarly many academics have prior beliefs that advertising must be effective because otherwise the large amount of advertising spending in the industry cannot be rationalized. Hence, for the reasons that we discussed above, advertising effect estimates that are small, negative, or not statistically different from zero are likely to be rejected in the publication process.

Our work avoids publication bias using clear research protocol for how the products in our sample are selected and by ensuring that 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 choose a large sample of 288 consumer packaged goods (CPG) for our analysis.

Robustness to the specific assumptions and choices made in our 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 the paper should generally be of interest to other researchers or analysts who use the Nielsen advertising (Ad Intel) data as a source for an advertising exposure measure. Second, our stated goal is to provide generalizable results on the *causal* effect of advertising sales. Hence, we need to ensure that the estimated advertising effects have a causal interpretation, and we analyze how robust the results are across different identification strategies. In general,

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.

advertising is not randomly assigned, and thus, in the presence of unmeasured confounders, the estimated advertising effects do not have a causal interpretation—this is the classic endogeneity problem in econometrics. Randomized controlled trials (A/B tests) such as the seminal IRI split cable experiments summarized in Lodish et al. (1995) establish causality, but conducting such field experiments is costly, and the split cable measurement technology is no longer available. Some recent papers have proposed instrumental variable (IV) strategies to infer causal advertising effects, such as the work on political advertising by Gordon and Hartmann (2013) that uses market-level advertising prices as instruments, or a recent paper by Sinkinson and Starc (Forthcoming) that proposes to use the timing of political campaigns as an exogenous shifter of brand advertising. However, these IV strategies have limitations due to standard concerns about instrument validity and weak instruments. For example, advertising prices are determined in equilibrium by the derived demand for advertising, and hence may fail the exclusion restriction. Furthermore, instrumental variables as a source of random variation are often case-specific and thus not useful to provide general results on advertising effectiveness. In this work we employ two identification strategies that are easily scalable across products. First, we consider threats to identification that arise if aggregate demand shocks and national advertising over time are correlated. We include a rich set of location and time fixed effects in our model to address these concerns. Given that the exact scheduling of advertising is at the discretion of television stations rather than advertisers, advertising variation net of these fixed effects may be argued to be as good as random. Second, we address the potential for correlation between local demand shocks and local advertising using an identification strategy that exploits a discontinuity in advertising at local television market (DMA) borders (Shapiro 2018). In this case, residual advertising variation is generated by borders, which are determined independent of demand for any particular consumer product, making advertising variation immediately at the border as good as random.

We show that the median of the distribution of the estimated long-run own-advertising elasticities is about 0.014 and varies a small amount depending on the exact specification and identification strategy. The corresponding mean is about 0.025. Between 65 and 75 percent of elasticities are not statistically different from zero or are negative. Compared to the prior literature, the magnitudes of the estimated elasticities are considerably smaller. Furthermore, few published papers in marketing or economics demonstrate null effects or negative elasticities. If the products with negative or statistically insignificant elasticities are excluded from the analysis, the mean and median of the advertising-elasticity distribution is substantially larger and more in line with the estimates in the extant literature. Overall, our results are consistent with a considerable degree of publication bias in the extant literature on advertising measurement.

To what extent are our results an artifact of statistical noise? Restricting our attention to estimates with 50% ex ante power to detect an advertising elasticity of 0.05 yields similar results on the median, mean and frequency of null and negative results, but significantly attenuates the extreme ends of the distribution. This suggests that the frequency of null results is not simply an artifact of noise, as the frequency is similar amongst "precisely estimated" estimates. However, the fact that the very high advertising elasticities go away when focusing on estimates with 50% power to detect an elasticity of 0.05 suggests that the largest estimates may be an artifact of noise.

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. In the sections that follow, we first discuss how our work relates to the existing body of work on advertising effects in section 2. In section 3, we describe the data used in our empirical analysis. Section 4 introduces our research design, and Section 5 presents the estimation results. Section 6 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 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. Lodish et al. (1995) analyzes a series of 389 household-level split cable TV advertising experiments that were completed between June 1982 and Dec 1988. They estimate an average ad elasticity of 0.13 across all products. When they split the data by new versus established products, they find that these positive ad effects are driven by new products that have an average ad elasticity of 0.26. The authors do not find a significant advertising effect for established products. In a follow-up paper, Hu et al. (2007) analyze the results of 241 TV ad tests carried out between 1989 and 2003. Contrary to the findings of Lodish et al. (1995), the authors find positive and significant effects of advertising weight tests for established products, and they find some evidence that advertising elasticities were increasing over time. The product or company-level results have low statistical power, so analysis is restricted to overall averages among a few small groups.

This work is very much in line with our motivation, and we seek to update and improve upon the results of Lodish et al. (1995) in a number of ways. First, the results in Lodish et al. (1995) are for small, specific markets. As a result, this study had low statistical power. 'Significance' for each product is reported as rejecting a one-tailed test at 80% confidence ($p<0.4$). The overall averages are jointly significant when pooled. Compared to Lodish et al. (1995), our study covers a longer time series and many more markets, through which we obtain considerably better statistical power. Additionally, the results of Lodish et al. (1995) may not apply decades later, with more and more forms of media competing for the attention of consumers. Finally, although we are not able to run large scale experiments, we can evaluate advertising effects using methods and existing datasets that ad agencies and product manufacturers may already have or can easily obtain.

Putting the issue of statistical power aside, it is interesting to note that all of these multi-product studies that use single-source data note a high prevalence of null effects, which are not represented in the meta-analyses of published studies discussed above. We believe this observation is suggestive of the existence of publication bias in the measurement of ad effects. We incorporate data from the entire United States and across many years, which helps us achieve better statistical power to highlight this lesson.

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 a few 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 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. 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.

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 carry-over. 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:

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

$\boldsymbol{a}_{d(s)t}$ is the flow of advertising in DMA $d(s)$ in week t, and δ is the advertising carry-over 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.[2] The second own advertising variable captures advertising messages for affiliated products

[2]We acknowledge that 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.

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 more thoroughly in Section 4. We also include advertising for up to three competitors in the model.[3]

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 ad stock elasticity is a form of long-run elasticity that represents the percent change in current period quantity sales that would result from having increased current and past advertising by 1%. A forward-looking manager might wish to forecast with the ad stock elasticity to answer the managerial question, "if I increase my advertising by 1% every week over the next year, how much higher will my sales be next year?" Appendix A discusses alternative short run and long run elasticity metrics that can be computed from this model.

3.2 Identification Strategy

The main challenge in estimating the model (1) is that advertising is not randomly assigned. Firms may target their advertising in DMAs and periods where they believe that advertising will be most effective. Such targeting strategies may involve advertising 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 jointly correlated with advertising and sales. Such targeting strategies or omitted factors lead to a spurious relationship between advertising and sales. Hence, to ensure that we estimate the causal relationship between advertising and 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 to which advertising practitioners could reasonably respond. Second, we use the quasi-random variation in advertising across the borders of television markets. We refer to the first approach as our *baseline* specification and the second as the *border strategy.*

The first approach, or 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 month fixed effects. To control for seasonality that occurs within the months in a given year, we employ week-of-year fixed

[3] Not all brands are sold at all stores. For each brand we determine the number of competitors that are included in the model based on the fraction of observations that would be lost by adding an additional competitor.

effects.[4] Furthermore, in some specifications we also include indicators for feature advertising and in-store display advertising. As already discussed, the model also includes prices and competitive advertising, 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 ad-buying process, whereby advertising agencies buy ad slots often many months in advance. The advertising agency 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 this approach is obtained by adding controls and fixed effects to equation (1):

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

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.[5]

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 inclusion of the fixed effects and controls discussed above 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 DMA's. This research design was first used in Shapiro (2018) to study the effects of television advertising on antidepressant demand, and has now been used in Tuchman (2018) to study e-cigarette advertising, as well as in Spenkuch and Toniatti (2018) and **?** 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 to very local areas. Intuitively, this approach simulates an experiment with two treatment groups.

[4]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.

[5]Other controls include variables indicating whether a product was on feature or display. These variables are only recorded for a sub-set of the stores in our data, so our preferred specifications omit these variables. See (https://advertising-effects.chicagobooth.edu/) for results that include these controls.

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. As an example, Figure 1 depicts the Louisville, KY and Lexington, KY DMAs and outlines the border counties between these neighboring DMAs. 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. Second, we adapt the model specified in equation 3 to include a time fixed-effect, $\gamma_{\mathcal{B}(s,t)}$, that is 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/).[6] As before, we use store fixed effects to control for persistent local factors related to demand. While focusing on borders significantly decreases the number of observations in each regression, the net effect of focusing on the borders on statistical power is ambiguous. Using the border strategy affects statistical power in three ways that do not work in the same direction. First, the border-specific time fixed effects work like adding additional control variables. This reduces residual variance in the dependent variable, which, all else equal, increases statistical power.[7] However, the additional fixed effects also reduce the residual variance in advertising stock, which all else equal, reduces statistical power. Finally, focusing on the border reduces the number of observations, which reduces statistical power, all else equal. The net effect depends upon the relative magnitudes of these forces.

The approach of focusing on cross-border variation will be useful if two conditions hold. First, the side of the border on which a household lives must be conditionally independent of omitted factors which influence changes in demand. We argue that it is unlikely for a household to decide where to live based on changes in expected advertising exposure for a set of relevant brands. This assumption is not directly testable, but Shapiro (2018) shows that differences in advertising across borders do not predict observable demographic variables. Second, there must be sufficient variation in advertising net of the fixed effects included in the model. Said differently, there needs to be significant cross-border differences in advertising, and these differences need to vary over time.

When estimating the regression models, standard errors are clustered to account for correlation in the error terms. The clustering varies somewhat by specification, as different specifications induce

[6]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.

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

different residual variation in advertising, which induces different correlation structures across error terms. In our main baseline specifications where the time fixed effects are not at the same periodicity of the data, we two-way cluster the standard errors by DMA and week. This accounts for (1) the correlation in error terms induced by repeated observations over time and for (2) the correlation in error terms induced by correlation in the treatment (residual of the controls and fixed-effects in the model). 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 main border strategy specification, we two-way cluster standard errors by border-side and by week.

4 Data

Data on both purchase volume and advertising intensity is necessary in order to tease out the effect of advertising on sales. We construct such a dataset by merging market-level TV advertising data with retail sales data at the brand level. These datasets 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 AdIntel and RMS scanner data. Merging these two large datasets is difficult. Advertised brands do not match up perfectly with RMS scanner UPC codes. Often the advertised brand is either more or less specific than the brand associated with a UPC code. An advertised brand typically intends to advertise for multiple UPC codes. Next, advertising data comes from a number of measurement devices at the local and national levels that must be reconciled in order to produce a coherent television timeline. Our data appendix shows, in detail, how to re-create our data build and merge.

Retail Sales Data

The AC Nielsen database includes weekly store sales data reporting prices and quantity sold at the UPC-level. During the period of our data, 2010–2014, 41,309 stores are tracked in the data and 575,617 distinct UPCs are sold by 1,421 brands. We focus our analysis on the top 500 brands in terms of dollar sales. In our case, we define a brand as all forms of the same consumable end product, as indicated by a brand code field in the RMS data. That is, Coca-Cola Classic would include 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 the UPC-level data, calculating total volume sold in equivalent units and average price per equivalent unit. After dropping some small stores and stores that are located in counties that switch DMAs over time, we are left with 12,671 stores in our final estimation sample.

Advertising Data

Product-level television advertising data from 2010–2014 comes from the Nielsen Ad-Intel database. The advertising data 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 eye-balls, that viewed each ad. In the top 25 DMAs, impressions are measured by set top box recording devices. For DMAs below the top 25, impressions are measured using diaries filled out by Nielsen households. For Cable ads, which are aired nationally, viewership data is only available at the national level. Spot ads are bought locally, and viewership measures are recorded separately for each DMA. Network and Syndicated ads are recorded in national occurrence files that can be matched with local measures of viewership. Thus, in our data, variation in a brand's ad viewership across markets can come from both variation in occurrences (one market has more Spot ad occurrences than another), as well as variation in viewership across markets (a Network or Syndicated ad aired in both markets, but viewership was larger in one market compared to the other). From this impression data, we calculate gross rating points (GRPs) for each ad occurrence. GRPs are a frequently used measure of advertising intensity, calculated as exposures per capita times 100. We aggregate the occurrence-level ad data up to the brand-week level by summing GRPs across all occurrences for a brand in a given week.

Combining Datasets

We merge the advertising and sales datasets at the store-brand-week level. Our merging procedure warrants some discussion because product sales are recorded at the granular UPC level while advertising is often categorized at a higher product or brand level. Furthermore, the brand variables in the Ad-Intel and RMS datasets 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 Coke 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, these tiers should have a positive effect on sales of the focal product. Alternatively, the sign of the effect of tiers 2 and 4 matches is ambiguous. In particular, in tiers 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. Seeing 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 in the data. Clearly, the exact matches between the advertising and sales datasets are thus very important to get right.

This was a manual process that required us to evaluate the brand descriptions in each dataset and determine the nature of the relationship between brands. The initial merge was carried out by two RA's, and any disagreements were resolved by the authors.

We were able to match 288 of the top 500 brands in the RMS data to TV advertising records in the AdIntel database. Although focusing on the top 500 brands in terms of dollar sales necessarily means we are selecting relatively established products, we find that even within this set of brands, there is substantial variation in the intensity at which these large brands advertise on TV. Figure 2 documents the variation across brands in average weekly GRPs, $\bar{a}_j = \frac{1}{M*T}\sum_{m=1}^{M}\sum_{t=1}^{T} a_{jmt}$.

Descriptive Statistics

Before introducing our empirical model, we first document the extent to which there is temporal and cross-sectional variation in advertising for each brand. This is important because our analysis relies on the extent to which there is variation in advertising intensity both across markets and over time. To illustrate this variation, for each brand we regress weekly DMA-level advertising GRPs on a set of DMA, month, and week-of-year FEs. We calculate the standard deviation of the residuals, which tells us the amount of residual variation in advertising that is not explained by these fixed effects. Finally, we calculate the ratio of the residual variation to the average DMA-level weekly GRPs for that brand. This "coefficient of variation" serves as a parsimonious way of quantifying the amount of variation in advertising that is left net of the market and time FEs. Figure 3 presents a histogram of this measure across brands for both advertising flow (current week GRP) and advertising stock assuming a carry-over parameter of $\delta = 0.9$. The median brand's coefficient of variation is 0.41 for advertising flow. In other words, the standard deviation of the residuals is 0.4 times the size of the average weekly advertising for that brand. This tells us that we observe reasonably large deviations from average advertising levels for most brands, which will help us identify the advertising effect of interest. Notably, this variability is reduced considerably when looking at advertising stocks. This happens for two reasons. First, there is considerable week over week variability in which markets have positive shocks to advertising. As a result, aggregating over more weeks in the building of a stock cancels out much of that variability. Second, the coefficient of variation measure mechanically gets smaller when viewed as a stock, since the denominator is larger.

Recall that our approach leveraging variation at the borders of television markets requires that there be sufficient variation in advertising net of the fixed effects in the model to pin down advertising effects. We report two analyses below that demonstrate that this second condition is satisfied in our 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.[8] The average absolute difference is about 14 GRPs, which is reasonably large relative to the average weekly GRPs documented in Figure 2. Second, in Figure 5, we show the distribution of brand-level residual advertising flows and advertising stocks with carry-over parameter δ set equal to 0.9. The residuals are obtained from a regression of advertising flow (and stock) GRPs on the fixed effects and control variables in our preferred specification. These include border-month fixed effects, week-of-year fixed effects, store fixed effects, own and competitor prices, and competitor advertising stock values. Both

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

distributions reveal a considerable degree of variation to estimate the advertising effects using the border strategy. Notably, the variation in residuals from the border-strategy model that we show in Figure 5 is quite similar to the variation in residuals from the more parsimonious baseline model reported in Figure 3. The residuals of advertising stock show a considerably smaller coefficient of variation than the residuals of advertising flow, for exactly the same reasons as before: i) there is variability week-over-week in which side of the border has a positive shock to advertising, which is cancelled out when aggregating to a stock, and ii) the denominator is mechanically larger in the stock conception of this coefficient of variation.

5 Results

5.1 Own Advertising Elasticities

We first discuss the estimation results for the two main model specifications that are given in equations (3) and (4). The baseline model in equation (3) controls for confounding factors using a rich set of fixed effects, whereas equation (4) implements the border strategy. To assess how the results are affected by confounding factors, we also present the results for the most parsimonious specification, equation (1), which includes no fixed effects at all. We refer to this specification as the *naive* model. The results discussed here are only a subset of all the models we have estimated. Please see (https://advertising-effects.chicagobooth.edu/) to explore the sensitivity of the results to different modeling choices.

5.1.1 Long Run Advertising Elasticities: $\delta = 0.9$

Here, we present the main results. We allow for long-lived advertising effects. Specifically, we incorporate an advertising stock with geometric decay and set the carry-over parameter to $\delta = 0.9$. That is, 90 percent of current advertising carries over to the advertising stock next week. Past work has shown evidence of long-lasting advertising effects. 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. We use this estimate of advertising decay as a starting point for our analysis.

Summary statistics for all three model specifications are provided in Table 1. When no fixed effects are included, the median long-run advertising elasticity is 0.0305 and the mean is 0.0417. About 19.4% of estimates are negative and significant. When we restrict the sample to products with estimated advertising effects that are positive and statistically significant, we drop 58 percent of all observations. The median of this selected distribution is 0.0846 and the mean is 0.1384. The mean is notably still smaller than the mean effect of 0.23 reported by Sethuraman et al. (2011), but the median is quite close to Sethuraman et al. (2011)'s median long-run elasticity of 0.10. However, when we account for potential confounding factors using a rich set of fixed effects in the baseline model, the median shrinks considerably to 0.0140 and the mean to 0.0239. Negative and significant results are reduced to 7.6% of the estimates. Focusing only on the positive and statistically significant results, we discard 73.3 percent of all observations. The corresponding median of the advertising elasticities is 0.0655 and the mean is 0.0967. The results using the border strategy – a median of

0.0144 and a mean 0.0257, with 7.3% negative and significant estimates in the full sample of products – are similar to the results with the baseline model. Focusing only on the products with positive and significant estimates, we drop 74.7 percent of all observations. In the restricted sample, the median is 0.0716 and the mean is 0.0990.

The main take-away from these results is that, on average, the estimated long-run advertising effects in our full sample are small compared to the reported estimates in the extant literature. We find that the estimated elasticities are more similar to the results in the literature if we are less careful to control for confounding factors that may bias the results or if we restrict the sample of products to those with positive and statistically significant estimates. In the latter case, we effectively select a population of results that are more likely to be publishable given a publication process that rejects "implausible," small, and statistically insignificant results. This finding highlights the short-comings of relying on meta-analyses of published estimates to learn about generalizable TV advertising effects.

We display the results in Figure 6. The left panel shows the histogram of advertising elasticities from the baseline model with store, month and week-of-year fixed effects. The right panel displays the results when we employ the border strategy. The two distributions are similar. Furthermore, the differences in the estimated advertising elasticities between the baseline and border strategies are quite small. Figure 7 displays the distribution of these differences among all brands in the sample. Most of the differences are not statistically different from zero, although there are also several exceptions. The statistically significant differences could be explained by, for example, confounding factors that are addressed by the use of the borders approach, by differences in local effects of advertising at the borders versus the rest of the DMA, or by different degrees of measurement error. The exact source of these differences is not identifiable from our data.

As expected, the magnitude of the mean and median coefficients change when we change the assumed decay 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 assumed δ between zero and 1. Furthermore, while there is variance in the magnitude of mean and median bias when we only consider positive and significant estimates across assumed δ, it is consistently large. Please see Table 2 for the relevant summary results for every δ from zero to 1 for the border strategy.

5.1.2 Results Using Estimated Advertising Carry-Over Factor

We allow for an additional degree of freedom in the estimated models and estimate the carry-over parameter, δ, using a grid search. We use a grid from 0 to 1 in increments of 0.01. For each point in the δ-grid, we calculate the implied advertising stock using equation (2) and then estimate the remaining model parameters via OLS. For each brand, the estimated $\hat{\delta}$ is the δ that minimizes the predicted root 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 level of the advertising carry-over. 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 should be uniformly distributed on $(-\infty, \infty)$. However, since we impose the constraint that $\delta \in \{0, 0.01, \ldots, 0.99, 1\}$, the δ that fits the data best will typically be on the bounds of the grid, $\delta = 0$ or $\delta = 1$. 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. For these reasons, we expect that the estimates will have more noise compared to the previous approach where we set the carry-over factor δ to a given value.

Table 3 summarizes the results when we estimate δ (the table does not include the results for the naive model without fixed effects). Using the baseline specification, model (3), the median advertising elasticity is 0.0090 and the mean is 0.0112. When we only keep the positive and significant estimates, we lose 63.2 percent of the product observations. In the restricted sample, the median of the estimated elasticities is 0.0540 and the mean is 0.1081. Using the border strategy, the median advertising elasticity is 0.0114 and the mean is 0.0268. If we only consider the positive and statistically significant estimates we drop 61.1 percent of the observations. Then the median elasticity is 0.0354 and the mean is 0.0823.

Figure 8 shows the distribution of the advertising effect estimates, separately for the baseline specification and the specification that implements the border strategy. We find that when we estimate δ, the distribution of the estimated advertising elasticities exhibits a larger spread compared to the case when we set $\delta = 0.9$. Both the number of negative and significant estimates and positive and significant estimates increase. The 90th percentile is larger and the 10th percentile is more negative. Just comparing the border strategy results, when estimating the best δ, the 90th percentile is 0.1441 but when $\delta = 0.9$ is assumed it is 0.1. The 10th percentile elasticity is -0.0351 when estimating the best δ but is -0.0307 when assuming $\delta = 0.9$.

Figure 9 shows the histogram of estimated δ. Highlighted are the estimates of δ for the brands that have advertising elasticity statistically significant and greater than 0.01 when $\delta = 0.9$ is assumed. For the brands that do not exhibit relatively large and precise estimates under the assumed δ, we see considerable bunching at the boundaries, likely driven by the model having a difficult time separately identifying β and δ.

5.1.3 Statistical Power and Selection

We have so far restricted any discussion of statistical power to whether or not estimates are statistically significant at $p<0.05$. In this section, we discuss two remaining issues regarding statistical power. First, we want to establish whether insignificant estimates are true "null" effects or simply noisy and whether large estimates are truly large or simply noisy. We explore this by identifying the brands that ex ante have sufficient power to detect a reasonable effect size. Second, we evaluate how sensitive our results are to the $p<0.05$ significance threshold. We explore this sensitivity analysis for a couple of reasons. First, we hypothesize that publication requirements that require authors to produce process evidence of advertising "mechanisms" may select on estimates that have a higher degree of precision than $p<0.05$. In addition, several prominent researchers have recently suggested that the publishable threshold for statistical significance should be moved to $p<0.005$ (Benjamin et al. (2018)). We discuss both of these issues below, focusing on the border strategy results with carry-over parameter $\delta = 0.9$.

First, we focus attention on brands where there is at least 50% ex ante power to detect an advertising elasticity of 0.05 at the 5% level.[9] This condition is met for 157 of the 288 brands, with results presented in Table 4. For this set of brands, the median elasticity is 0.0075 and the mean is 0.0083. Notably, this distribution is compressed, with zero advertising elasticities larger than 0.1 and zero advertising elasticities smaller than -0.1. The 90th percentile of the distribution is 0.0353 and 9.6% of estimates are negative and significant. In terms of "zeros," 68.8% of elasticities are not statistically significant, which is comparable with the unrestricted set of advertising elasticities. This suggests that the set of null results *does not* simply indicate noise, but a significant number of true null effects. In terms of "large" effects, restricting to the set of brands with 50% power to detect an advertising elasticity of 0.05 completely eliminates the set of very large estimated advertising elasticities in the full sample. This suggests that the set of large effects *does* indicate a significant degree of noise rather than a truly large advertising effect. As a result, trying to explain what variables predict a large effect of advertising estimated in our data is unlikely to be a fruitful endeavor.

Next, we focus attention on brands that are not only positive and significant at $p<0.05$, but are also positive and significant at $p<0.005$. This tighter requirement of $p<0.005$ selects on even larger effect sizes, and therefore exacerbates publication bias and our ability to draw generalizable results from the distribution of "publishable" ad effects. The results are in Table 4. Only 32 of the 288 brands (11%) are positive and $p<0.005$. The median ad elasticity among these brands is 0.0768 and the mean is 0.1118, or about 7% and 13% larger than the set that is positive and $p<0.05$, respectively.

5.1.4 Discussion

We draw a few key take-aways from this analysis. First, the discrepancy between the models with and without fixed effects shows that advertising is correlated with the unobserved component of demand and illustrates the importance of accounting for such potential correlations in a flexible way. However, the similarity between the results obtained with the baseline and the border specifications indicates that confounds due to the coordination between local advertising and local demand shocks are not of first-order concern when estimating TV advertising effects for the brands covered in our sample. Second, the majority of our estimated short-run ad elasticities are smaller than the mean elasticity of 0.13 reported in Lodish et al. (1995) and the mean elasticity of 0.12 reported in Sethuraman et al. (2011). That said, the median short run elasticity in Sethuraman et al. (2011) is 0.05, which suggests that the distribution is highly skewed. Third, the estimated advertising effects are not statistically different from zero for a large percentage of the brands in our sample, and these null results do not seem to be due to a lack of statistical power. Finally, our analysis shows that only publishing positive and statistically significant results could substantially bias our understanding of the distribution of advertising effects across brands. Such bias could further explain the difference between the magnitude of our estimates and the extant meta-analyses of published results.

Overall, the main point about publication bias presented here is robust to i) using the border approach versus including a rich set of fixed-effects in the baseline model, ii) considering short-

[9] 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 (**?**).

run versus long-run advertising effects, and iii) assuming versus estimating advertising carry-over. Analyzing rich TV advertising and store sales data with flexible model specifications, we estimate positive and significant ad effects for only a fraction of the brands in our dataset. Further, only considering positive and significant advertising elasticities biases our estimates of the median and mean advertising elasticities by a considerable amount.

5.2 Cross Advertising Elasticities

In the section above, we reported the estimated own-advertising elasticities for the brands in our sample. All model specifications control for competitor advertising in the product category, and we now discuss the estimated competitive advertising effects. While theory predicts that own-ad 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 (Forthcoming)). 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 5 shows summary statistics for the estimated cross-advertising elasticities corresponding to the baseline and borders model specifications in equations (3) and (4), and Figure 10 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 1 and 3 competitors (see footnote 3). In Table 5 and Figure 10 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 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.

6 Discussion

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.

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.

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.

7 Conclusion

In this paper, we present generalizable estimates of the advertising elasticities for 288 large, national brands. To ensure robustness of the results we consider a variety of specification choices and identification strategies. We document that the median of the distribution of the 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. The magnitudes of the estimated elasticities are considerably smaller than what has been found in prior literature. The discrepancy with respect to the extant literature is consistent with publication bias. In particular, the estimated advertising effects are either negative or not statistically different from zero for more than half of all brands in our data. If these brands are excluded from the analysis, the mean and median of the advertising-elasticity distribution is substantially larger and more in line with the estimates in the extant literature.

References

Andrews, I. and M. Kasy (2017): "Identification of and correction for publication bias," Tech. rep., NBER Working Paper 23298.

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.

Benjamin, D. J., J. O. Berger, M. Johannesson, B. A. Nosek, E.-J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer, et al. (2018): "Redefine statistical significance," *Nature Human Behaviour*, 2, 6.

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. H. 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*.

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.

Hu, Y., L. Lodish, and A. Krieger (2007): "An Analysis of Real World TV Advertising Tests: A 15-Year Update," *Journal of Advertising Research*, 47, 341–353.

Johnson, G., R. Lewis, and E. Nubbemeyer (2016): "The Online Display Ad Effectiveness Funnel & Carry-Over: A Meta-Study of Ghost Ad Experiments," *Working Paper*.

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*, 1941–1973.

Lodish, L., 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.

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.

Sinkinson, M. and A. Starc (Forthcoming): "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals," *Review of Economic Studies*.

Spenkuch, J. L. and D. Toniatti (2018): "Political advertising and election outcomes," *Quarterly Journal of Economics, forthcoming.*

Tuchman, A. E. (2018): "Advertising and Demand for Addictive Goods: The Effects of E-Cigarette Advertising," Working Paper.

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

$$\frac{dQ_t}{da}\frac{a}{Q_t} = \beta\frac{\rho a}{1+\rho a}. \tag{5}$$

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 $t, \dots, t+L$:

$$\left(\frac{\partial}{\partial a_t}\sum_{\tau=t}^{t+L} Q_\tau\right)\frac{a_t}{Q_t} = \beta\frac{\rho a}{1+\rho a}. \tag{6}$$

The effect of permanent percentage increase in advertising (5) is equivalent to the total, long-run increase in demand (6). 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.

To study the long-run effect of the advertising stock, we consider specifications where we fix

$\delta > 0$ prior to running the analysis, as well as specifications where we jointly estimate δ using a grid search.

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 datasets.
   (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 {AdIntel 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 "Sweep Months": February, May, July, and November. Therefore, we need to impute the impressions for the non-sweep 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 datasets 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 {AdIntel 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 {AdIntel 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 dataset 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.

Figure 1: Border Counties Between the Louisville and Lexington DMAs

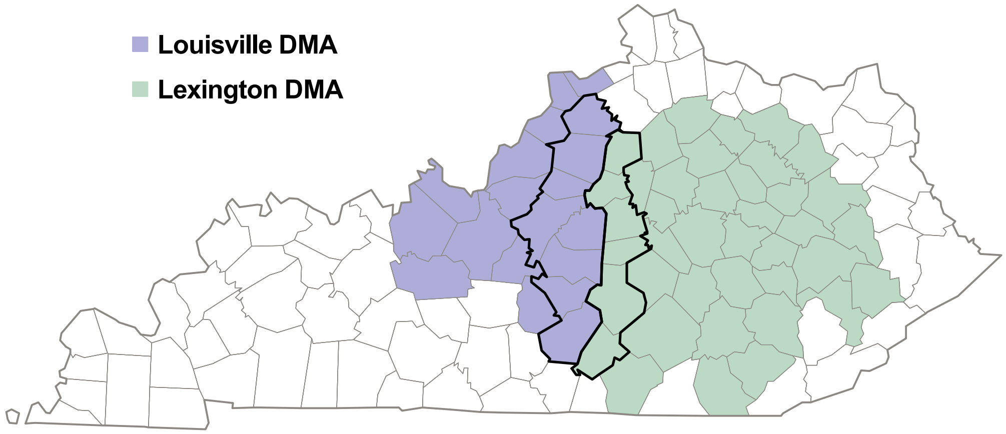


Figure 2: Variation in Advertising Intensity Across Brands

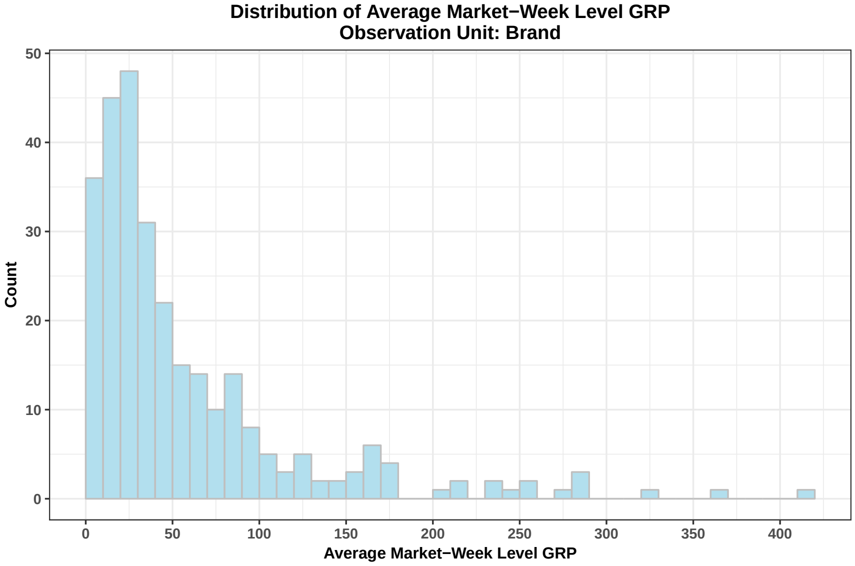


Figure 3: Residual Variation in Advertising Net of Fixed Effects

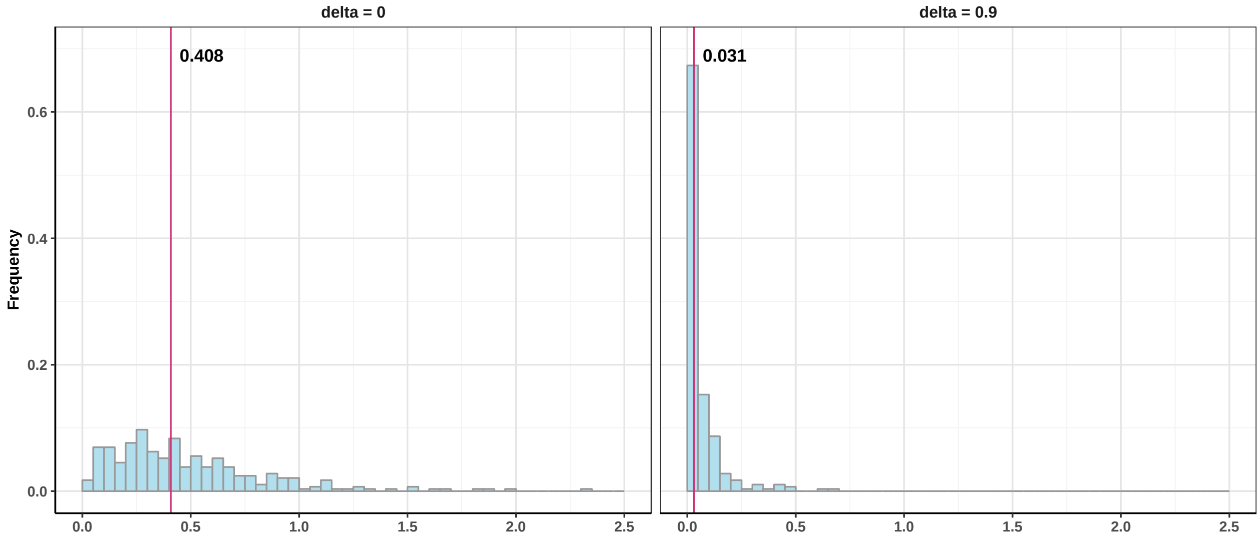


Figure 4: Brand Average of Weekly Absolute Difference in GRP Across Borders

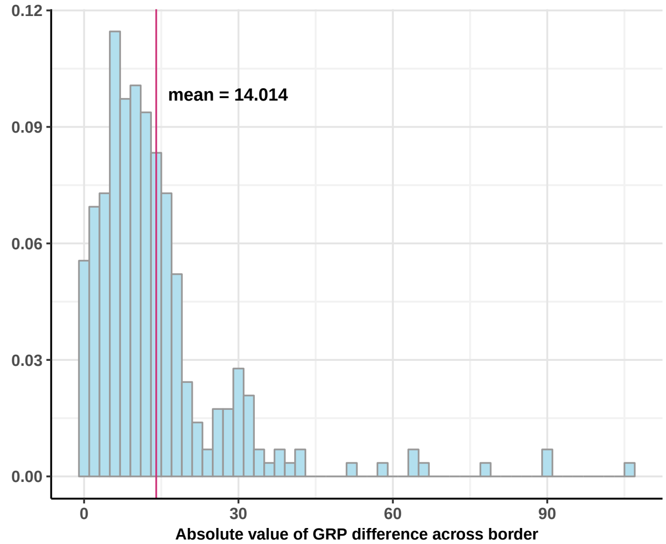


Figure 5: Residual Variation in Advertising Net of Border Fixed Effects

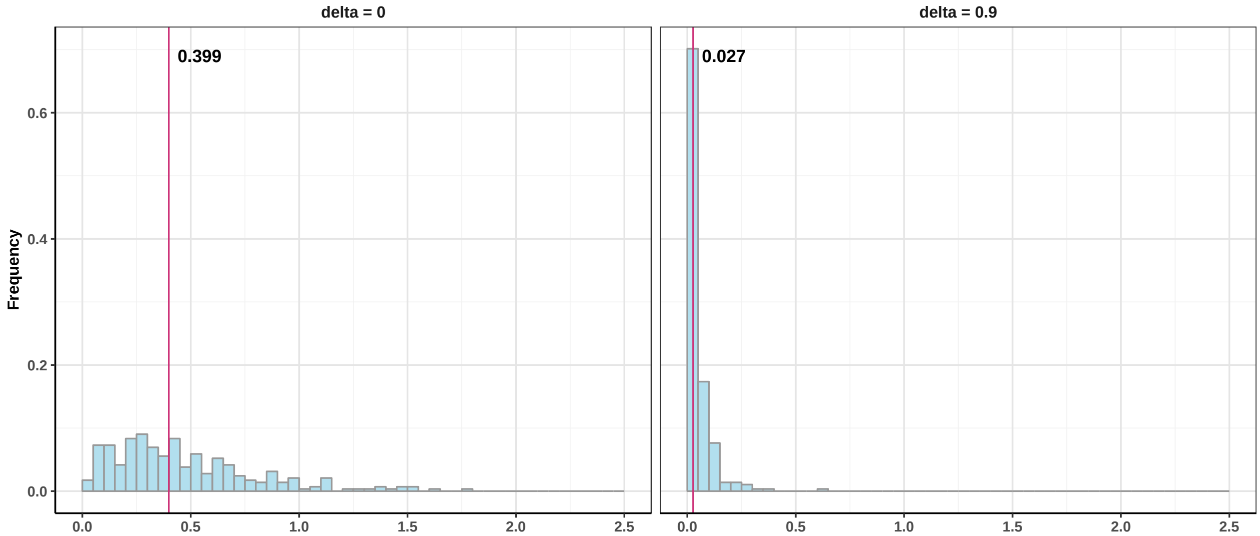
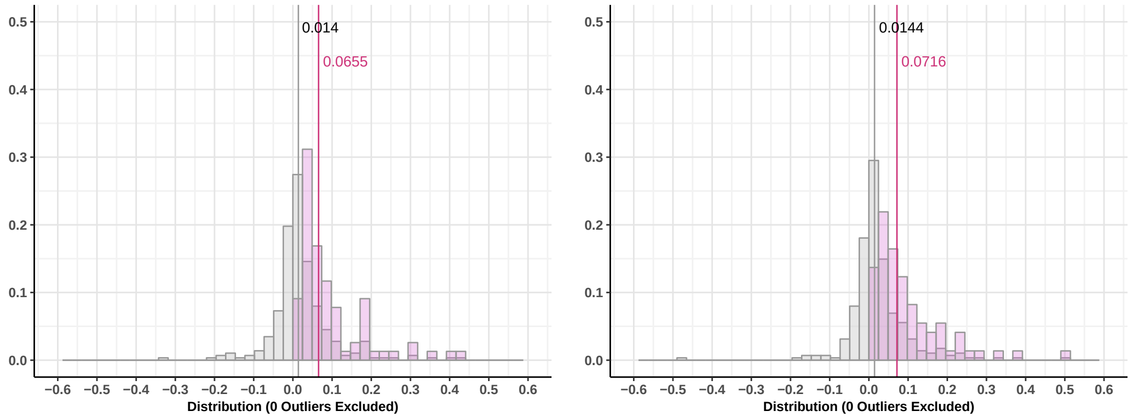


Figure 6: Ad Elasticities in Ad Stock Model with Carry-Over Fixed at $\delta = 0.9$



Colored part is the distribution of positive and siginifcant estimates, normalized to have the same density.

Colored part is the distribution of positive and siginifcant estimates, normalized to have the same density.

Figure 7: The Distribution of the Difference in Advertising Elasticities Border vs. Non-Border

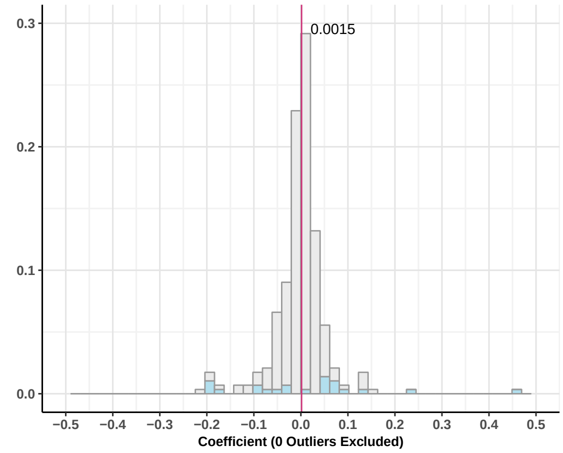


Figure 8: Ad Elasticities in Ad Stock Model with Estimated Rate of Ad Carry-Over

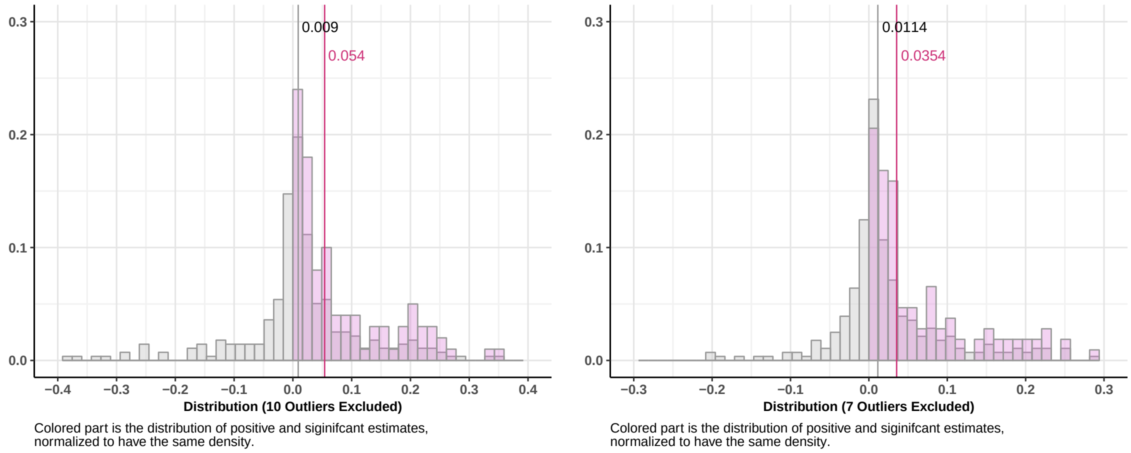


Figure 9: Distribution of Estimated Best δ

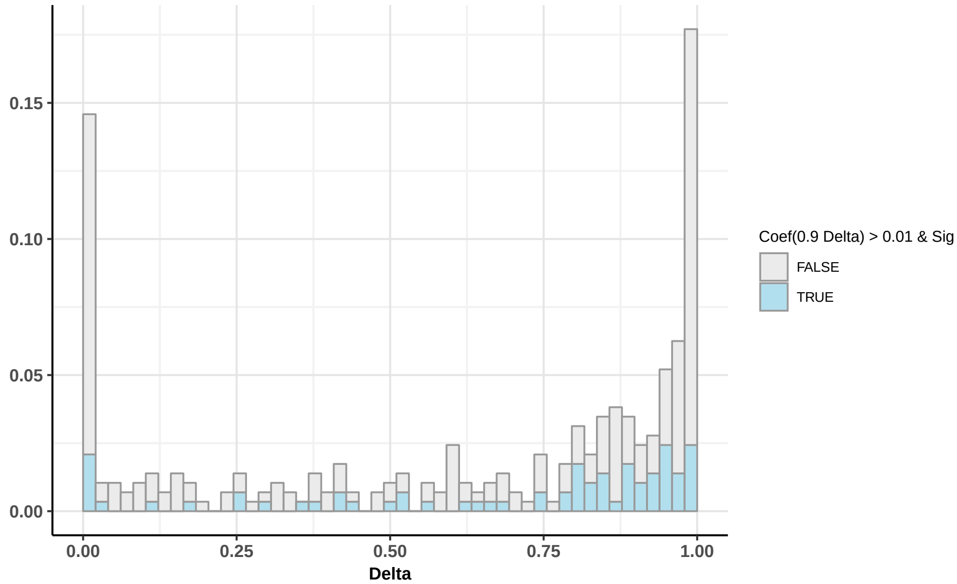


Figure 10: Competitor Ad Elasticities in Ad Stock Model with $\delta = 0.9$

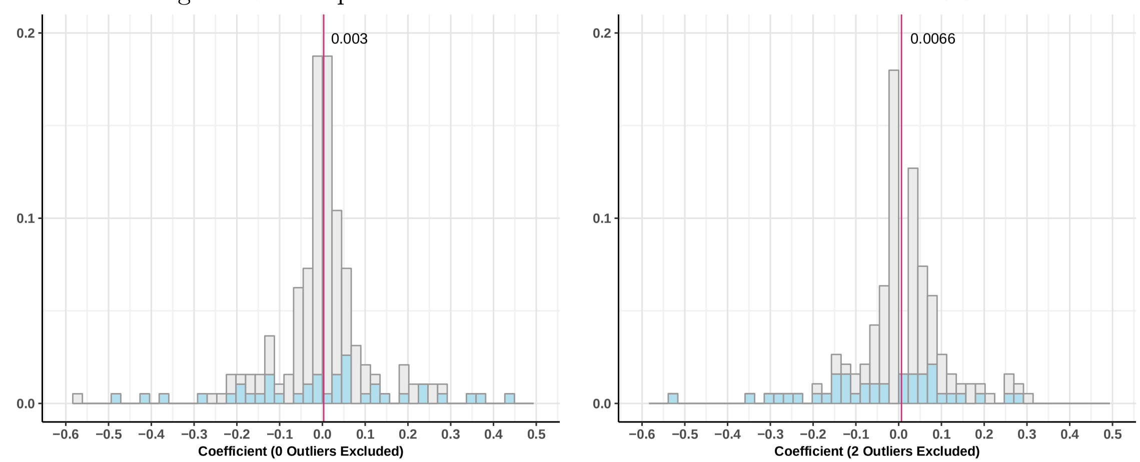


Table 1: Summary of Results: Long Run Ad Elasticities w/ Carryover Parameter $\delta = 0.9$

	Naive		Baseline		Border Strategy	
	All	Positive & Significant	All	Positive & Significant	All	Positive & Significant
Number of Brands	288	121	288	77	288	73
10th Percentile	-0.0704	0.0370	-0.0397	0.0253	-0.0307	0.0196
Median	0.0305	0.0846	0.0140	0.0655	0.0144	0.0716
Mean	0.0417	0.1384	0.0232	0.0967	0.0257	0.0990
90th Percentile	0.1834	0.2983	0.0918	0.2002	0.1007	0.2171
Fraction Not Significant	0.385	-	0.660	-	0.674	-
Fraction Positive Significant	0.420	-	0.267	-	0.253	-

Note: Descriptive statistics of estimated advertising elasticities reported for three model specifications and 288 brands. Elasticities derived from regressions of log quantity on log advertising GRP stock (own and competitor) and log prices (own and competitor). The naive model does not include any additional controls. The baseline model includes store, month, and week-of-year fixed effects. The border strategy i) restricts the sample to those stores that are located in a county that shares a border with a different DMA and ii) includes store, border-month, and week-of-year fixed effects. Regressions are estimated separately for each brand. The unit of observation in each regression model is a store-brand-week. Standard errors are two-way clustered at the DMA level and the week level in the Naive and Baseline specifications. Standard errors are two-way clustered at the border-side level and the week level in the Border Strategy specification.

Table 2: Summary of Results: Ad Elasticities w/ $\delta = \{0, 0.25, 0.5, 0.75, 0.9, 0.95, 1\}$

	$\delta = 0$	$\delta = 0.25$	$\delta = 0.5$	$\delta = 0.75$	$\delta = 0.9$	$\delta = 0.95$	$\delta = 1$
Number of Brands	288	288	288	288	288	288	288
10th Percentile	-0.0074	-0.0102	-0.0134	-0.0177	-0.0307	-0.0507	-0.0719
Median	0.0019	0.0036	0.0054	0.0073	0.0144	0.0148	0.0114
Mean	0.0029	0.0039	0.0059	0.0112	0.0257	0.0378	0.0352
90th Percentile	0.0128	0.0156	0.0233	0.0415	0.1007	0.1431	0.1827
Fraction Not Significant	0.778	0.750	0.722	0.684	0.674	0.701	0.764
Fraction Positive Significant	0.174	0.198	0.222	0.236	0.253	0.215	0.160
Median Bias of Positive and Significant Estimates	515.79%	261.11%	203.70%	290.41%	397.22%	685.81%	1412.28%
Mean Bias of Positive and Significant Estimates	565.52%	420.51%	366.10%	325.89%	285.21%	309.52%	496.02%

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

Table 3: Summary of Results: Long Run Ad Elasticities w/ Best Delta

	Baseline		Border Strategy	
	All	Positive & Significant	All	Positive & Significant
Number of Brands	288	106	288	112
10th Percentile	-0.1130	0.0104	-0.0351	0.0104
Median	0.0090	0.0540	0.0114	0.0354
Mean	0.0112	0.1081	0.0268	0.0823
90th Percentile	0.1750	0.2548	0.1441	0.2169
Fraction Not Significant	0.486	-	0.486	-
Fraction Positive Significant	0.368	-	0.389	-

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

Table 4: Summary of Results: Statistical Power of Border-Strategy w/ Carryover Parameter $\delta = 0.9$

	Border Strategy	
	50% power to detect elasticity = 0.05	Positive & p < 0.005
Number of Brands	157	32
10th Percentile	-0.0214	0.0268
Median	0.0075	0.0768
Mean	0.0083	0.1118
90th Percentile	0.0353	0.2645
Fraction Not Significant	0.688	-
Fraction Positive Significant	0.217	-

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

Table 5: Summary of Results: Top Competitor Ad Elasticities with $\delta = 0.9$

	Baseline	Border Strategy
Number of Brands	192	191
10th Percentile	-0.1288	-0.1039
Median	0.0030	0.0066
Mean	-0.0023	-0.0087
90th Percentile	0.1082	0.0965
Fraction Not Significant	0.797	0.785
Fraction Positive Significant	0.104	0.105
Fraction Negative Significant	0.099	0.110

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