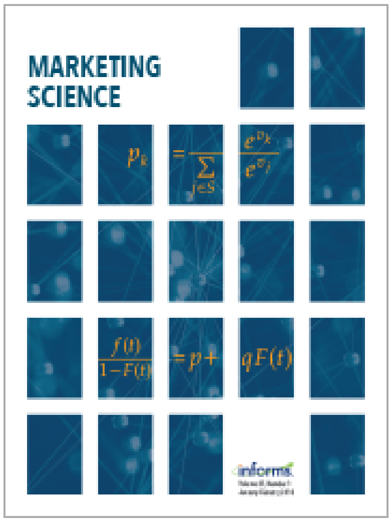


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

How and When to Use the Political Cycle to Identify Advertising Effects

Sarah Moshary, Bradley T. Shapiro, Jihong Song

http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Vol. 40, No. 2, March–April 2021, pp. 283–304
ISSN 0732-2399 (print), ISSN 1526-548X (online)

How and When to Use the Political Cycle to Identify Advertising Effects

Sarah Moshary,[a] Bradley T. Shapiro,[a,b] Jihong Song[c]

[a] Booth, University of Chicago, Chicago, Illinois 60637; [b] National Bureau of Economic Research, Cambridge, Massachusetts 02138; [c] Department of Economics, Princeton University, Princeton, New Jersey 08544

Contact: sarah.moshary@chicagobooth.edu, https://orcid.org/0000-0002-4624-6941 (SM); bradley.shapiro@chicagobooth.edu, https://orcid.org/0000-0003-1938-6727 (BTS); jihongs@princeton.edu (JS)

Received: September 19, 2019
Revised: March 9, 2020; June 2, 2020
Accepted: June 10, 2020
Published Online in Articles in Advance: November 23, 2020

https://doi.org/10.1287/mksc.2020.1258

Abstract. A central challenge in estimating the causal effect of television (TV) advertising on demand is isolating quasirandom variation in advertising. Political advertising, which topped $14 billion in expenditures in 2016, has been proposed as a plausible source of such variation and thus, a candidate for an instrumental variable (IV). We provide a critical evaluation of how and where this instrument is valid and useful across categories. We characterize the conditions under which political cycles theoretically identify the causal effect of TV advertising on demand, highlight threats to the exclusion restriction and monotonicity condition, and suggest a specification to address the most serious concerns. We test the strength of the first-stage category by category for 274 product categories. For most categories, weak-instrument robust inference is recommended, as first-stage F statistics are less than 10 for 221 of 274 product categories in our benchmark specification. The largest first-stage F statistics occur in categories that typically advertise locally, such as automobile dealerships and restaurants. Failure to use the suggested specification leads to results that suggest violations of exclusion and monotonicity in a significant number of categories. Finally, we conduct a case study of the auto industry. Despite a very strong first stage, the IV estimate for this category is imprecise.

History: Avi Goldfarb served as the senior editor and Bart Bronnenberg served as associate editor for this article.

Supplemental Material: The data/replication files and online appendix are available at https://doi.org/10.1287/mksc.2020.1258.

Keywords: advertising • advertising effectiveness • political advertising • causal effects • instrumental variables

1. Introduction

In 2016, political groups in the United States spent in excess of $4 billion on television (TV) advertising. By increasing the price of advertising time, this large disruption in the TV advertising market may generate exogenous variation in nonpolitical commercial advertising.[1] Using this variation in an instrumental variables (IVs) framework was first suggested by Sinkinson and Starc (2019), which estimates the effect of advertising for cholesterol-lowering drugs. In this paper, we provide a critical evaluation of that approach using data across all product categories that advertise on television in the United States. Our intention is to guide practitioners on how and when to use political advertising as an instrument. We also seek to provide researchers a road map for scrutinizing IV strategies that are meant to have broad applicability by establishing the properties and regularities of this candidate instrument. This study may be read as a cautionary one. Our findings suggest that despite the intuitive appeal of the political advertising IV strategy, the plausibility of the exclusion restriction is sensitive to the specification, and the first stage is often weak. These findings highlight pitfalls that may plague other IVs as well.

We begin by characterizing the conditions under which the political advertising IV strategy is valid. This first step is crucial for any IV strategy because the exclusion restriction and monotonicity condition are not directly testable. Thus, a sound theoretical argument is needed to pinpoint a specification that satisfies both conditions. In this case, the justification of the exclusion restriction rests on the premise that political advertising reduces commercial advertising by increasing the price of commercial advertising. Operationally, we argue that this requires the inclusion of television market fixed effects and time fixed effects at the periodicity of the data in order to isolate the price mechanism and prevent contamination from other sources that might violate exclusion or monotonicity. These arguments apply to any past or future election where the television advertising market continues to operate as it does in our data.

Second, we characterize where the IV strategy has a sufficiently strong first stage to limit weak-instrument bias. We use data on political advertising and commercial advertising for 274 product categories from 2010 to 2016. These data cover two U.S. presidential elections (2012 and 2016) and several midterm, state, and local elections. Overall, we find that relatively few categories have a sufficiently strong first stage to alleviate concerns about weak-instrument bias. For example, in a simple log-linear specification, we find that political advertising has a first-stage F statistic less than 10 in 221 of 274 product categories and an F statistic greater than 25 in only 28 categories.[2] Although the exact number changes across specifications, the qualitative result is the same: most categories exhibit a weak first stage.[3] Overall, political advertising has a very concentrated effect on a few categories that advertise almost exclusively locally; for example, advertising by car dealerships, hospitals, household furnishings outlets, and appliance stores is strongly offset by political advertising, whereas advertising for national brands in consumer packaged goods is not.[4]

Although the exact quantitative results on category-by-category first-stage F statistics do not necessarily generalize to political advertising associated with all future elections, they are useful for two reasons: first, they guide practitioners who hope to use variation in political advertising between 2010 and 2016 as an instrument for commercial advertising, and second, they guide our prior for how much and for which types of product categories a given level of political advertising is likely to disrupt the commercial advertising market in future elections.

Finally, we fully implement the strategy for one product category, automobiles, where both (1) the first stage is particularly strong and (2) the short-run effect of advertising on sales is expected to be zero. The purpose of this exercise is twofold. First, we want to characterize the properties of the IV estimator in an instance where the F statistic exceeds conventional thresholds that limit weak-instrument bias. Second, because we believe that the true advertising effect is zero, the IV estimate provides a placebo test of the exclusion restriction. We find that the IV strategy produces a confidence interval that contains the presumed truth (zero), but the confidence interval is wide, containing improbably large negative and positive values. Thus, this case study underscores that political advertising may not identify precise advertising elasticities, even for the 28 categories with a strong first stage. For these categories, our findings suggest that reducing noise in sales is of first-order importance to obtaining a precise estimate.[5]

To demonstrate the strength of the first stage, we focus on estimating category-level advertising effects because political advertising constitutes only one source of variation. As a result, it may be used to estimate a parameter on a single endogenous variable. In oligopoly settings, own advertising and rival advertising are potentially confounded by competitive responses, so that estimating brand-level effects using political advertising as an instrument requires either an additional source of variation or a particular functional form assumption.[6] As an example, in Section 2.6, we describe how to use political advertising in conjunction with a logit model of demand to estimate brand-level advertising effects.

Category-level advertising effects are also interesting and important in their own right in many circumstances. For many research questions (e.g., Shapiro 2018, Sinkinson and Starc 2019), separating category expansion from the business-stealing effects of advertising is important and requires an instrument like political advertising that can identify category-expansive advertising effects. Also, several policy debates revolve around the category-level effect of advertising, such as bans on television advertising for junk food and smoking, as in Dubois et al. (2018) and Tuchman (2019).

This paper contributes to the literature that seeks to identify the effects of television advertising using observational data (including Hartmann and Klapper 2018, Shapiro 2018, Thomas 2018, and Sinkinson and Starc 2019, among others). Shapiro (2018) exploits the borders of television markets to generate quasirandom variation in advertising. Hartmann and Klapper (2018) uses the ex ante uncertain identity of the teams in the Super Bowl, which provides randomness in which households watch television during the game. Thomas (2018) combines data on how TV advertising bundles viewers of a single show with information on ideal targeting to construct an instrument for advertising. Li et al. (2019) compare IV and non-IV approaches in estimating the effect of political advertising on votes. Sinkinson and Starc (2019) pioneered the approach in this paper, using political advertising as an exogenous shifter of television advertising. Lovett et al. (2019) show that a version of this instrument provides little power in determining the effect of advertising on word of mouth. All of these approaches are clever ideas implemented in case studies. This paper goes beyond a single case study to characterize the usefulness of the instrument more generally. Thus, it also adds to the literature moving beyond case studies to characterize empirical results in a generalizable way (Shapiro et al. 2019).

Beyond the case of advertising, this paper adds to the literature on instrumental variables, in particular by providing a road map for scrutinizing instruments meant to have broad applicability. This adds to recent work on Bartik shift-share style instruments (Borusyak et al. 2018, Goldsmith-Pinkham et al. 2018) and instruments based on competitor attributes (Gandhi and Houde 2017) by characterizing the circumstances under which IV approaches, which are meant to be generally implementable, are valid.[7] Finally, this paper also adds to the literature thinking carefully about the use of instrumental variables in marketing contexts (Rossi 2014). Our results validate many of the concerns of Rossi (2014) with the implementation of IVs in marketing—in particular, the necessary conditions for validity in this case are generally not innocuous. As a result, researchers wishing to use the instrument require a strong theoretical argument that the assumptions hold in their particular settings.

The paper proceeds as follows. Section 2 describes the political advertising IV strategy, including the theoretical justification for the instrument, exclusion restriction, and monotonicity condition. Section 3 describes the Nielsen Ad Intel data and illustrates how political advertising is broadly related to commercial advertising. Section 4 presents the results for the first stage across categories. Section 5 shows a case study as a proof of concept using the automobile category, which has a strong first stage. Section 6 discusses limitations of the approach and concludes.

2. Conditions for Validity

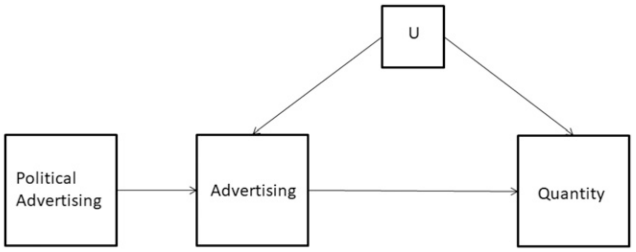
In this section, we present the conditions required for political advertising to identify the causal effect of commercial advertising at either the category level or brand level (for a monopolist). First, political advertising and commercial advertising must be correlated strongly enough to avoid weak-instrument bias. This condition is testable in the data. Second, we need any correlation between political advertising and commercial sales to operate exclusively through changes in commercial advertising or an exclusion restriction. This condition is not directly testable, and we will use theory to justify it. Third, we require a monotonicity condition: that each treatment unit exhibits changes in commercial advertising in the same direction for a given change in political advertising. This condition is also not directly testable and will require theoretical justification.

The goal is to estimate the causal effect of category-level advertising on category-level sales. For the purposes of exposition, we will employ a log-log functional form as governing the true relationship of interest:

$$\log(1 + Q_{jmt}) = \beta \cdot \log(1 + A_{jmt}) + \omega_{jmt}, \tag{1}$$

where j indexes category, m indexes television market, and t indexes time in months. To be clear, we would like to estimate this regression equation category by category. An ordinary least squares (OLS) regression of sales on advertising will not recover the causal parameter of interest (β) because advertising is likely a function of the unobservable ϵ_{mt}. As a result,

Figure 1. Political Advertising and Commercial Demand—A DAG



we propose an instrumental variables approach using political advertising as an exogenous shifter of category-level commercial advertising.

To fix ideas about how the relationship between commercial advertising and political advertising maps into the necessary conditions for instrument validity, we employ a series of directed acyclic graphs (DAGs). Figure 1 illustrates the basic problem. Although we want to estimate the effect of advertising on quantity, there exist unobservables, U, that simultaneously drive advertising and demand. We can think of these unobservables as demand shocks. We propose using political advertising as a source of variation in advertising that is unrelated to those unobservables. In each subsection, we will break apart this figure to identify sources of validity and sources of potential threat.

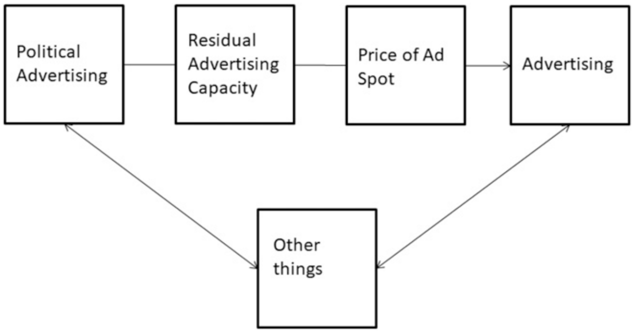
2.1. Relevance

Figure 2 shows the relationship between the instrument, political advertising, and the endogenous variable, commercial advertising. The intuition for the political IV strategy is that there is a maximum capacity of advertising time. Thus, when candidates advertise, they reduce the residual supply of advertising facing other product categories.[8] The negative supply shock increases equilibrium prices. Because demand curves slope downward, higher prices lower the quantity of airtime demanded by commercial advertisers. Thus, we predict a negative sign on the first stage of commercial advertising on political advertising. In other words, an increase in political advertising in market m at time t should decrease commercial advertising.

It is important to note that although this argument holds directionally for each commercial advertiser, it need not hold with equal magnitude across potential commercial advertisers. The magnitude depends on exactly which ad spots are of the most interest to political advertisers as well as the willingness to pay of commercial advertisers. A main contribution of this paper is to document when and where the first stage is sufficiently strong so as to be useful.[9]

It is also important to note that there may be other sources of correlation between political advertising and commercial advertising. These factors may threaten

Figure 2. Political Advertising and Commercial Advertising



either the exclusion restriction or the monotonicity condition, which we discuss. These potential threats are depicted in Figure 2 below the main proposed channel of the relationship between political advertising and commercial advertising.

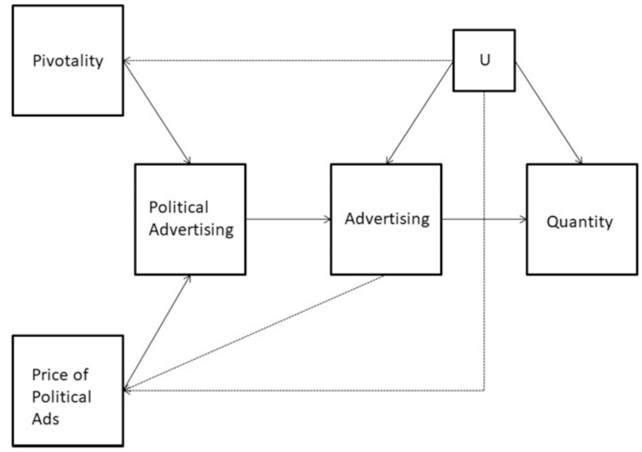
2.2. Exclusion

To recover an unbiased estimate of the causal effect of commercial advertising on sales, political advertising must be correlated with product sales exclusively through its correlation with the endogenous variable, commercial advertising. Because political advertising is not truly randomized across markets by campaigns, understanding political demand for airtime is important for justifying this exclusion restriction.

Political advertisers wish to maximize their chance of winning an election given a fixed budget. To accomplish this goal, they must advertise to viewers who meet two criteria: viewers must be persuadable and also potentially pivotal in the election of interest. A viewer is potentially pivotal if changing his or her vote could change the result of the election. That is, political campaigns do not want to advertise in a market where the election outcome is a foregone conclusion whether they advertise or not—advertising in these markets would waste their scarce budgets. Thus, this campaign objective suggests two determinants of political advertising: (1) the price of airtime and (2) the likelihood that the vote of the marginal ad viewer will flip the election.

Figure 3 illustrates these sources of demand for political advertising. The potential threats to exclusion are depicted with dotted lines. In particular, we must assume that there is no path leading from the unobservable demand shocks, U, to political advertising. The first threat is a direct path from U to the pivotality of a market. This would be the case if the product market in question was especially politically relevant. For example, if the product market was health insurance on the Affordable Care Act "Obamacare" exchanges and access to health insurance was a particularly important political issue in a market,

Figure 3. Political Advertising and Commercial Advertising



demand shocks for health insurance could directly impact pivotality. In this case, the exclusion restriction is violated because some of the variation in political advertising would be contaminated by the unobserved demand shock U.

The second potential threat to exclusion acts through the equilibrium in the market for advertising. That is, a demand shock, U, could lead to higher (lower) commercial advertising for a product category. This increase (decrease) in advertising decreases (increases) the residual supply of advertising, raising (lowering) the price of advertising faced by political campaigns. This price change would lower (raise) the amount of political advertising. In this scenario, there is a negative correlation between political advertising and commercial advertising, but that correlation operates through U, which has a direct effect on commercial demand, violating the exclusion restriction.

We argue that making use of market and time fixed effects alleviates these concerns to an extent. Nonelection time periods serve as a "pretreatment" period, whereas high and low political advertising markets serve as "treatment" and "control" units, respectively. Exclusion requires that political competitiveness (pivotality) does not drive the relative changes in commercial product demand conditions across markets between election and nonelection time periods. This type of violation only seems plausible for very special product categories. As an example, consider a market where a manufacturing plant closed in October of an election year, leaving a large number of people without health insurance. The closure stimulates demand for individual insurance and potentially makes the political election more/less competitive. However, such stories are difficult to tell for the typical product category that advertises on television.

In terms of paths from U to prices faced by political campaigns, the difference-in-differences style variation alleviates these concerns to a large degree. Much of the variation in political advertising comes from variation in the political cycle—there is nearly zero political advertising in the pretreatment period, not because advertising prices are high but because there is no impending election.[10] Even during election season, many markets are completely noncompetitive and hence, see almost no political advertising. The difference in commercial advertising between those markets and politically competitive markets is significantly larger than the difference in political advertising among the competitive markets. These observations combined lead us to argue that the majority of residual variation in political advertising net of market and time fixed effects is driven by the likelihood that political advertising will shift the election rather than by relative prices.

Another consideration is whether firms modify their other promotional activities in response to the political advertising shock. For example, if TV advertising becomes expensive, a firm might substitute from TV to digital advertising, confounding the relationship between TV advertising and sales. Alternatively, a firm might reduce complementary promotional activities in conjunction with a reduction in TV advertising. Fortunately, this potential confound is directly testable with data on these other promotional activities.

Finally, if demand shocks U alter the prices faced by campaigns, then these shocks must also change aggregate commercial advertising. Most commercial categories constitute a small fraction of total advertising, so that even a large change in a single category's advertising would be unlikely to disrupt the market as a whole. As a result, the type of demand shock U that would be most problematic is one correlated across all commercial product categories, such as a negative income shock. If residents of a particular designated market area (DMA) cut spending across all categories, then total commercial advertising demand would fall, increasing the residual supply facing political campaigns and thus, increasing political advertising. In this scenario, more political advertising is spuriously correlated with lower commercial demand, leading to an overestimate of the effect of commercial advertising on demand. In Section 3.2, we estimate the total disruption in commercial advertising that can be attributed to political advertising, which provides some perspective on the likelihood that the exclusion restriction is violated in this fashion. However, if researchers are particularly concerned about the price mechanism in their specific case, they may wish to find a more direct measure of political pivotality to serve as an instrument for commercial advertising.

Overall, we cannot fully rule out these threats to the exclusion restriction. As a result, we instead provide a road map for the required theoretical arguments that must be made to justify exclusion. If a researcher is willing to assume a priori that an advertising effect must be greater than or equal to zero, a positive and significant reduced form coefficient from the regression of commercial demand on political advertising would indicate a violation of the exclusion restriction. Such an assumption might be considered reasonable in the context of category-level advertising effects. In Online Appendix D, we present reduced form regression results for a set of products in which we have demand data from AC Nielsen's retail measurement services (RMS) scanner data. For 5 of 36 product categories, we estimate a positive and significant reduced form coefficient, indicating a possible violation of the exclusion restriction in those categories (although we would expect 1–2 categories to be positive and significant by chance).[11] The frequency of these results may suggest to researchers that this exclusion restriction is not entirely benign and requires careful justification on a category-by-category basis.

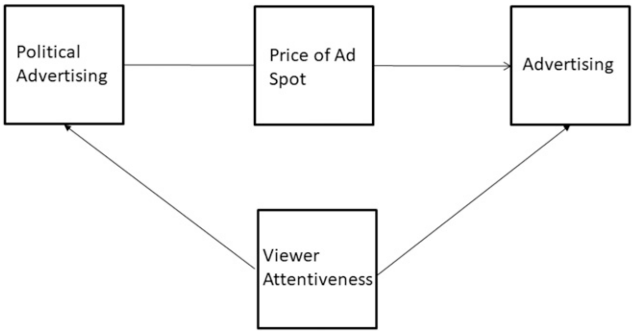
2.3. Monotonicity

Finally, the application of standard instrumental variable techniques requires a monotonicity condition, following Angrist and Imbens (1995). Under the theory outlined, where political advertising crowds out commercial ads, monotonicity means that an increase in political advertising must never increase commercial advertising. There are two potential threats to this assumption, and neither involves a relationship between political advertising and commercial advertising through the price mechanism (the lower path in Figure 3).

First, if there are time-varying differences across markets in viewer attention, then political advertising and commercial advertising could exhibit a positive relationship. Both political advertisers and commercial advertisers would like to reach more attentive audiences. If this effect outweighs price effects for any market-time observations, monotonicity will fail. We note that attentiveness threatens the monotonicity condition, even if it does not affect product market sales directly (if attentiveness drove sales, then it would constitute a violation of the exclusion restriction). This mechanism is illustrated in Figure 4. To deal with this potential issue, we include market fixed effects. Although this does not solve any issues surrounding attentiveness shocks at the market-time level that can be anticipated by political and commercial advertisers, it does account for the fact that viewers in some markets pay more attention to their televisions. Any residual attentiveness effects on both political and commercial advertising must be assumed to be zero.

The second threat to monotonicity relates to the fact that both political advertisers and commercial

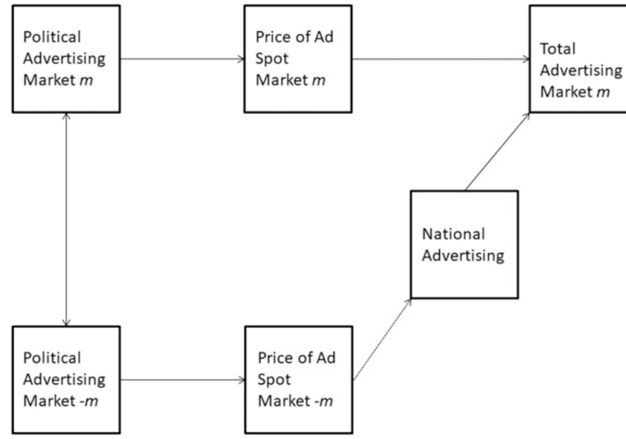
Figure 4. Political Advertising and Commercial Advertising



advertisers operate in multiple, interrelated markets. First, political advertisers tend to advertise in ways that are correlated across months. In particular, there is more political advertising in all markets in months just before an election than in months far from an election. Figure 5 illustrates how this can be a problem for the monotonicity condition. In particular, consider advertising in market m. If our intuition holds, then political advertising in market m increases the price of a local ad spot in market m, reducing local commercial advertising in that market. At the same time, political advertising also increases in market $-m$. Again, the price of a local ad spot in market $-m$ increases, reducing the quantity demanded for local ad spots in market $-m$. The reduction in local commercial advertising in market $-m$ is not concerning by itself. However, if the viewers in market $-m$ are sufficiently important to the advertiser, it may choose to advertise *nationally*. National advertising is seen by viewers in all markets. So although demand for local ad spots in market m decreases, viewers in market m may see more ads in total because more national ads are run. Monotonicity is violated if the increase in national ads dominates the decrease in local ads in market m for the category in question.

This problem is driven by the combination of two factors. First, political advertising is correlated across markets within a time period. Second, national advertising is seen by viewers in all markets. We can control for both of these factors by including a *time fixed effect at the periodicity of the data*.[12] These fixed effects allow us to partial out both national advertising and the correlation in political advertising between markets within a time period, so that only changes in local ad spots in market m identify advertising effects. Conditional on time fixed effects at the level of periodicity in the data, monotonicity violations coming from substitution from one local market to another local market are still possible but are less obviously plausible than substitution to national advertising.

Figure 5. Political Advertising and Commercial Advertising



One way to check the plausibility of the monotonicity condition is to inspect the sign of the first-stage estimate. The theory highlighted in Section 2.1 indicates that political advertising and commercial advertising should be negatively correlated. A positive correlation indicates a possible violation of the monotonicity condition, either coming from shocks to viewer attentiveness or from substitution across markets.[13] In Online Appendix F, we show that with market fixed effects and year-month fixed effects, only about 3% of first-stage estimates are positive and significant, roughly the proportion that should occur by chance under a null hypothesis of a zero first stage. However, if we remove the year-month fixed effects or use coarser time controls, nearly 50% of first-stage estimates are positive and significant, and more are positive than negative, violating the theory outlined. These patterns highlight the importance of time fixed effects at the periodicity of the data to implement this research design.

Here, we note again that a researcher may impose further structure on the problem and explicitly model heterogeneous treatment effects rather than use the local average treatment effect (LATE) formulation. This could, in principle, make the monotonicity assumption unnecessary. That said, a positive relationship between political advertising and commercial advertising remains a warning sign that the instrument does not work as the theory prescribes (through offset). As this theory justifies the exclusion restriction, evidence against that theory casts doubt on the exclusion restriction, even if the monotonicity assumption is not needed.

2.4. LATE Considerations

In the presence of heterogeneous treatment effects, a valid instrumental variable identifies an LATE. That is, the treatment effect identified is the average effect among "compliers" to the instrument, those ads that are crowded out by political advertising. To interpret the effects of crowd out, it is important to consider how the market clears. If TV stations do not anticipate political advertising demand when selling in the up-front market to commercial advertisers (typically in May or June), then they may displace commercial ads in response to an influx of political dollars. Indeed, many contracts include provisions for the priority of the airtime: a low-priority ad is more likely to get displaced than a more expensive high-priority ad. If an ad spot is displaced, then the station issues a "make good," and the spot is aired at another time. Although make goods supposedly offer comparable viewership, industry wisdom is that they are generally of inferior quality. Because priority is explicitly priced into commercial advertising contracts, political ads presumably displace spots for which commercial willingness to pay (WTP) is low.[14]

Such selection may be more severe if stations anticipate political demand. In this case, stations should set higher prices in the up-front market in election years, particularly in contested states. High prices reduce advertising by the lowest WTP commercial firms. Therefore, political gross rating points (GRPs) instruct us about the treatment effect from the bottom of the distribution of advertising efficacy.

An issue related to both LATE considerations and the exclusion restriction is the potential for commercial firms to coordinate other marketing levers, such as sales force or price promotions, with its TV advertising strategy. In this case, even if political advertising does not directly affect this other lever, it may indirectly change other marketing efforts through its effect on commercial advertising. Any indirect effects would change the interpretation of the LATE. Political advertising would then identify the "total" effect of advertising plus other marketing activities, including the direct path from advertising to sales as well as a path from advertising to other marketing levers to sales. To be clear, the LATE would not identify how TV advertising affects sales holding all else equal. This concern must be evaluated on a case-by-case basis: it is a nonissue for firms that set marketing levers independently but a potential problem for others that link them algorithmically. If possible, we advocate directly testing the relationship using data on these other marketing activities.

2.5. Preferred Specification

Our regression of interest, described in Equation (1), is a standard log-log advertising regression for category j in market m on contemporaneous sales at time t. The goal of this paper is to use political advertising as an instrument for category advertising flow, A_{jmt}.[15] In order to address the most plausible threats to monotonicity and the exclusion restriction, our final specification includes market fixed effects and time fixed effects at the periodicity of the data. We specify our first-stage equation as

$$\log(1 + A_{jmt}) = \gamma \cdot f(P_{mt}) + \alpha_m + \alpha_t + \epsilon_{jmt}. \qquad (2)$$

In the equation, A_{jmt} is the GRP for category j in market m in month t; P_{mt} is political advertising market m in month t; γ governs the relationship between political and commercial advertising; and α_m, α_t are market and month fixed effects, respectively. We specify category advertising on the left-hand side as a log because researchers typically want to identify an advertising elasticity. Here, we leave the relationship between the endogenous variable and the instrument reasonably general, as we will show robustness to many different first-stage functional form specifications, including a flexible one using machine learning. Next, we specify the reduced form equation as

$$\log(1 + Q_{jmt}) = \pi \cdot g(P_{jmt}) + \alpha_m + \alpha_t + u_{jmt}. \qquad (3)$$

2.6. Brand-Level Advertising Effects

Our preferred specification estimates the effect of advertising on sales for an entire category and in the case of monopoly, for a particular brand. Even in oligopoly settings, category-level advertising elasticities are of interest because they reveal the extent to which advertising is market expanding. For a good like cigarettes, the importance of an advertising ban hinges precisely on whether advertising increases smoking. However, in other instances, it is important to quantify business stealing; such quantification is challenging with the political advertising instrument because it affects all firms within a market simultaneously. If the researcher is willing to impose additional structure on the relationship between advertising and sales, however, then more progress can be made. As an example, one approach is to model s_{kmt}, the share of brand k in market m at time t, using the logit[16] as follows:

$$s_{kmt} = \frac{\exp\{X'_{kmt}\beta - \alpha \cdot p_{jkt} + \gamma \cdot Ads_{kmt} + \xi_{kmt}\}}{1 + \sum_{k' \in K} \exp\{X'_{k'mt}\beta - \alpha \cdot p_{k'mt} + \gamma \cdot Ads_{k'mt} + \xi_{k'mt}\}}.$$

Using the standard Berry (1994) inversion, the difference between the share of brand k and the outside option s_{omt} can be written as

$$\ln s_{kmt} - \ln s_{omt} = X'_{kmt}\beta - \alpha \cdot p_{kmt} + \gamma \cdot Ads_{kmt} + \xi_{kmt}, \qquad (4)$$

which we can estimate using linear IV with political advertising as an instrument. The key restriction is that a single parameter (γ) governs both own- and crossadvertising elasticities:

$$\epsilon_k = \gamma \cdot Ads_{kmt} \cdot (1 - s_{kmt})$$
$$\epsilon_{kk'} = -\gamma \cdot s_{k'mt} \cdot Ads_{k'mt} \text{ for all } k' \neq k.$$

The idea is analogous to estimating a price coefficient in a logit demand function using a single industry-wide cost shock.

Finally, we note that in this framework, the category-level F statistics presented are informative about the brand-level F statistics. To be clear, the first stage that corresponds to Equation (4) is

$$Ads_{kmt} = X'_{kmt}\pi_o + \pi_1 \cdot PGRP_{mt} + \epsilon_{kmt}. \quad (5)$$

It is important to cluster standard errors at least at the market-time ($m \times t$) level in this brand-level regression (Equation (5)) precisely because this is the level at which political advertising shocks occur.[17] The category-level regressions exploit exactly this market-time variation, and so, the category-level F statistics speak to the magnitude of this variation. Said differently, disaggregating the data to the brand level rather than the category level does not buy the researcher any additional variation in political advertising and so, will not typically buy the researcher additional first-stage strength.

3. Data and Aggregate Analysis

3.1. Ad-Intel Data

Our main source of data is the Nielsen Ad-Intel database that records all TV advertisements in 130 DMAs in the contiguous United States from 2010 to 2016.[18] The ads are recorded at the occurrence level, where an occurrence is the placement of an ad by a specific advertiser on a given channel in a specific market for a specific duration on a given date and time. Nielsen uses its propriety technology to collect TV programming information and identify advertising occurrences based on their unique audio and video content.[19] Occurrences are therefore measured in the same way in all of the 130 DMAs used in our analysis.

Because occurrences can differ substantially in their reach, estimated impressions (the number of eyeballs viewing an ad) can serve as a scaling factor to help in comparisons between ad occurrences. For each ad occurrence, we calculate GRPs, a frequently used measure of advertising intensity, as the number of impressions for the ad as a percentage of all TV-viewing households in a DMA.

The Ad-Intel database provides impressions estimates across markets and media types. For local media types, the database provides impressions at the "station–month–day of week" level in 5- to 15-minute time intervals.[20] In the top 25 DMAs, Nielsen measures impressions using "Local People Meters" (LPMs) that capture all TV-viewing activities of Nielsen households, so the data are available in all months.[21] In all other DMAs, Nielsen measures impressions using diaries filled out by Nielsen households, and the data are only available in February, May, July, and November ("sweep months").[22] For nonsweep months in those DMAs, we impute impressions by taking the average between the two closest sweep months.[23] We show in Online Appendix C that our results are robust to a more flexible imputation method. For national media types (we only use cable TV), the impressions are measured at the program level but are only available nationally; we thus compute GRPs for those ads at the national level and assume that they are the same in every market.[24]

Nielsen groups advertisers into 343 categories. Political campaigns and unions are classified as "B181: Organization Advertising: Political, Union," and independent expenditure-only political action committees (or super PACs) are found in a separate category titled "B189: Miscellaneous Organization Advertising." We identified 616 advertisers in the latter category as super PACs by manually matching the advertiser names with the list of super PACs created by OpenSecrets.[25] We then calculate total political advertising at the DMA-month level by summing up the GRPs for all category B181 ads and super PAC ads. Table 1 provides summary statistics for political advertising at the DMA-month level. Of particular note is that political advertising makes up a larger share of total GRPs than of total duration. This means that, on average, political advertisers buy spots with high viewership. Although political advertising accounts for as much as 9.39% of total advertising views in a market-month, it never makes up more than 1.53% of the total advertising airtime.

Among the 342 nonpolitical categories,[26] we use the top 274 that constitute 99.9% of total GRPs for analysis. We aggregate the data to the category-DMA-month level by summing up the GRPs of individual ads.

3.2. Political Advertising and the Commercial Advertising Market

In this section, we document how the commercial advertising market as a whole fluctuates with movements in political advertising. We begin by overlaying the time series of political advertising and commercial advertising for the Columbus, Ohio, DMA. Figure 6(a) shows these time series in levels. It plots the time series of average daily GRP in each month; both series are demeaned (the means are shown in the legend). Here, it appears that political advertising offsets commercial advertising almost one to one. Figure 6(b) shows the time series in log scale for both commercial and

Table 1. Summary Statistics: Political Advertising at DMA-Month Level ($N = 10,917$)

			Quantiles								
	Mean	SD	Min	5%	10%	25%	50%	75%	90%	95%	Max
Political GRP	4,484.5	9,017.1	0.0	5.4	19.3	85.8	723.0	4,421.7	13,698	22,446	102,791
Political/total GRP (%)	0.53	1.03	0.00	0.00	0.00	0.01	0.08	0.54	1.65	2.64	9.39
Political duration (hours)	14.92	27.39	0.00	0.09	0.27	1.17	4.03	15.26	42.74	68.59	362.37
Political/total duration (%)	0.08	0.14	0.00	0.00	0.00	0.01	0.02	0.08	0.23	0.36	1.53

Note. SD, standard deviation.

political advertising. Here, we see that large percentage changes in political advertising lead to hardly any change in the percentage of commercial ads. Changes in log commercial advertising are nearly imperceptible in the picture. Panels (a) and (b) can be reconciled by the fact that each political ad GRP crowds out roughly one commercial ad GRP, but because political advertising is a relatively small share of total advertising, the total disruption of the commercial ad market is small in percentage terms.

Next, we analyze the relationship between political and commercial advertising systematically across all markets in regression form. In particular, we estimate regressions of the form

$$\sum_{j \in J} A_{jmt} = \gamma \cdot P_{mt} + \alpha_m + \alpha_t + \epsilon_{mt}, \qquad (6)$$

where J denotes a set of categories; A_{jmt} is the amount of advertising in commercial category j, in market m, in month t; and P_{mt} is the amount of political advertising in market m, and in month t. We include market and time fixed effects. We consider A and P measured both in duration (hours) and in GRPs. The coefficient γ measures the crowd-out effect in levels.

Panel (A) of Table 2 describes the results where A and P are measured in GRPs. Column (1) presents the results for all commercial advertising—that is, all nonpolitical advertising, except for ads that promote the television station's own programming. Our estimates suggest that each political GRP offsets 0.85 commercial GRPs. Column (2) shows the effect of political advertising on advertising for the television station's own programming. It is negative and significant but small—each political GRP offsets about 0.045 programming GRPs. Column (3) shows the effect of political advertising on all nonpolitical advertising. Each political GRP offsets about 0.86 nonpolitical GRPs. Notably, this number has a confidence interval that excludes one, which means that political GRPs lead to an expansion of total advertising GRPs (shown in column (4)). Each political GRP expands total GRPs by roughly 0.14. This suggests that although political advertising offsets commercial advertising, it also expands the total amount of advertising viewed on television. Panel (B) of Table 2 describes the analogous results where A and P are measured in duration rather than in GRPs. The findings are qualitatively and quantitatively similar for each column. To give a sense of how the results look in percentage terms, we also report the approximate percentage effects of a 1% increase of political advertising from its mean in Table 2. We observe that the magnitudes in percentage terms are considerably smaller. For example, column (1) of panel (A) shows that a 10% increase in political advertising corresponds to a 0.06% decrease in commercial advertising. In other words, although political ads crowd out commercial ads almost one to one, the effect is still small in percentage terms because there are far fewer political ads than there are commercial ads.[27]

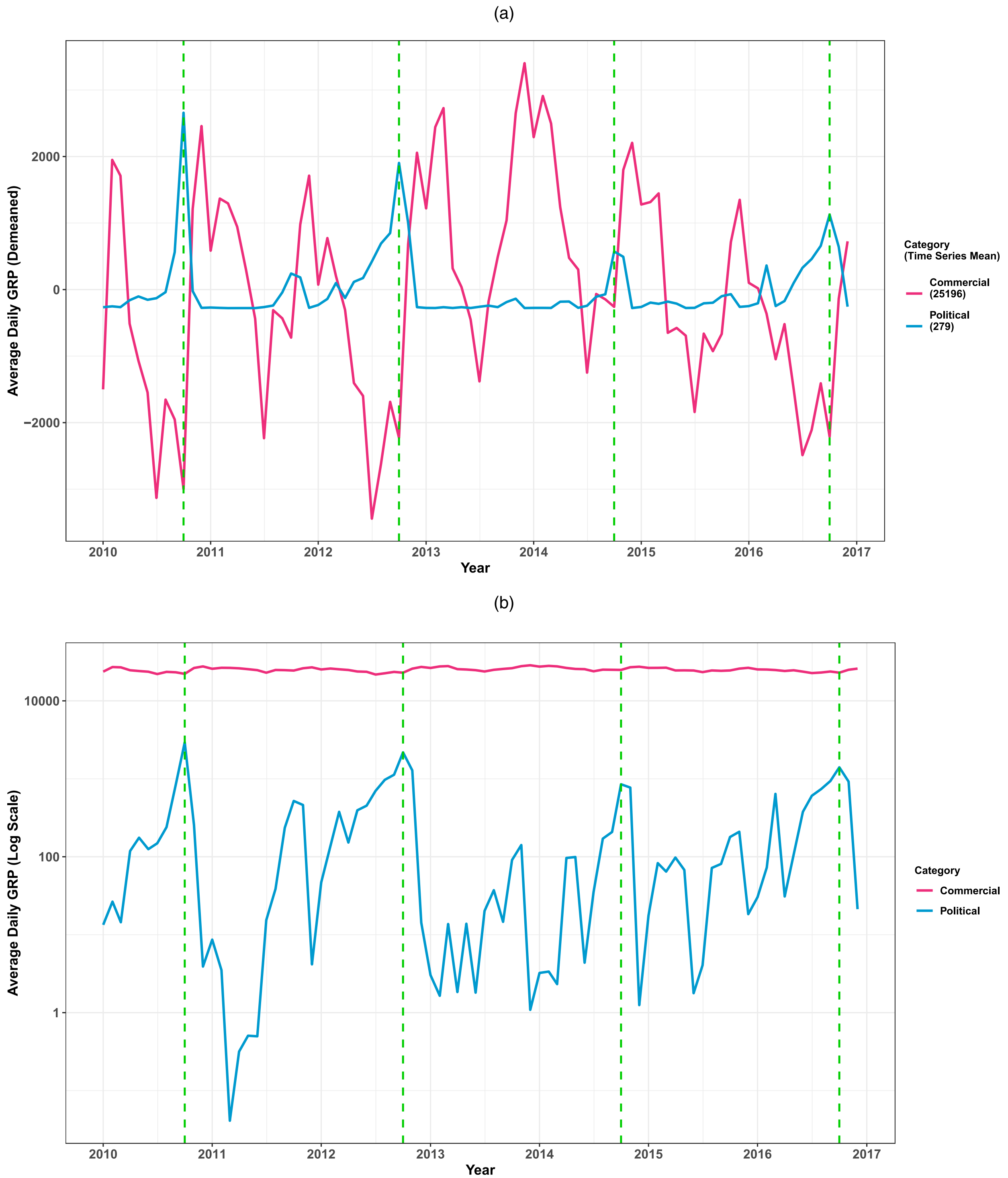
Overall, it seems that although political advertising offsets commercial advertising, it has a relatively minimal impact (percentagewise) on the advertising market as a whole.

4. Political Advertising as an Instrument

In this section, we test the strength of the political advertising instrument across 274 categories. We first employ Equation (2) and set $f(P_{mt}) = P_{mt}$ in our benchmark "log-linear" specification. We specify category advertising as a log in our benchmark model because researchers often want to identify an advertising elasticity, which is most natural with the endogenous variable in log form. We specify political advertising in levels in our benchmark model for simplicity.[28]

To ensure that our qualitative results are not driven by our somewhat arbitrary choice of functional form, we take two additional approaches. First, we estimate a second specification where both category advertising and political advertising are measured in levels. Second, we allow the first-stage shape to be guided by the data using a machine learning specification to pick the functional form of political advertising. We do so using the method in Belloni et al. (2012) to select the optimal $f(P_{mt})$ by combining 40 nonlinear functions using the least absolute shrinkage and selection operator (LASSO). Overall, our results are robust across alternative specifications and measurements, indicating that the qualitative results are not driven by the

Figure 6. (Color online) Time Series of Political and Commercial Advertising in Columbus, Ohio, DMA



Notes. Time series means for panel (a) are taken out and listed. (a) Average daily GRP in each month (time series mean taken out). (b) Average daily GRP in each month (log scale).

Table 2. Crowd-Out Effect for Aggregate Categories

Ad category	Commercial (1)	Programming (2)	Nonpolitical (3)	Total (4)
Panel A: Ads measured in GRP				
Political	–0.850***	–0.045***	–0.859***	0.141**
	(0.054)	(0.013)	(0.063)	(0.063)
Partial F	247.6	11.5	184.1	4.9
Partial R^2	0.062	0.002	0.043	0.001
Percentage effect (%)	–0.0056	–0.0014	–0.0046	0.0008
Panel B: Ads measured in duration				
Political	–0.798***	–0.021*	–0.770***	0.230***
	(0.049)	(0.011)	(0.050)	(0.050)
Partial F	265.2	3.8	233	20.8
Partial R^2	0.05	0.001	0.036	0.003
Percentage effect (%)	–0.0008	–0.0001	–0.0006	0.0002

Notes. "Programming" includes all categories for TV programs and TV networks/stations. "Commercial" includes all categories except "Programming" and "Religious, Charitable, and Humanitarian Organizations." "Nonpolitical" includes all categories except political ads. "Percentage effect" is the coefficient multiplied by 1/100 of the mean of independent variable and then divided by the mean of dependent variable. It approximates the crowd-out effect for a 1% increase of political advertising in percentage terms and is close to the estimate from a log-log specification. Standard errors are in parentheses, clustered at the DMA level. DMA and month × year fixed effects are included. $N = 10{,}917$.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

choice of functional form. Finally, we document several category-level characteristics that are linked to a strong first stage.

We cluster standard errors at the television market level to account for the potential serial correlation of errors within a market over time. We assess the instrument strength using the effective F statistic of Olea and Pflueger (2013), which is identical to the Kleibergen–Paap partial F statistic in the case with one endogenous variable and one instrument. For each specification, we classify categories into three bins: $F > 25$ (strong), $F \in [10, 25]$ (semistrong), and $F < 10$ (weak). We use these bins because researchers employ a variety of rules to determine whether an F statistic is "large enough" to allow for standard inference.[29] One common rule of thumb suggests that $F > 10$ is sufficient. However, Olea and Pflueger (2013) suggest a critical value of 23.11 for rejecting H_0 : Nager Bias < 10% with one instrument. As it is generally accepted that $F < 10$ constitutes a weak instrument, weak-instrument robust inference (e.g., Anderson and Rubin 1949, Andrews 2016) should be used for the $F < 10$ categories. For categories where $F \in [10, 25]$, whether the instrument is sufficiently strong depends on the relevant critical value to the researcher on a case-by-case basis, but for $F \in [10, 25]$, we suggest that researchers consider using weak-instrument robust inference. Although these two break points are arbitrary, they correspond roughly to the critical values for rejecting the null of weak instruments at different tolerance levels. For our benchmark specification, the Olea and Pflueger (2013) critical value is 23.11 for rejecting H_0 : Nager Bias < 10% and 12.05 for rejecting H_0 : Nager Bias < 30% at 5% level; the Stock and Yogo (2005) critical value is 16.38 for rejecting H_0 : Maximal Size < 10% at 5% level.

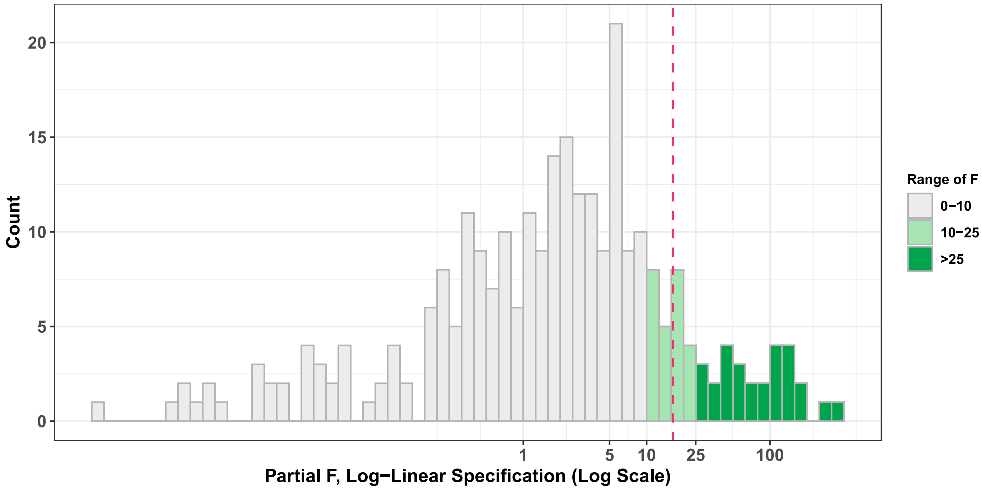
4.1. First-Stage Results

We begin by evaluating the first stage employing the level of political advertising (in GRPs) as a single instrument for the log of category advertising, which we call the "log-linear" specification. We will discuss these results in some detail and then present the results for alternative functional form choices.

The distribution of partial F statistic is reported in Figure 7, in which the vertical line is at the Stock and Yogo (2005) critical value of 16.38. The partial F ranges from 0.0004 to 327.5, and it is greater than 25 for 28 categories, between 10 and 25 for 25 categories, and below 10 for 221 categories. The categories with $F > 25$ are listed in panel (A) of Table 3, the categories with $F \in [10, 25]$ are listed in panel (B) of Table 3, and the categories with $F < 10$ are listed in Online Appendix A, Table 7.

The coefficient β_j in the log-linear first stage represents the crowd-out effect in log points of one extra political GRP. To make the coefficient easier to interpret, we scale it by the median political GRP across markets in October 2016 (which is 23,438). The scaled coefficient thus represents the median crowd-out

Figure 7. (Color online) Distribution of Partial F Statistic, Log Linear



effect in log points across markets in the most heated month before election. Figure 8 plots the distribution of scaled coefficients, which range from −0.264 to 0.088 and have a median of −0.008 across all categories. The median effect is −0.064 for the $F > 25$ categories, −0.028 for the $F \in [10, 25]$ categories, and −0.005 for the $F < 10$ categories.

The signs of the first-stage coefficients can also be used as a check on the theory used to justify the exclusion restriction and monotonicity condition. For the set of 28 categories with $F > 25$, the first-stage sign is negative for 27—consistent with crowd out—with the sole exception of "Miscellaneous Organization Advertising." Although not easily identified by our hand matching, many of these organizations may be political, so that a positive first stage reflects the importance of political pivotality for these advertisers. Overall, for the categories where the instrument appears strongest, we find no indication of a violation of our theory in the first-stage coefficient signs. For the set of 25 categories with $F \in [10, 25]$, only one has a positive sign, and that is "Cellular Radio Systems & Accessories." For the set of 221 categories with $F < 10$, 64 have a positive point estimate on the first-stage coefficient. Many of these may be because of chance, as the first stage is quite weak. For our specifications, in the categories for which we have a strong first stage, we only see a "wrong sign" first stage for one commercial advertising category. It appears that the market and time fixed effects do a reasonable job of correcting for the potential monotonicity concerns. Although these patterns do not prove that there are no monotonicity violations (this assumption is fundamentally untestable), it does speak to perhaps the most plausible ones.

4.2. Robustness to Specification and Measurement

In this section, we examine the robustness of the first-stage results to alternative specifications and measurements.

First, we show that the first-stage results are stable if we specify the category ads (left-hand side of Equation (2)) in levels instead of logs. Next, we use a machine learning method to capture the potential nonlinear effect of political advertising. Finally, we show that our results are robust to measuring advertising in total duration of advertising, which is measured consistently across markets and time (in lieu of GRP).

We start by specifying category advertising in levels rather than in logs in the first stage. Column (3) of Table 4 summarizes the results for the "linear-linear" specification. Comparing columns (3) and (1), changing the category GRPs from logs to levels adds five categories to the $F > 25$ bucket and five categories to the $10 < F < 25$ bucket. Figure 9 compares the partial F for all categories between the two specifications and shows that the identities of the strongest categories are not changed.

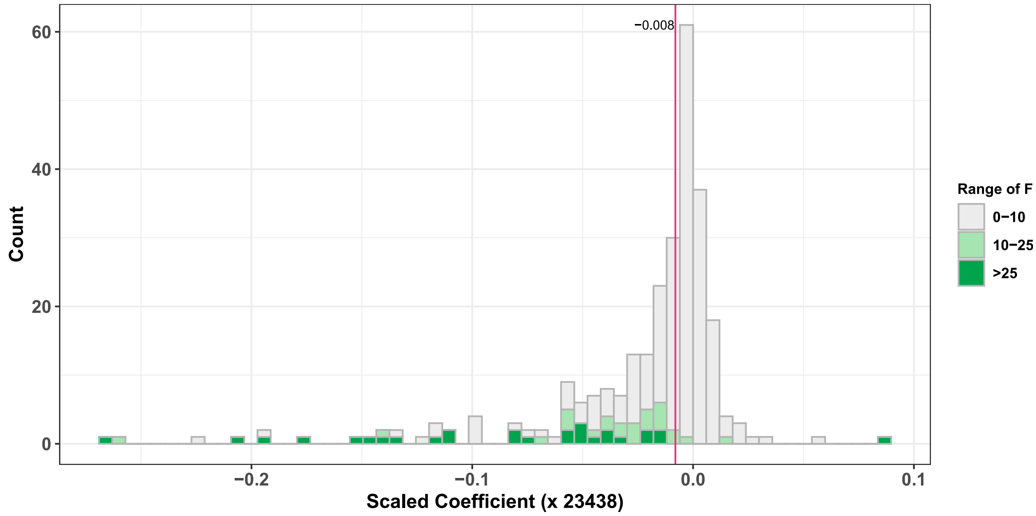
Next, we employ an LASSO to try and get additional first-stage strength through a better functional form fit. Technical details of the LASSO formulation are presented in Online Appendices B.1 and B.2.[30] The basic statistics of the "Log-Lasso" and "Linear-Lasso" specifications are listed in columns (2) and (4) of Table 4, respectively. Comparing with the "Log-Linear" specification, 3 categories are "downgraded" from $F > 25$ to $10 < F \leq 25$, but 28 categories are "upgraded" from $F < 10$ to $10 < F \leq 25$.[31] This is potentially a notable improvement, depending on which critical values are relevant to the researcher. However, zero categories are "upgraded" from $F < 25$ to $F > 25$. The identities of the strong instrument categories remain stable.

Table 3. List of Categories with Strong or Semistrong Log-Linear First Stage

Category	Partial F	Partial R^2	Scaled coefficient	Size rank	Spot share	GRP/ occurrence
Panel A: 28 Categories with $F > 25$						
Household Furnishings & Appliance Stores	327.5	0.058	−0.134	13	0.74	2.8
Autos & Light Truck–Dealerships	280.2	0.035	−0.197	16	0.96	2.9
Hospitals, Physicians & Misc. Physical Culture	197.2	0.031	−0.117	25	0.68	2.6
Optical Goods and Services	184.6	0.018	−0.139	70	0.59	2.6
Misc. Entertainment & Combination Copy	144.6	0.020	−0.264	78	0.80	2.3
Passenger Cars–Factory: New & CPO	143.0	0.026	−0.040	6	0.24	2.9
Miscellaneous Professional Services	141.2	0.028	−0.051	3	0.36	1.8
Restaurants, Hotel Dining & Nightclubs	135.4	0.026	−0.031	2	0.22	2.7
Construction, Engineering & Architecture	122.2	0.018	−0.154	61	0.84	2.7
Miscellaneous Retail	114.4	0.023	−0.043	7	0.25	2.5
Passenger Cars–Dealer Assn: New & CPO	106.8	0.018	−0.206	55	1.00	3.1
Light Trucks & Vans–Factory: New & CPO	105.9	0.024	−0.051	15	0.29	3.0
Misc. Organization Advertising	96.5	0.028	0.088	26	0.31	1.9
Dance, Theater, Concerts, Opera	91.3	0.014	−0.149	76	0.54	2.6
Automotive	72.2	0.011	−0.083	58	0.41	2.2
Schools & Camps, Seminars	71.9	0.014	−0.054	18	0.37	1.5
Light Trucks & Vans–Dealer Assn: New & CPO	63.0	0.011	−0.176	77	1.00	3.2
Data Communications Networks	53.5	0.010	−0.050	28	0.29	2.0
Misc Financial Inst. Services & Products	52.4	0.009	−0.042	33	0.25	1.6
Banks	45.0	0.006	−0.073	65	0.43	3.3
Cakes, Pies, Pastries & Donuts	43.1	0.005	−0.022	116	0.10	1.6
Cable Television Stations	42.9	0.007	−0.054	42	0.36	2.1
Direct Response Products	40.3	0.011	−0.012	8	0.05	0.9
Apparel, Footwear & Accessory Stores	36.8	0.004	−0.020	31	0.14	2.5
TV Station	35.4	0.006	−0.113	51	0.99	2.1
Fix-it Supplies	31.3	0.002	−0.113	203	0.13	1.8
Cereals	30.1	0.007	−0.013	20	0.06	1.8
Real Estate, R.E. Brokers & Developers	27.2	0.004	−0.083	123	0.43	2.5
Panel B: 25 Categories with $F \in [10, 25]$						
Plumbing & Sanitary Equipment	23.6	0.003	−0.026	113	0.11	2.4
Amusement Parks & Sporting Events	23.0	0.002	−0.059	71	0.48	2.2
Appetizers, Snacks & Nuts	21.4	0.003	−0.009	35	0.04	1.7
Automobile Insurance	20.0	0.004	−0.010	11	0.07	1.9
Medical & Dental Insurance	18.9	0.002	−0.055	69	0.43	2.0
Life Insurance	18.3	0.004	−0.023	97	0.07	1.2
Golf Equipment	17.9	0.004	−0.028	216	0.00	1.9
Breads, Rolls, Waffles & Pancakes	17.6	0.001	−0.033	95	0.18	1.6
Hotels & Resorts	17.2	0.003	−0.039	45	0.32	2.7
Food & Liquor Stores	16.8	0.003	−0.072	57	0.77	3.2
Religious, Charitable & Humanitarian Org.	16.1	0.003	−0.040	49	0.33	1.3
Lotteries	15.9	0.002	−0.262	142	1.00	3.2
Magazines, Newspapers, Misc Media	15.1	0.002	−0.034	107	0.15	2.1
Other Insurance & Combination Copy	14.4	0.003	−0.015	19	0.13	2.0
Medical Appliances & Equipment	14.3	0.001	−0.029	159	0.07	2.4
Cold, Cough & Sinus Remedies	12.7	0.001	−0.012	21	0.04	1.8
Misc Accessories, Supplies & Hardware	12.6	0.001	−0.054	175	0.11	2.2
Coffee, Tea, Cocoa & Derivatives	12.2	0.002	−0.016	85	0.06	2.5
Cellular Radio Systems & Accessories	11.8	0.003	0.013	46	0.03	2.8
Jewelry, Gift Stores & Galleries	11.2	0.001	−0.042	56	0.28	2.2
TV Program: Late Night News	11.0	0.002	−0.139	155	0.92	2.9
Comb Copy & Misc Major Appliances	10.4	0.001	−0.023	162	0.05	2.0
Cheese Products	10.2	0.002	−0.017	101	0.09	2.2
Computerized Games, Accessories & Software	10.2	0.001	−0.006	27	0.02	2.0
Drugs, Toiletries & Salons	10.2	0.001	−0.021	72	0.16	2.6

Notes. "Size rank" is the category rank by total GRP across all markets and years. "Spot share" is the share of ads bought in spot markets. "GRP/ occurrence" is the average GRP per occurrence. See Section 4.3 for further discussions about those category characteristics. "Scaled coefficient" is the log-linear regression coefficient times 23,438. CPO, certified preowned.

Figure 8. (Color online) Distribution of Scaled First-Stage Coefficients Colored by Range of Partial *F*



The "Linear-Lasso" compares similarly with the "Linear-Linear" model. Overall, the results of the first-stage analyses employing an LASSO are not qualitatively different from the benchmark models with a single instrument and do not produce any additional categories where political advertising is clearly a strong instrument.[32]

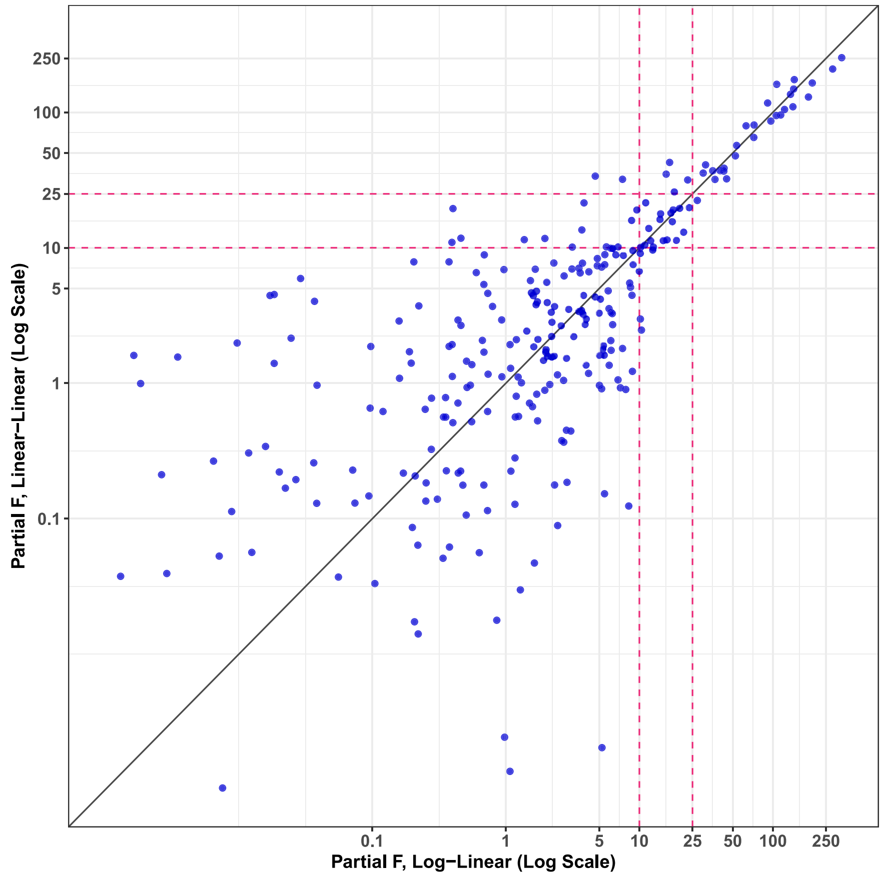
Finally, we show robustness to using advertising measured in duration rather than in GRP. This exercise is helpful for two reasons. First, as we discuss in Section 3.1, the impressions data for the 105 non-LPM DMAs are only measured by Nielsen in four sweep months (February, May, July, and November), so we must impute the impressions for nonsweep months. Because the vast majority of political ads occur from August to October in election years, one might worry that this induces measurement error in our analyses based on GRPs. In contrast, advertising duration is measured by Nielsen using the same technology across all 130 DMAs. Second, the intuition of crowd out is that a political ad takes the place of a commercial ad, a mechanism that operates through ad duration rather than viewership. Specifying the

Table 4. Comparison Across Different Specifications

Measure	GRP				Duration			
LHS	Log		Linear		Log		Linear	
RHS	Linear	Lasso	Linear	Lasso	Linear	Lasso	Linear	Lasso
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Number of categories in partial *F* bins								
> 25	28	25	33	32	19	18	19	18
10–25	25	49	30	56	33	45	40	56
< 10	221	13	211	14	222	10	215	12
NS		187		172		201		188
Panel B: Top three categories for each specification								
1	FA (327)	HP (197)	FA (253)	FA (253)	Car (195)	Car (195)	AD (152)	Car (115)
2	AD (280)	FA (163)	AD (209)	AD (209)	FA (161)	FA (118)	Car (115)	AD (87)
3	HP (197)	Car (143)	Ent. (174)	Opt. (130)	AD (116)	LT (105)	FA (97)	LT (72)

Notes. The left-hand side (LHS) is category advertising, and the right-hand side (RHS)is political advertising. Both sides are either measured in GRP or in duration. "NS" means that the Lasso algorithm does not select any of the 40 nonlinear transformations. The top three categories for each specification are listed in abbreviations, with their partial *F* statistics in parentheses. The abbreviations are "FA" for "Households Furnishings & Appliance Stores"; "HP" for "Hospitals, Physicians, & Misc. Physical Culture"; "AD" for "Autos & Light Truck: Dealerships"; "Car" for "Passenger Cars–New & Certified Preowned"; "LT" for "Light Trucks & Vans–New & Certified Preowned"; "Ent." for "Misc. Entertainment & Combination Copy"; and "Opt." for "Optical Goods and Services."

Figure 9. (Color online) Partial F: Linear/Log Linear



Note. This plot compares columns (3) and (1) of Table 4.

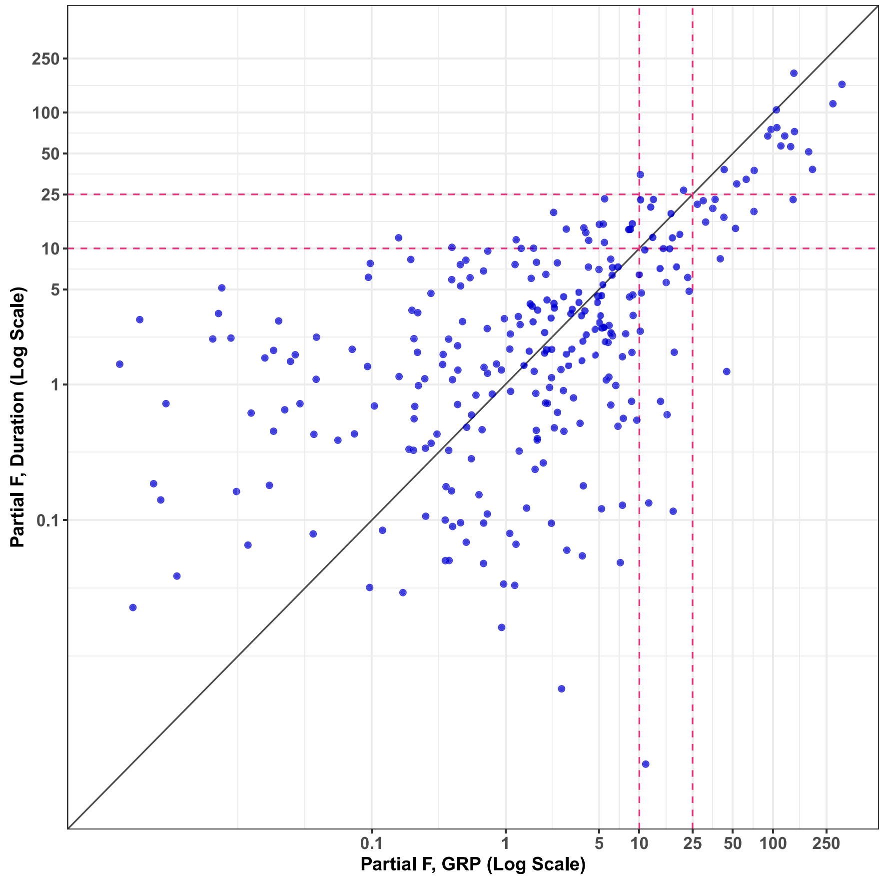
model in duration as such may increase first-stage strength.[33]

Columns (5)–(8) in Table 4 summarizes the first-stage results when both category ads and political ads are measured in duration (minutes). By comparing columns (5)–(8) with columns (1)–(4) in panel (A), we see that the first-stage results using durations are generally weaker. However, panel (B) suggests that the identities of the strongest categories are not changed, although the partial F statistics become smaller. Figure 10 confirms this finding by showing that the strong categories in the GRP-log-linear specification are also relatively strong in the duration-log-linear specification (the correlation is strong and positive in the right tail of the scatterplot). In total, these results suggest that weak first stages in the benchmark model are not driven by the interpolation of viewership that comprises part of GRP measurement.

4.3. Category Characteristics Related to First-Stage Strength

In this section, we examine the characteristics associated with first-stage strength. Figure 11 plots the distributions of four characteristics for each F-statistic bin. Panel (a) shows that larger categories tend to have stronger crowd-out effects. Per Section 2.4, these may be categories with lower willingness to pay on the margin because of an already built-up advertising stock; they may be on the relatively flat part of the advertising response curve. Panels (b) and (c) show that categories that rely more on spot markets (higher spot GRP shares) and advertise more in high-impression periods (higher GRP per occurrence) tend to have stronger first stages. For the $F > 25$, $10 < F < 25$, and $F < 10$ categories, the median spot GRP shares are 0.37, 0.11, and 0.04, respectively, and the median average GRPs per occurrence are 2.46, 2.16, and 1.99, respectively. Because political ads are primarily bought on spot markets (spot share 0.94) and are very eyeball heavy (3.54 GRPs per occurrence), the associations between these two characteristics and first-stage strength are intuitive. The categories that advertise primarily locally likely have less flexibility to substitute to national television because of their inherently local nature (e.g., car dealerships, appliance stores, and hospitals). The association with high-impression day parts may reflect a preference for reach over frequency, as might be the case for

Figure 10. (Color online) Partial F: Duration/GRP



Note. This plot compares columns (5) and (1) of Table 4.

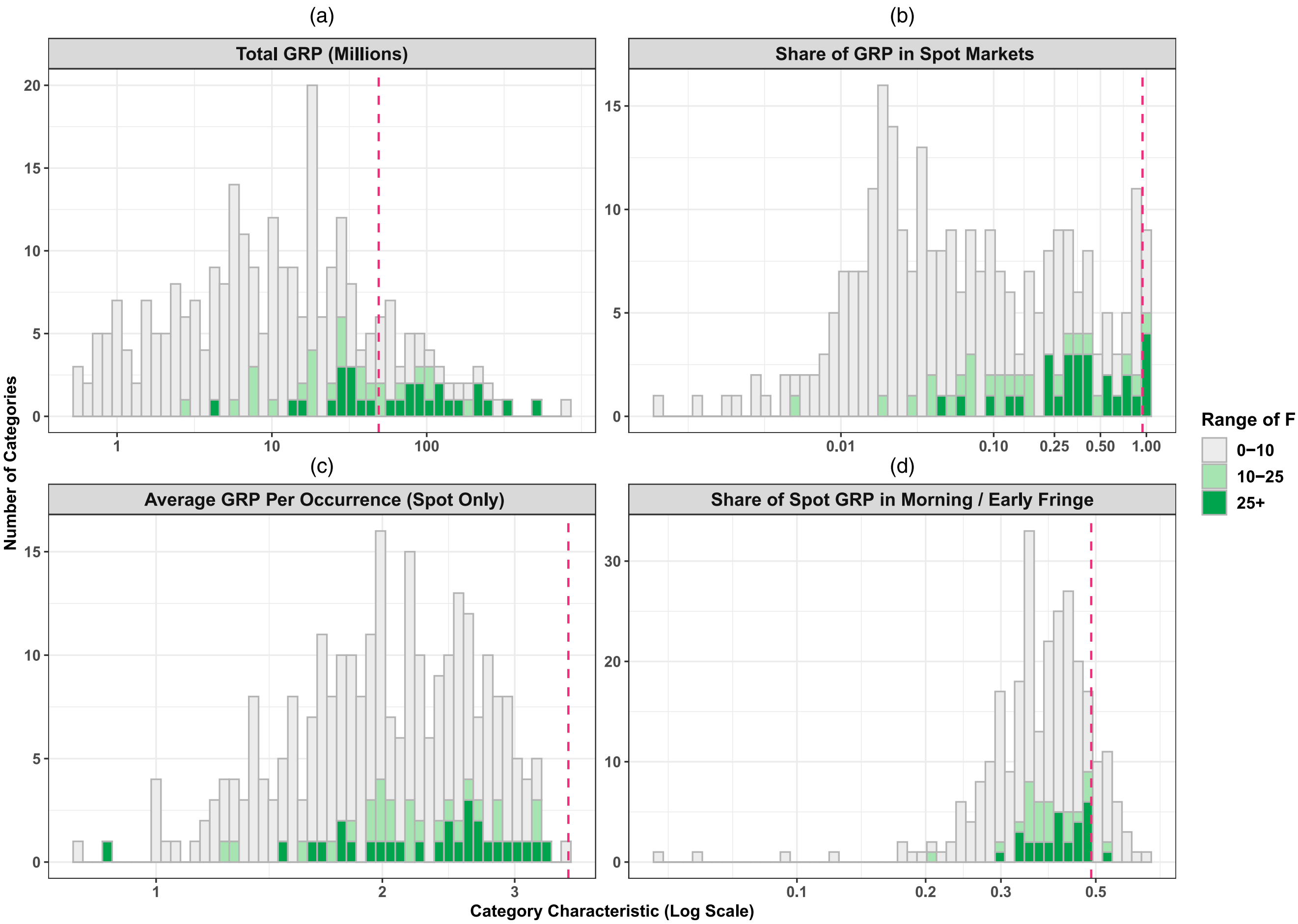
durable products. For these types of goods, frequent reminders may be less useful than broadly distributed information. Because there are relatively few ad slots with high reach, these types of categories likely have less flexibility to substitute their ad dollars elsewhere. Finally, panel (d) plots the share of spot GRPs in the "morning" and "early fringe" (which corresponds to the late afternoon/early evening) day parts, in which political ads are the most concentrated. The shares in those two day parts are slightly larger for the $F > 25$ categories (median 0.41 versus 0.38), but its association with first-stage strength is weak. This pattern might indicate that some categories substitute to other day parts, offsetting low willingness-to-pay advertisers in those day parts indirectly. The first three category characteristics are also listed along with the first-stage results in Table 3 and Online Appendix A, Table 7.

5. Proof of Concept: Auto Advertising

We implement the political advertising IV to understand ad effects for motor vehicles. The purpose of this exercise is twofold. First, we want to characterize the properties of the IV estimator conditional on having a sufficiently strong first stage to limit weak-instrument bias. Second, our prior is that the short-run category-level auto advertising elasticity is likely zero. That is, we believe auto advertising is unlikely to change an individual's need for a new car (although it may change an individual's choice of car, so that advertising is primarily business stealing). Thus, we interpret the IV estimate of the short-run category-level ad effect as a placebo test of the exclusion restriction.

Automotive advertising is large. Dealers and manufacturers spent an estimated $51 billion worldwide in 2017.[34] Automotive advertising has also been studied in the economics and marketing literature (e.g., Murry 2017). Anecdotal evidence from the popular press suggests that political advertising is particularly relevant for car dealerships, so we interpret this case study as a near-ideal case for the theory of the identification strategy.[35] We sum GRPs across manufacturers, dealers, and dealer associations for motorcycles, light trucks, passenger cars, and vans.[36] Figure 12 shows the comovement of GRPs aired by political groups and the automotive industry for four media markets: Columbus, Miami, New York, and Pittsburgh. It is easy to see the ebb and flow of the political cycle across all media markets, where GRPs

Figure 11. (Color online) Distributions of Four Category Characteristics Colored by Partial *F*



Notes. Each observation is a category. Panels (c) and (d) use data from local spot markets only. The vertical dotted lines show the characteristics for political advertising. The "morning" day part is from 6 to 9 a.m., and the "early fringe" day part is from 4:30 to 7 p.m.

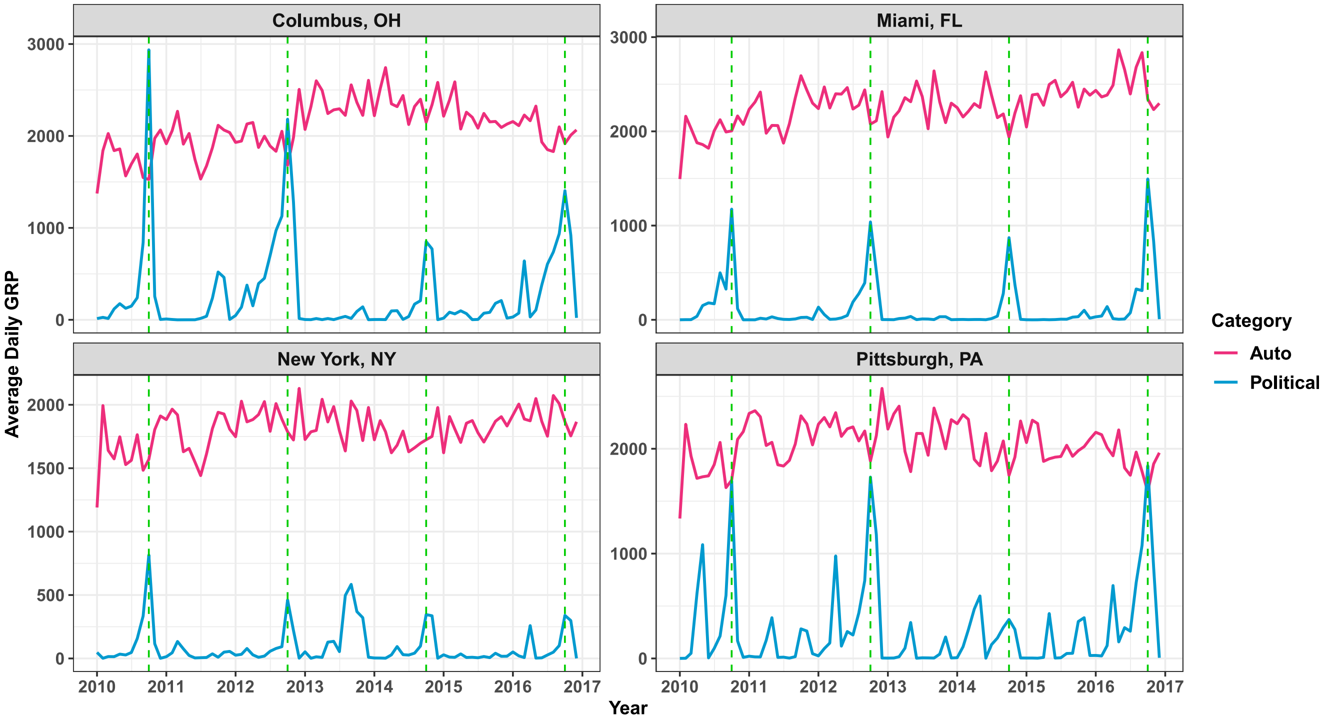
spike in the fall of even years. Advertising cycles appear to be high frequency for auto advertisers but low amplitude. Our sales data come from RL Polk and include the monthly quantity of motor vehicle purchases by buyer zip code from 2010 to 2016. Summary statistics are available in Table 5.[37]

An OLS regression of sales on GRP that includes month × year fixed effects and zip code fixed effects suggests an ad elasticity of 0.108 (Table 6), which is likely spurious because of the usual endogeneity concerns that dealers and manufacturers strategically target advertising at areas with high latent demand for motor vehicles. That association is depicted graphically in Figure 13(a).

To obtain a causal estimate of the effect of category advertising on category sales for automobiles, we turn to the political advertising instrument. Based on our results in Section 4, the auto industry seems amenable to the political ad strategy; six auto categories number among the categories where political advertising has a strong first stage (see Table 3).

Panel (b) of Figure 13 reproduces that first-stage relationship graphically. As predicted by our theory, there is a negative relationship between log political GRPs and log auto GRPs. Panel (c) of the figure shows the reduced form relationship between auto sales and political GRPs that appears nearly flat and is not statistically significant. Table 6 displays the corresponding regression results, which include month × year fixed effects and zip code fixed effects (standard errors are clustered at the DMA level).[38] The preferred specification is in column (4), where we estimate a strong first stage, with a partial *F* statistic of 402, which is far above conventional thresholds for limiting weak-instrument bias. The IV estimate of the ad elasticity is 0.07 and is not statistically significant (column (4) of Table 6). From the perspective of the exclusion restriction, it is reassuring that zero is in the

Figure 12. (Color online) Auto and Political GRPs over Time in Four Example Markets



confidence interval, as we expect the true short-run effect of car advertising on category car sales to be zero. However the 95% confidence interval of [−0.024, 0.16] is reasonably wide and spans a considerable portion of the distribution found in Shapiro et al. (2019). It also contains point estimates obtained from estimating the model with automobile advertising stock (rather than flows) as the explanatory variable. The results using stock are reported in columns (5) and (6) of Table 6, which differ in how political advertising is measured (as a flow and stock, respectively).[39]

We take a few main lessons from this exercise. First, it demonstrates that a very strong first stage does not guarantee a precise estimate in the two-stage least squares regression. In this case, the dependent variable is noisy, even with a battery of fixed effects. Individuals do not buy cars frequently, and the exact timing of their purchases is likely driven by many omitted factors. Our estimated confidence interval includes both zero and meaningfully positive advertising elasticities.[40] Second, that we find a statistically significant short-run effect of advertising on category-level car purchases in the OLS (when the true effect is likely zero) reinforces our belief that the endogeneity of advertising in this setting is a first-order problem. Further, our confidence in the exclusion restriction for political advertising is incrementally increased by the fact that using the IV puts the true effect (zero) into the confidence interval.

6. Conclusions

This paper investigates when and how the political cycle can be used to estimate the causal effect of commercial advertising on sales. Absent a source of quasirandom variation, observational data are likely to yield biased estimates on the return to commercial advertising. We carefully enumerate the necessary assumptions for political advertising to identify causal effects and discuss considerations of which LATE it identifies. Using Ad Intel data from 2010 to 2016, we present descriptive evidence that political advertising increases sharply during election seasons and offsets commercial advertising almost one to one. We also show that the offset constitutes only a small overall change in the aggregate commercial advertising market. We find that political advertising moves commercial advertising levels (first-stage $F > 25$) for 28 of 274 product categories that advertise on television. We show how to use LASSO to obtain optimal instruments and document that it marginally improves

Table 5. Summary Statistics on Auto Sales

	Count	Mean	SD	Min	Max
Sales	1,646,659	37.38261	287.9025	0	76,646
GRPs	1,681,092	55,012.33	11,941.77	18,757.73	108,048.8

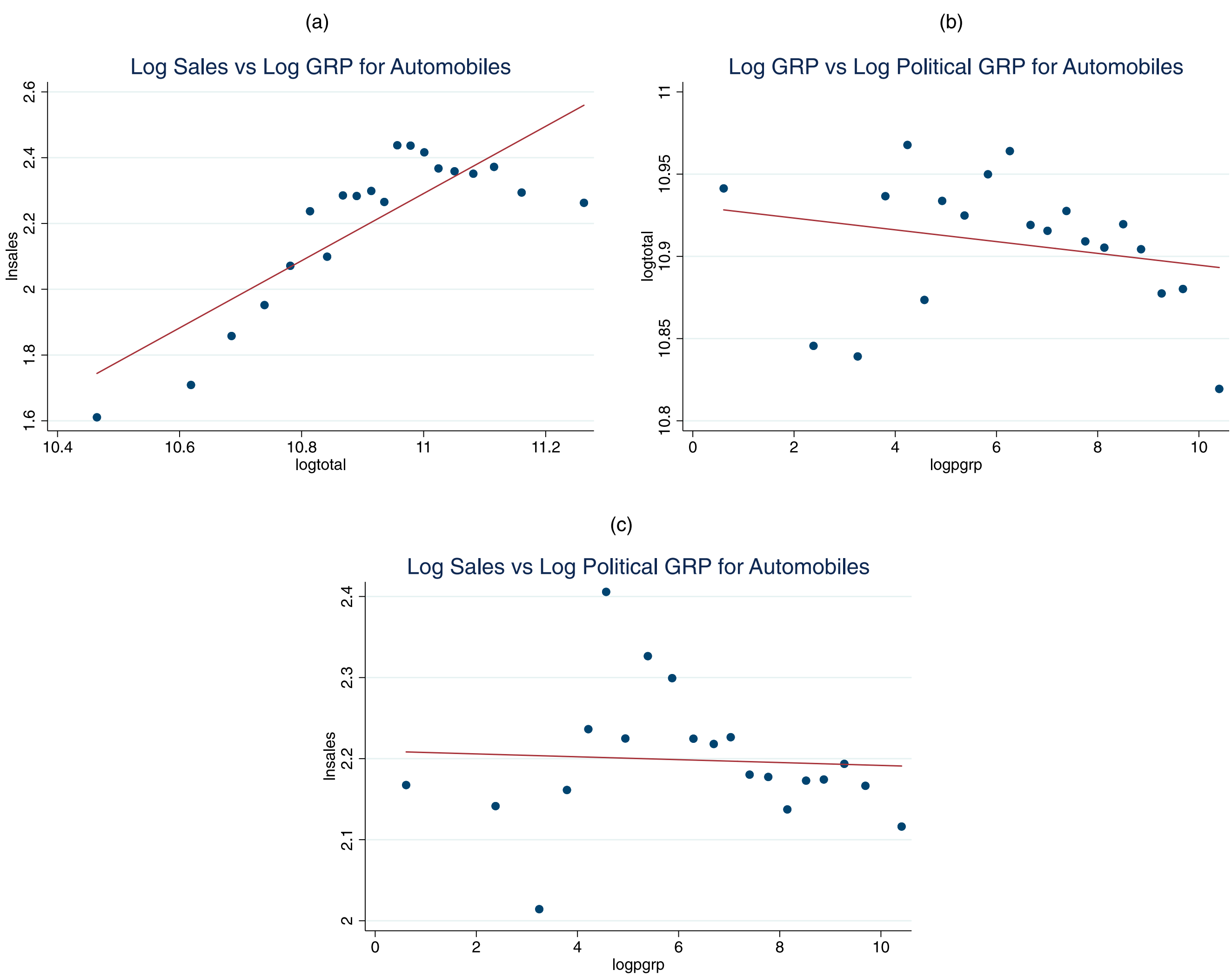
Notes. Summary statistics are on automobile purchases from 24,608 zip codes 2010–2016. SD, standard deviation.

Table 6. Automotive Advertising Effects

	OLS		IV				Optimal IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(*auto GRP*)	0.779	0.108	2.463	0.068			0.054
	(0.163)	(0.048)	(1.447)	(0.047)			(0.049)
log(*stock auto GRP*)					0.022	0.100	
					(0.077)	(0.083)	
N	1,646,415	1,646,415	1,646,415	1,646,415	1,361,372	1,361,372	1,646,415
Instrument			Flow	Flow	Flow	Stock	Flow
FE	No	Yes	No	Yes	Yes	Yes	Yes
First-stage *F*	—	—	15.05	402.85	136.48	142.37	159.59

Notes. Standard errors are clustered at the DMA level and are provided in parentheses. FE indicates DMA and month × year fixed effects. *Instrument* refers to whether political GRPs are measured as a flow or stock variable. The first-stage Montiel Olea and Pflueger (MOP) *F* statistic is reported in the last row.

Figure 13. (Color online) Political Advertising and Auto Sales



Notes. All panels show bin-scatter plots with 20 equipercentile bins. (a) OLS. (b) First stage. (c) Reduced form.

first-stage power for several particularly weak categories.

An important consideration in employing the political advertising instrument is whether the exclusion restriction holds. We highlight two potential threats: first, through the price mechanism, as negative product-market demand shocks might reduce commercial advertising demand, lowering prices for airtime and luring political advertisers; and second, through fluctuations in viewer attentiveness across markets and over time, which might attract both political and commercial advertisers. We argue that integral to the implementation of the political advertising instrument is the inclusion of time and market fixed effects. Using this preferred specification, we estimate the advertising elasticity for the automotive industry using monthly vehicle sales data. Although the confidence interval of the IV estimate contains the presumed truth of zero short-run advertising effect on category-level demand of automobiles, the estimate is imprecise.

Given our results, our subjective recommendation to researchers interested in employing political advertising as an instrument is threefold. First, given that political advertising is a weak instrument in the vast majority of cases, researchers should consider using weak identification robust inference. Second, researchers may want to avoid the instrument in cases where the first stage is the "incorrect" sign because it suggests that the theory underlying the validity of the IV approach is violated. In these cases, political advertising and commercial advertising are related in some way other than crowd out. It could be that there is substitution between markets or positively correlated preferences for particularly attentive viewers occurring in conjunction with crowd out, violating monotonicity. Alternatively, there could be some path from unobservable demand shocks to both political advertising and commercial advertising, violating the exclusion restriction. Although we stop short of recommending complete avoidance of the instrument for the categories where LASSO returns zero terms, it signals to the researcher that the instrument is particularly weak. Although Belloni et al. (2012) provide a way forward in such cases, it is unlikely that the political cycle will produce an informative IV. Researchers may wish to find other sources of variation in such cases. Finally, we want to emphasize that the exclusion restriction might be difficult to justify in some cases. Online Appendix D documents that for 36 categories of grocery products, 5 of them show signs of potential exclusion restriction violations. When employing this instrument, researchers and practitioners should make sure to carefully consider the possible exclusion restriction violations and decide how likely such violations are in their particular case. It should not be taken for granted that the instrument satisfies the exclusion restriction in all cases.

Acknowledgments

The authors thank Sylvia Hristakeva, Mitch Lovett, Sanjog Misra, and participants at the 2019 Marketing Science Conference and the 12th Workshop on the Economics of Advertising and Marketing for valuable feedback. Results calculated (or derived) were based on data from The Nielsen Company (United States), 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 Company data are those of the researchers and do not reflect the views of The Nielsen Company. The Nielsen Company is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Endnotes

[1] See https://adage.com/article/media/2016-political-broadcast-tv-spend-20-cable-52/307346/.

[2] We specify advertising as a log because researchers typically want to identify an advertising elasticity, where the independent variable of interest is the log of advertising and the dependent variable of interest is the log of quantity.

[3] This includes using machine learning as in Belloni et al. (2012) to find the functional form that gets the strongest possible first stage.

[4] The distribution of first-stage F statistics that we recover raises a question of inference with weak instruments. The F statistic for many categories falls near the critical value of 10, proposed by Stock and Yogo (2005), although more recent work highlights that this value may be too permissive (for example, Olea and Pflueger 2013). We suggest that researchers justify their choice of critical values on a case-by-case basis, depending on the research question. If the first-stage relationship appears weak, as measured by the F statistic, rather than disregard the political advertising instrument, researchers should consider using weak identification robust inference methods (e.g., Anderson and Rubin 1949, Andrews 2016).

[5] Our results in this case study also underscore that the problem of weak instruments is one of bias rather than precision. A strong instrument circumvents bias because of weak instruments, but it need not increase precision.

[6] For example, researchers may interact political advertising with brand fixed effects. This would require an assumption that the amount of displacement on one brand is independent of the amount of displacement on a rival brand. This might be reasonable in some cases, but it is not an assumption we are willing to apply broadly across 274 categories for the purposes of this demonstration. Sinkinson and Starc (2020) solve this problem by using political advertising in conjunction with a temporary ban on advertising for the popular product, Lipitor, to identify brand-level advertising effects.

[7] Bartik instruments originate in Bartik (1991). The idea is to instrument for employment growth rate with the product of national industry growth rates with local industry shares to estimate the elasticity of labor supply. Because the endogeneity is posited to occur on the local industry growth *rates*, this instrument satisfies exclusion.

[8] Note that this condition does not require a hard capacity constraint, as long as the marginal cost of advertising time increases in the quantity of ads. In theory, local television stations have some flexibility in the airtime they devote to advertising compared with programming (e.g., run shorter or fewer local news segments). If viewers exhibit increasing disutility of advertising time, then the presence of

political ads increases the marginal cost of commercial airtime. In our sample, we find very little variation in total advertising time both across stations and over time, suggesting that crowd out of commercial advertising is nearly one to one with the presence of political ads.

[9] Here, it is worth pointing out that Lovett et al. (2019) employs a version of this instrument in studying the effect of advertising on word of mouth. Instead of estimating effects category by category, it estimates a single effect of advertising on word of mouth with a random coefficient. The instruments employed are political advertising interacted with brand intercepts. Overall, it finds that the set of instruments has a weak first stage in that regression and produces several positive coefficients in the first stage, an indication that there might be some monotonicity violations. We will break apart the problem in detail to see exactly where the instrument is weak and where the theory might fail.

[10] This is shown graphically in Section 3.2.

[11] We do not adjust our inference for multiple hypothesis testing.

[12] Here, we note that Lovett et al. (2019) find many first-stage coefficients that are the wrong sign. That paper also does not employ time fixed effects at the periodicity of the data because of limitations in its data, which we suggest could explain some of the apparent violations of the theory.

[13] A positive correlation would indicate that a nonprice mechanism was contributing to the first stage, either by itself or in conjunction with the posited price mechanism. In the former case, it is likely that more political advertising increases commercial advertising in some circumstances while reducing commercial advertising in others. This would be a violation of monotonicity. Alternatively, a positive first-stage relationship could be driven entirely by a nonprice mechanism that violates the exclusion restriction.

[14] Category-level average WTP may change between different political cycles if the nature of competition in those categories changes over time. For example, in one election, it may be that a category has a new entrant firm that has a very high WTP for advertising space because of a need to inform customers of its existence. In a future election, that firm is no longer a new entrant and as a result, has a lower WTP for advertising space. In such an example, the category-level first stage could increase in a future example. These kinds of considerations are important for researchers to keep in mind when they interpret the LATE associated with the political advertising IV.

[15] We focus on estimating the effect of advertising flow even though advertising is typically thought to have accumulating stock effects. If, in August, firms anticipate higher prices of advertising in October because of political campaigns, they may increase advertising in August to take advantage of stock effects that linger into November. Such anticipation and dynamic effects would pose problems for the validity of this specification. If we specify our model with an accumulating advertising stock as the endogenous variable to account for this problem, the dynamic behavior would serve to weaken the first stage, as current period crowd out would be partially offset by previous period anticipatory effects. In Online Appendix E, we specify the model using advertising stock as the endogenous variable of interest and find that the instrument is generally weaker. As a result, the main analysis should be viewed as perhaps an overoptimistic view of the strength of the first stage.

[16] The utility of the outside option is normalized to zero.

[17] In practice, because of repeated observations in the outcome, practitioners will typically want to cluster standard errors at the market level. For that reason, we cluster at the market level throughout. In either case, the relevant clustering for the brand-level formulation and the category-level formulation is the same, as the relevant quasiexogenous variation is at the same level.

[18] The data are missing for North Platte, Nebraska DMA in June 2011 and for Yakima-Pasco-Richland-Kennewick, Washington DMA in December 2014.

[19] See "Nielsen Monitor-Plus Methodology by Medium," page 19 at http://en-us.nielsen.com/sitelets/cls/documents/adviews/AdViews-Methodology-by-Medium-InfoKit.pdf.

[20] These include spot TV, network TV broadcast locally, and syndicated TV broadcast locally.

[21] In 2016, Nielsen started providing impressions data in all months for 70 markets, including the 25 LPM markets and 45 "Set Meter" or "Code Reader" markets.

[22] The impressions data are missing for Birmingham (Ann and Tusc), Alabama DMA in May 2011.

[23] We weight the data from two closest sweep months by time difference. For example, for March, we use two-thirds February and one-third May.

[24] For more description of processing the Nielsen Ad-Intel database and the data computed from it, please see Shapiro et al. (2019).

[25] See https://www.opensecrets.org/pacs/superpacs.php.

[26] We redefine the "B189: Miscellaneous Organization Advertising" category after taking out the 616 identified super PACs.

[27] Calculation based on Figure 6 and the point estimates in Table 2 shows that about 7% of total commercial advertising was offset in Columbus, Ohio, in the month with the most political advertising (October 2010). In all other months and in most other markets, the offset is considerably lower.

[28] Ex post, the linear term is the most frequently selected term in the LASSO specification, suggesting that it is typically the best single first-stage predictor in terms of functional form.

[29] It might be suggested that we adjust our F statistics to account for the fact that we are testing 274 hypotheses at one time. On one hand, we are indeed testing 274 categories at the same time in this paper, so the adjustment may seem necessary. On the other hand, other researchers using this instrument will likely focus on one category at a time, so the unadjusted F statistics should be the objective of interest. To the extent that the reader would prefer adjusted F statistics, our F statistics reported here would represent an upper bound on the adjusted F statistics.

[30] Intuitively, we employ this approach to make sure we choose a functional form of political advertising to get every last bit of available strength out of the first stage. If we specify a linear model and the first stage is highly nonlinear, we may find a weak instrument, but there exists a specification where that instrument might be strong. Conducting this LASSO exercise comes at a cost, as it requires further assumptions and may make final estimates difficult to interpret in the presence of heterogeneous treatment effects. For the purposes of our exercise, the LASSO is simply meant to demonstrate how much additional strength could be added to the first stage. Authors should carefully consider the additional required assumptions of the LASSO if they choose to employ such a formulation in order to increase first-stage strength.

[31] Further analysis in Online Appendix B.3 shows that the LASSO is indeed picking up crowd-out relationships that are very different from log linear for those 28 "upgraded" categories, though for some categories, the first-stage improvement comes at the cost of monotonicity concerns.

[32] The number of optimal instruments selected is one for 49 categories, two for 25 categories, three for 11 categories, four for one category, and five for one category. The linear term $f(P_{mt}) = P_{mt}$ is most likely to be selected: it is the optimal instrument for 19 categories, and numbers among the terms selected for another 31 categories.

This result supports our choice of the log-linear specification as the benchmark and suggests that the crowd out is stronger when the level of political advertising is high.

[33] A final reason to show the duration analysis is that some researchers may have access to advertising occurrence data, but not viewership data, and this would provide a guide in such cases.

[34] See https://adage.com/article/cmo-strategy/world-s-largest-advertisers-2017/311484.

[35] See Higgins (2016).

[36] The seven categories are T111 "Passenger Cars–Factory, New & CPO"; T112 "Light Trucks & Vans–Factory, New & CPO"; T115 "Motorcycles and Misc. Vehicles–Factory, New"; T121 "Passenger Cars–Dealer Association, New & CPO"; T122 "Light Trucks & Vans–Dealer Association, New & CPO"; T161 "Autos and Light Truck–Dealerships"; and T163 "Motorcycles and Misc. Vehicles–Dealerships."

[37] The data come from vehicle registrations.

[38] We use zip code fixed effects instead of DMA fixed effects because of the fact that we have outcome data at the zip code level. Zip code fixed effects reduce more noise in the dependent variable than do DMA fixed effects.

[39] Note that we omit price in our regression specification (Equation (1)) because we aim to estimate the total effect of advertising, which is potentially mediated through price. That is, if advertising affects pricing, then price is a "bad control" (Angrist and Pischke 2009). If our goal was to estimate a "partial equilibrium" effect of advertising that held prices fixed, we would instead include price as a covariate, requiring a second instrument for price.

[40] This further highlights that the point of testing for weak instruments is limiting finite sample *bias* and not about ensuring precision in the IV estimate.

References

Anderson TW, Rubin H (1949) Estimation of the parameters of a single equation in a complete system of stochastic equations. *Ann. Math. Statist.* 20(1):46–63.

Andrews I (2016) Conditional linear combination tests for weakly identified models. *Econometrica* 84(6):2155–2182.

Angrist JD, Imbens GW (1995) Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *J. Amer. Statist. Assoc.* 90(430):431–442.

Angrist JD, Pischke J-S (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Bartik TJ (1991) *Who Benefits from State and Local Economic Development Policies?* (W. E. Upjohn Institute for Employment Research, Kalamazoo, MI).

Belloni A, Chen D, Chernozhukov V, Hansen C (2012) Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80(6):2369–2429.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.

Borusyak K, Hull P, Jaravel X (2018) Quasi-experimental shift-share research designs. Preprint, submitted June 4, https://arxiv.org/abs/1806.01221.

Dubois P, Griffith R, O'Connell M (2018) The effects of banning advertising in junk food markets. *Rev. Econom. Stud.* 85(1):396–436.

Gandhi A, Houde J-F (2019) Measuring substitution patterns in differentiated products industries. Working Paper 26375, National Bureau of Economic Research, Cambridge, MA.

Goldsmith-Pinkham P, Sorkin I, Swift H (2018) Bartik instruments: What, when, why, and how. NBER Working Paper 24408, National Bureau of Economic Research, Cambridge, MA.

Hartmann WR, Klapper D (2018) Super bowl ads. *Marketing Sci.* 37(1):78–96.

Higgins T (2016) No one hates political ads more than car dealers. *Bloomberg News* (January 8), https://www.bloomberg.com/news/articles/2016-01-08/no-one-hates-political-ads-more-than-car-delaers.

Li X, Hartmann WR, Amano T (2019) Identification using border approaches and IVs. Preprint, submitted June 19, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3402187.

Lovett MJ, Peres R, Xu L (2019) Can your advertising really buy earned impressions? The effect of brand advertising on word of mouth. *Quant. Marketing Econom.* 17(2019):215–255.

Murry C (2017) Advertising in vertical relationships: An equilibrium model of the automobile industry. Preprint, submitted January 25, http://dx.doi.org/10.2139/ssrn.2549247.

Olea JLM, Pflueger C (2013) A robust test for weak instruments. *J. Bus. Econom. Statist.* 31(3):358–369.

Rossi PE (2014) Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Sci.* 33(5):655–672.

Shapiro B, Hitsch GJ, Tuchman A (2019) Generalizable and robust TV advertising effects. Preprint, submitted January 20, http://dx.doi.org/10.2139/ssrn.3273476.

Shapiro BT (2018) Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *J. Political Econom.* 126(1):381–437.

Sinkinson M, Starc A (2019) Ask your doctor? Direct-to-consumer advertising of pharmaceuticals. *Rev. Econom. Stud.* 86(2):836–881.

Stock J, Yogo M (2005) Testing for weak instruments in linear IV regression. Andrews DWK, ed. *Identification and Inference for Econometric Models* (Cambridge University Press, New York), 80–108.

Thomas M (2018) Spillovers from mass advertising: An identification strategy. Preprint, submitted May 31, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3182092.

Tuchman AE (2019) Advertising and demand for addictive goods: The effects of e-cigarette advertising. *Marketing Sci.* 38(6):994–1022.