informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–19
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Preference Externality Estimators: A Comparison of Border Approaches and IVs

Xing Li,[a] Wesley R. Hartmann,[b,*] Tomomichi Amano[c]

[a] Peking University Guanghua School of Management, Marketing, Beijing 100871, China; [b] Stanford University, Stanford, California 94305; [c] Harvard Business School, Boston, Massachusetts 02163

*Corresponding author

Contact: xingli@gsm.pku.edu.cn (XL); wesleyr@stanford.edu, https://orcid.org/0000-0002-0211-2386 (WRH); tamano@hbs.edu, https://orcid.org/0000-0003-3594-5031 (TA)

Received: October 9, 2021
Revised: September 13, 2022
Accepted: November 28, 2022
Published Online in Articles in Advance: January 23, 2024

https://doi.org/10.1287/mnsc.2023.4977

Abstract. This paper compares two estimators—the Border Approach and an Instrumental Variable (IV) estimator—using a unified framework where identifying variation arises from "preference externalities," following the intuition in Waldfogel (2003). We highlight two dimensions in favor of the IV approach. First, an econometric model of the data-generating process reveals that the border approach requires a set of identification assumptions that are not easily satisfied in practice: the ignorance of some payoff-relevant information and conflicting spatial correlation assumptions. The IV approach, in contrast, exhibits greater internal validity because it is derived from the model that generates the data. Second, the border approach suffers from representative issues when the true effect sizes are different between border and off-border regions. We use a common political advertising example to evaluate these estimators and suggest ways to evaluate or limit the above concerns, such as excluding localities that are a large share of the policy making region and evaluating spatial correlations of observables. We find the border approach's representative issue to be substantial when the ignorance assumption is most plausible and observe that spatial correlations do not reflect those needed in the unobservables for consistency of the estimator. The IV, in contrast, does not exhibit concerns related to local average treatment effects. We also derive the specific conditions when the border approach can reduce bias relative to OLS.

History: Accepted by Raphael Thomadsen, marketing.
Funding: This work was supported by National Natural Science Foundation of China (NSFC) [72072004, 72131001].
Supplemental Material: The data files are available at https://doi.org/10.1287/mnsc.2023.4977.

Keywords: marketing • advertising and media • economics • econometrics • causal inference

1. Introduction

Two otherwise comparable individuals can face different policies, product offerings, prices, or other treatments when their geographic markets or other group identities differ. When governments and firms incorporate group composition in their policymaking, the treatment any single individual experiences is affected by preferences of the rest of the group. This "preference externality," as coined by Waldfogel (2003), forms the basis of two causal inference approaches: a preference externality-based instrumental variable (PEIV) (Gentzkow and Shapiro 2010, Fan 2013, Berry and Haile 2022) and the border approach (Card and Krueger 1994, Dube et al. 2010, Shapiro 2018). The former is derived explicitly from a model of the data-generating process, whereas the latter is typically cast as "model-free" yet imposes its own set of assumptions. We compare these estimators.

We begin with an economic model under a layering structure in which policy decisions are made at a more aggregate level than individual outcomes. The model formalizes the source of "preference externalities" that motivate both estimators. We show that the PEIV is consistent with the behavior implied by the model, whereas the border approach instead assumes that the decision maker ignores payoff relevant information—a behavioral assumption we term the *ignorance assumption*. The ignorance assumption is a necessary condition for the statistical *unconfoundedness* that guarantees the validity of the border approach.

Establishing statistical unconfoundedness in the border approach is not achieved solely by assuming ignorance. It requires an additional *zero spatial correlation assumption* between contiguous units of regions lying off the border. Because the merit of the cross-border comparison in the border approach is the existence of unobserved common components between contiguous units crossing the border, this assumption imposes another ad hoc restriction that spatial correlation exists only across borders but not elsewhere off the border. We suggest evaluating the spatial correlation of observables to evaluate the plausibility of both i) zero spatial correlation between borders and regions influencing policies and ii) strong cross-border spatial correlations. In summary, the border approach relies on a series of behavioral and econometric assumptions that are challenging to justify in common empirical applications, particularly in the recent analysis of firms as policymakers where ignorance implies sacrificing profits.

One alternative assumption to validate the border approach is through the econometrics of discontinuities, that in the counterfactual without the treatment of interest, the outcome is continuous across the border. However, we choose the econometric assumptions of unconfoundedness for two reasons. First, implementations using the border approach often do not isolate inference locally at the discontinuities but rather equally weight all individuals within an administrative unit adjacent to a border, thereby retaining biases that likely increase with individuals' distances from the border.[1] It is more reasonable to interpret the border approach as a matching estimator using nearest neighbors in the geographic sense, thus in the unconfoundedness framework. Second, because the border approach has extended to competitive contexts, discontinuity econometrics cannot apply because competition typically involves the discontinuity of more than one competitor's action at the border, leaving no way other than unconfoundnedness to attribute a discontinuous outcome for one competitor to a discontinuity in its own actions versus its competitors'.[2]

Aside from the ability to resolve endogeneity concerns under certain behavioral models, the two estimators also present different trade-offs regarding statistical power and representativeness. The PEIV may lack statistical power when instruments are weak (Bound et al. 1995, Rossi 2014), and the local average treatment effects (Imbens and Angrist 1994, Masten and Torgovitsky 2016) can raise representativeness concerns. The border approach focuses on a subset of markets located on borders, so it reduces the sample size and can therefore also sacrifice statistical power. Moreover, borders may be unique, thereby raising concerns about the representativeness of estimated effects. This may be particularly true in "small" counties where the ignorance assumption is more plausible. Therefore, the question arises as to whether the border estimator's power and representativeness concerns are sufficiently small, relative to the IV approach, to justify its less appealing identifying assumptions.

We compare the performance of the two approaches in the context of U.S. presidential election advertising, where the advertising exposures observed in a county may include externalities from other counties in the same media market and potentially a different state. For example, Huber and Arceneaux (2007) pointed out that the strong advertising incentives in "battleground states" can spill over to non-battleground states if they share a media market. We use Gordon and Hartmann (2013)'s presidential advertising data and analysis as a baseline for comparing the preference externality estimators. An important advantage of these data is that they allow us to set econometric endogeneity biases aside to evaluate the representativeness concerns that otherwise are inseparable from endogeneity biases when comparing estimators. Gordon and Hartmann (2013) argued that endogeneity concerns are not severe because a candidate's advertising decreases with unobserved demand shocks when leading its competitor in a market but increases when trailing; that is, a mix of positive and negative biases may be offsetting. Their estimates support this argument because a Hausman test cannot reject the null that the fixed effect estimate is the same as an estimate using supply-side instrumental variables.

We find limited evidence of representativeness concerns for the PEIV but substantial evidence for the border approach. When we compare the fixed effect estimate of presidential advertising effects with the PEIV, the latter is 13% higher but, as in the case of supply-side IVs, a Hausman test does not reject the null that effects are the same. When we combine observable preference externalities with Gordon and Hartmann (2013)'s supply-side IVs, the estimates are identical to the fixed effect (a Hausman *p* value of 0.993).

We evaluate representativeness for two versions of the border approach: one using all border counties as in Shapiro (2018) and the other with less-populated "small" border counties where the ignorance assumption may be more plausible.[3] When we compare the border approach using all border regions to the fixed effect, the estimate is 23% smaller. It does not appear to be statistically different, but a formal test is elusive because of the oversampling of some border counties in the border approach. However, if we include only border regions that are more plausibly ignored by decision makers, that is, small counties, the point estimate drops by 83% to near zero and becomes insignificant despite a standard error that is identical to the lowest IV specification. We further provide evidence that the difference in effects can be fully accounted for by the border regions' representativeness concerns, because other estimators applied to only these small counties also produce near-zero insignificant ad effects.

To the question of statistical power, we find that neither IVs nor the border approach increase standard errors enough to yield insignificant estimates. In our empirical context, the preference externality-based IVs are fairly strong, with a first-stage excluded F of 48.7 and a partial R^2 of 0.48. When we include both demand and supply-side determinants of advertising in the instruments, the first stage F increases to 56.9 with a partial R^2 of 0.64, yielding standard errors midway between the fixed effect and using either instrument strategy alone. Statistical power does not seem to be a concern with the border approach either. The largest standard errors in the border approach occur when we use only the small counties where the identifying assumptions are more likely to hold, but as stated above, these standard errors are rather small. We suspect that the modest losses in statistical power despite dropping so many county-level observations result from the fact that the variation of advertising comes at the DMA level that is mostly retained in the subsample of border counties.

In summary, the PEIVs exhibit greater internal validity (they assume no ignorance) and empirically produce more representative effects than border implementations. We hope our discussion and analysis encourage future work to consider the IV approach over the border approach when the determinants of preference externalities are observable. For cases when IVs are unavailable, we propose methods to assess the validity of the border approach's identifying assumptions as well as an implementation of the border approach that demands less ignorance of the decision maker. These methods rely on studying the behavior of observables and may not be conclusive in characterizing the nature of unobservables. We have also defined conditions when the border approach is less biased than OLS. If those conditions do not hold or are difficult to assess, or if IVs are hard to find and justify, a descriptive OLS with fixed effects may be a reasonable alternative when accompanied by cautions about causal interpretations. In fact, in our context, a fixed effects analysis likely would have sufficed to generate causal estimates, with the IVs simply making it easier to argue for a causal interpretation.

More broadly, we hope our analysis encourages researchers applying "model-free" approaches to articulate the identifying assumptions imposed by a behavioral model that plausibly generates the data. This paper places a greater emphasis on identifying assumptions of the border approach for exactly this reason. Its assumptions in the presence of a model had not been articulated, whereas the PEIV derives from a model where the identifying variation is explicitly defined. Finally, we comment on the applicability of the estimators for firm decisions and the relevance of preference externalities in the age of "micro-targeting" at the end of our paper.

The remainder of this paper is structured as follows. In Section 2, we develop an illustrative econometric model of advertising decisions and demand response within a layering structure to motivate our analysis of group-level policymaking but local measurement of effects. Section 3 discusses the different identification strategies utilizing this layering structure for identification. Section 4 presents the empirical application and results, and the final section concludes the paper.

2. An Advertising Model in a Layering Structure

Preference externalities arise when heterogeneous individuals face a common policy. In advertising, for example, a group of customers is exposed to the same ads if they belong to the same group targeted by the focal advertiser. Studies of advertising response exploit such a layering structure whereby the unit of analysis for conversion (an individual, zip code, county, etc.) is smaller than the policymaking region (a state or designated market area, a set of counties that define media markets). We define a general model of advertising allocation and conversion (analogous to a supply- and demand-side model) in which allocation decisions occur at a more aggregate level than conversions.

We begin by formalizing the layering structure commonly used in advertising studies. Let $L = \{l_1, l_2, \dots, l_M\}$ be a partition of individuals (e.g., consumers), representing the layer in which advertisers make advertising decisions, and $K = \{k_1, k_2, \dots, k_C\}$ be a more granular partition by which the researcher analyzes conversion. By granular partition, we mean for any $k \in K$ and $l \in L$, either $k \subset l$, or $k \cap l = \emptyset$. For example, if ads were targeted by binary genders and we observed individual level conversion, each partition $k \in K$ includes only one consumer, and $L = \{l_1, l_2\}$, where l_1 is the set of female consumers, and l_2 is the set of male consumers. Groupings can be based on demographics such as geographic location, gender, age, past purchases, or any combination of the observed characteristics. Groupings can also be based on factors unrelated to individual types, such as search terms or any other targeting variables. For example, one group could be a time period when a given advertising intensity is held fixed. Typically, grouping in layer-K is determined by the data availability, whereas the grouping in layer-L is determined by ad technologies or institutional features.

In this study, we focus primarily on the case of DMA-based advertising decisions with observed county-level conversion, such that K is the set of counties and L is the set of DMAs.

2.1. Advertising Exposures and Decisions

Individuals' conversion decisions are influenced by the level of advertising they are exposed to, A_k, as well as x_k and ξ_k, which are determinants of the choice that are, respectively, observable and unobservable to the researcher. The share of individuals in group k who convert can be denoted as a general function

$$q_k(A_k, x_k, \xi_k). \tag{1}$$

No restrictions are made on the distribution of ξ_k. In the context of television advertising that targets different geographies, it may be that ξ_k is correlated across partitions (counties).

Individuals' exposure to advertising is a function of the advertiser's decisions as well as other determinants of exposure. It is denoted as

$$A_k(d_{l(k)}, x_k, \tilde{v}_k),$$

and determined by the coarser-level advertising decision d_l by the advertiser and individual-level characteristics x_k as well as unobserved determinants $\tilde{v}_k$ that are unknown to advertisers when making the advertising decisions.

The advertiser makes advertising decisions d_l to each group $l \in L$ with information set $(X_l, \Xi_l) = (\{x_k\}_{k \subset l}, \{\xi_k\}_{k \subset l})$,

$$d_l(X_l, \Xi_l, w_l), \tag{2}$$

as well as w_l, the marginal cost of advertising.[4] More specifically, the advertiser may maximize its total payoffs[5]

$$\max_{d_{l_1}, \ldots, d_{l_M}} \pi(Q, N, P) - \sum_{l \in L} \sum_{k \subset l} w_l \cdot A_k, \tag{3}$$

where $Q = \{q_k(A_k, x_k, \xi_k)\}_{k=1}^{C}$ is the vector of k-level conversion shares, $N = \{n_k\}_{k=1}^{C}$ is the vector of respective populations' sizes, and P is a set of variables determining how conversions translate to payoff. In the case of a profit-maximizing firm, we might expect P to include prices, p_k, and marginal cost, c, such that $\pi = \sum_{l \in L} \sum_{k \subset l} n_k((p_k - c) q_k(A_k, x_k, \xi_k))$. In the political advertising context that we study below, Gordon and Hartmann (2016) defined $\pi = P \cdot W(Q, N)$, where P is the "returns to winning" the election, and winning status, $w(Q, N)$, is a function of how county-level vote shares are aggregated in either the electoral college or a counterfactual popular vote. Like Gordon and Hartmann (2016), the political context allows us to focus on advertising as the endogenous choice and abstract away from jointly determined advertising and prices.

2.2. Preference Externalities

Preference externalities arise because local-level advertising, A_k, is affected by the external determinants of preferences $X_{-k} = X_l \backslash \{x_k\}$ and $\Xi_{-k} = \Xi_l \backslash \{\xi_k\}$ through ad allocation decisions at the aggregate level, d_l.[6] We see this when we combine the above ad exposure and ad decision models

$$A_k(d_l(x_k, X_{-k}, \xi_k, \Xi_{-k}, \cdot), \cdot, \cdot). \tag{4}$$

This also exposes the identification challenge; the conventional endogeneity bias in advertising studies emerges from unobserved shocks, ξ_k, that affect ad exposures A_k through ads decisions d_l.

2.3. Empirical Specification

Building on this general notation, we describe a simple empirical model of conversion choices that we can take to data. We assume a discrete choice model for individual i residing in county k facing a single advertiser, that is,

$$u_{ik} = \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k + \epsilon_{ik}, \tag{5}$$

where u_{ik} is the utility level, and $g(\cdot)$ is a function that translates advertising levels to utility. Individual i will convert if and only if he or she gains positive utility, so the county-level market share is

$$q_k = \int I(u_{ik} > 0) f(\epsilon_{ik}) d\epsilon_{ik}.$$

Assuming logistically distributed ϵ_{ik}, this model implies shares

$$q_k = \frac{\exp(\alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k)}{1 + \exp(\alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k)},$$

or equivalently, an equation we can take to data,[7]

$$y_k = \ln(q_k) - \ln(1 - q_k) = \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k. \tag{6}$$

We use a static model for illustration, but similar characterizations would exist in a dynamic advertising model.

3. A Comparison of Identification Strategies

The above model on preference externalities motivates two approaches for identification: one implementing an instrumental variable estimator, the preference externality instrumental variable (PEIV), and the other focusing on cross-border comparisons, the border approach. The PEIV is naturally motivated when a researcher has access to some observables X_{-k} that can drive advertising exposure independent of ξ_k, with the assumption that

$$E(\xi_k | X_{-k}, x_k) = 0.$$

The border approach works on an alternative assumption that the endogenous components of the ad decision can be differenced out across some neighboring geographic regions. In this section, we focus on the empirical model in (6) to describe existing advertising identification strategies and evaluate corresponding identifying assumptions in the context of preference externalities.

Table 1 summarizes representative advertising studies leveraging preference externalities, namely the PEIV and border approaches, with specific focus on the different layers where ad decisions and conversions reside.[8]

3.1. Preference Externality Instrumental Variables

The preference externality IV explicitly integrates the externality we laid out in (4) to form the first stage of a

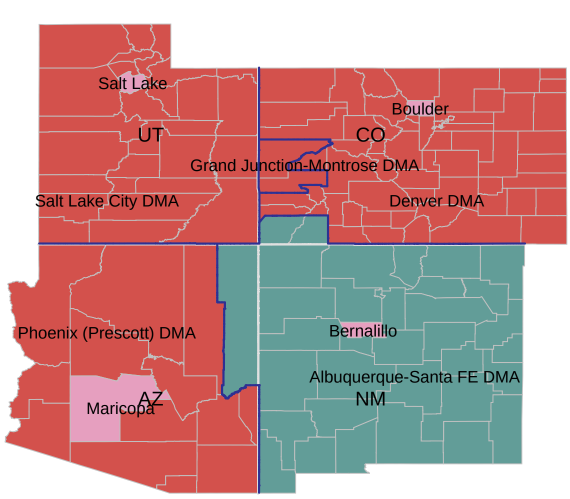
Table 1. Extant Advertising Papers Using Preference Externalities

Approach	Paper	Level of ad decision	Level of conversion analysis
Border	Huber and Arceneaux (2007)	DMA	Individuals in state-DMA/sub-DMA
Border	Shapiro (2018)	DMA-month/week	(DMA-border/sub-DMA)-month/week
Border	Spenkuch and Toniatti (2018)	DMA/year	Border county – year
PEIV	Thomas (2020)	Nation-week	DMA-week
PEIV	Gordon et al. (2023)	DMA	County

two-stage least squares estimator. Counties that might otherwise be identical receive different levels of advertising because they reside in different DMAs, and the ad levels in these DMAs are influenced by observables from other counties in their DMAs, that is, X_{-k}. This approach is applied in non-advertising contexts by Gentzkow and Shapiro (2010) and Fan (2013) and generalized in the context of differentiated products by Berry and Haile (2022). Not realized by us at the time or perhaps subsequently, Thomas (2020) developed a preference externality IV that focused on national advertising and DMA level measurement as opposed to the DMA advertising and county-level measurement that we consider.[9]

To set ideas, consider Figure 1, which depicts the Four Corners states (Utah, Arizona, New Mexico, and Colorado) in the American Southwest. The environment in four purple-shaded urban counties in these states, namely Salt Lake, Maricopa (Phoenix), Bernalillo (Albuquerque), and Boulder (Denver), may be comparable but receive different advertising because of the externalities that are realized from other parts of their DMAs. At the time of the 2004 presidential elections, Utah, Arizona, and Colorado were considered right-leaning (Republican) states (depicted in red) in which election outcomes were predicted to favor Republican candidates, whereas

Figure 1. DMAs and the Preference Externality IV



Notes. Four states and five DMAs in the south west. Blue lines are DMA borders. Red-shade DMA primarily covers right-leaning population, whereas green-shaded DMA constitudes largely battle-ground states (New Mexico) in 2004.

New Mexico was considered to be a battleground state, a state in which the Republican and Democratic candidates both had a reasonable chance of winning the state. The political leanings of other counties in the same state's DMA are likely to affect a firm's advertising decision. For example, although a low-emissions automobile manufacturer may not find it profitable to advertise to right-leaning customers in general because of lower concerns for carbon, right-leaning customers in Bernalillo may still receive advertising if there are more left-leaning emission-conscious customers elsewhere in their DMA. This general intuition can be extended to all counties in a DMA, not just urban ones. Rural county advertising exposures can similarly be influenced by urban county characteristics such as political leanings.

To formalize this argument, consider the outcome equation

$$y_k = \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k, \tag{7}$$

with the instruments $Z_l = \{z_k, Z_{-k}\} \subset X_l$, including market-level variables that explain advertising and are excluded from the equation after conditioning on the local realization of those variables, that is, x_k. The identifying assumption is thus

$$E(\xi_k | x_k, \{z_k, Z_{-k}\}) = 0,$$

which holds naturally because the instruments contain only exogenous demand shifters in X_l. In other words, some observable determinant of preferences, Z_{-k}, can be treated as excluded and exogenous upon conditioning on that same variable for the region of analysis.

In practice, we include some aggregate function of the set of instrumental variables $h(Z_l)$, such as a variable's population-weighted average. An alternative is to include the entire vector of Z_l combined with statistical methods for dimension reduction (i.e., Gordon et al. 2023).

We generally expect local outcomes to be influenced by their own observable characteristics, z_k, in ways that other regions' values of those same characteristics, Z_{-k}, have no direct effect, so the above will constitute our primary specifications. That said, there may be cases where Z_{-k} might influence local outcomes through a route other than advertising, thus generating correlation with ξ_k. We will refer to these as potential spillover examples and analyze them with a set of instruments tailored to circumvent confounds from spillovers. For example, consider the low-emissions vehicle case again. The left- or right-leaning preferences in other counties (Z_{-k}) might also influence whether a state offers carpool stickers for hybrid vehicles and thus the demand shock in the focal county (ξ_k). Researchers have noticed that borders defined by media markets (DMA) are unique to the television advertising context, determined by other factors such as the ability of television signals to traverse geographical features. They commonly do not overlap with other jurisdictional borders, such as state borders. The difference between DMAs and other groupings of counties such as states helps separate advertising from other policies to use demand shifters in same-DMA but different-state counties, $Z_{l(k)\setminus s(k)}$, as instrumental variables, as long as $s(k) \neq l(k)$. In the outcome equation, we can add state fixed effects to capture the state-specific public policies. We illustrate the value of DMA's lacking perfect overlap with potentially confounding regions in Section 4.6.

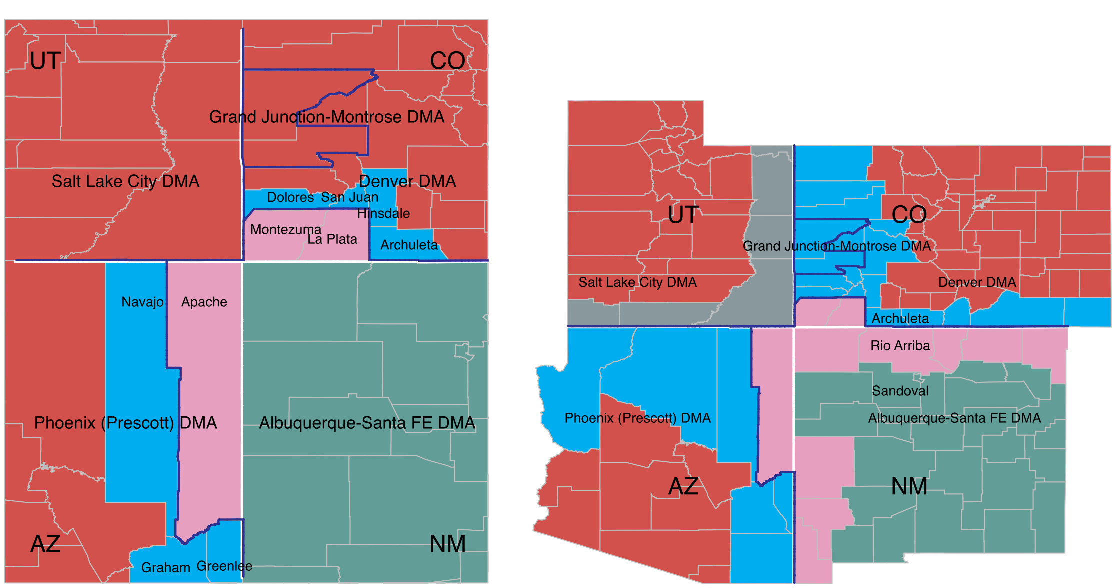
3.2. Border Approach

The border approach is a quasi-experimental method that compares two otherwise identical neighboring regions (indexed by k in our model) that incur different treatments because they reside in different treatment regions (indexed by l in our model). The treatment differences occur because the treatment regions differ and hence, impose different externalities.

For illustration, consider Apache County in Arizona (shaded in purple) on the left panel of Figure 2. It is contiguous with three other Arizona counties (Navajo, Graham, and Greenlee); however, it receives different advertising in the political advertising context because it resides in the Albuquerque-Santa Fe DMA primarily in New Mexico (battleground state shaded in green), whereas the other three counties reside in the Phoenix (Prescott) DMA in Arizona. As Huber and Arceneaux (2007) pointed out, counties in non-battleground states receive little or no advertising, yet a battleground state receives a lot of advertising. Consequently, Apache receives heavy advertising, whereas its similar neighboring Arizona counties receive no advertising. The essence of the border approach is to treat Apache and its neighboring region as a quasi-experimental "matched" pair, and many more such examples as other matched pairs, to form a quasi-experimental estimator of advertising effects. If we thought of advertising as a binary treatment, the purple-shaded areas on the left of Figure 2 would be treatment groups with advertising, and the blue-shaded neighboring areas would be control groups with no advertising.

The border approach was first introduced by Huber and Arceneaux (2007) in this presidential advertising context and later extended by Shapiro (2018) to nonpolitical contexts. The latter also expands the number of borders that can be analyzed, highlighted on the right panel of Figure 2, although a researcher may choose to drop counties when the DMA border coincides with a state border. Spenkuch and Toniatti (2018) applied border approach extensions similar to Shapiro (2018) to the political context. Applying our model to illustrate the econometrics of the border approach, the outcome

Figure 2. The Border Approach



Notes. Four states and five DMAs in the Southwest. Blue lines are DMA borders. Red-shaded states are right-leaning states, whereas the green-shaded state (New Mexico) was a battleground state in 2004. Areas shaded in pink, blue, and dark blue are focal areas in different implementations of the border approach.

equation for county k is

$$y_k = \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k \tag{8}$$
$$= \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \gamma_{b(k)} + e_k, \tag{9}$$

where $b(k)$ is the collection of counties lying on the same border. The unobserved demand shock ξ_k is decomposed into two components: one that is common within the border and captured by the fixed-effect $\gamma_{b(k)}$ and the other that is the residual local shocks e_k.

The statistical identifying assumption for a consistent estimate of α_1 is the following, which has commonly been referred to as the *unconfoundedness assumption*:

$$E(e_k | A_k, x_k, \gamma_{b(k)}) = 0. \tag{10}$$

This statistical assumption further implies two necessary assumptions detailed in the following subsections.

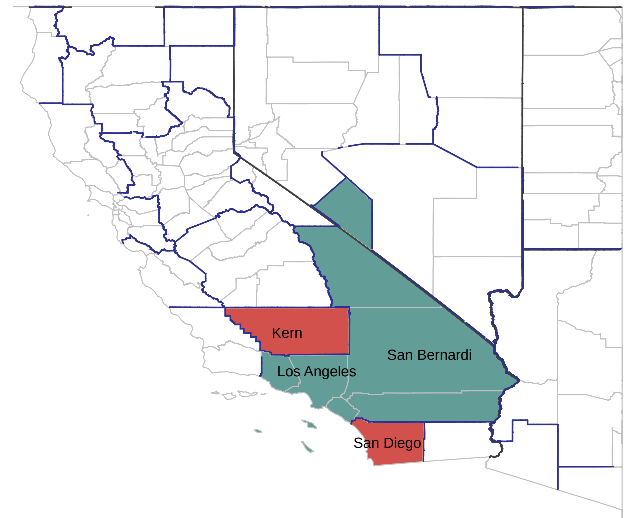
3.2.1. Ignorance Assumption. One assumption is behavioral in that that the advertiser ignores some local preferences e_k in the decision-making process to determine the ad level $d_{l(k)}$, which is a part of A_k. In our example above, the residual unobservable in Apache County is excluded from the ad decision in the Albuquerque DMA despite the model illustrating its relevance. We call this behavioral assumption the *ignorance assumption*.

Institutional details can increase the tenability of the ignorance assumption. The admissibility of the ignorance assumption in Huber and Arceneaux (2007), that the ad level d_l in the Albuquerque DMA ignores e_k in Apache, likely derives from the structure of the Electoral College. Any effect on voting behavior caused by advertising in Apache County has a minimal likelihood of affecting the electoral outcomes of the state of Arizona, a decidedly Republican-leaning state. That is to say, the influence that border regions such as Apache County have on the decision to advertise in the Albuquerque DMA is limited; that is, there may be bias, but it is likely small. However, concerns arise when we generalize the approach outside of political advertising where there is not a similar distortion mechanism to the Electoral College.

The ignorance assumption may also be justified by the border region representing a small share of the larger region at which treatment is defined. The challenge with the extended border approach is that many border regions are not small in practice (i.e., right panel of Figure 2). One obvious case is when the DMA and county are one in the same, as in Kern and San Diego, as shaded in red in Figure 3. Other border regions that challenge the ignorance assumption are border counties that represent a large share of their DMA, such as Los Angeles and San Bernardino (shaded in green).

More generally, we find that border and off-border counties are comparable in size, further questioning the validity of the border approach. In Table 2, we use

Figure 3. Southwest DMAs



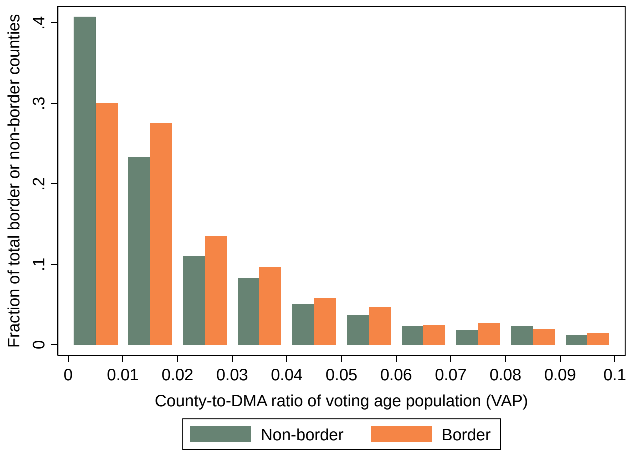
Note. Three color-shaded DMAs: Bakersfield (constitutes only one county: Kern), Los Angeles, and San Diego (constitutes only one county: San Diego).

Gordon and Hartmann (2013)'s data on voting age population (VAP) by county to show that border counties are more likely to represent a very large population share of their DMA than are non-border counties. In the data, two border counties represent more than 90% of their DMA, and two more between 70 and 80% of the DMA share. The majority of counties represent less than 10% of their DMA. Figure 4 illustrates that within these smaller counties, border and off-border counties are fairly evenly distributed, with off-border counties being more likely to be small, representing less than one percent of the DMA. We do not see a sharp difference between the two groups, suggesting that border counties are no smaller than off-border counties. In light of this concern, in our empirical application in Section 4, we apply the estimators using all border counties as well as only those that represent a small share of the DMA where the

Table 2. Border vs. Non-Border Counties' Percent of DMA Population: Political Ads

County VAP/DMA VAP	%Non-border counties	%Border counties
0–0.1	88.4	91.7
0.1–0.2	5.9	4.4
0.2–0.3	2.4	1.5
0.3–0.4	1.4	0.9
0.4–0.5	0.9	0.7
0.5–0.6	0.6	0.0
0.6–0.7	0.3	0.0
0.7–0.8	0.1	0.4
0.8–0.9	0.0	0.0
0.9–1	0.1	0.4
N	1,065	542

Figure 4. Distribution of Counties' Share of DMA Population by DMA Border or Not



Note. This figure plots the distribution of county-to-DMA ratio of voting age population (VAP) among all border counties and non-border counties whose county-to-DMA VAPs are below 0.1.

ignorance assumption is plausible. Notably, using a share of the DMA to justify the ignorance assumption could in fact justify using all small counties instead of just those on the border.

3.2.2. A Zero Spatial Correlation Assumption. Ignorance is only necessary but not sufficient for the unconfoundedness assumption in (10). Even if the decision maker is ignorant to the local shocks e_k in Apache County, if the shocks in Apache County are correlated with those in counties in the same DMA (e_{-k}), for example, San Juan in New Mexico, e_k affects d_l in an indirect way, and the unconfoundedness assumption would not hold. To guarantee the validity of the unconfoundedness assumption in (10), we therefore require an additional assumption of zero spatial correlation for unobservables between focal border counties and off-border counties within the same DMA.

Typically, it is difficult to empirically examine the validity of the zero spatial correlation assumption because we do not observe e_k. One suggestive exercise is to analyze observed factors to see whether they are correlated within DMAs. For example, unemployment and earnings data are highly correlated at 0.84 and 0.70, respectively, between focal counties lying on the border and the most populated county in DMAs, casting doubt on the zero spatial correlation assumption within DMAs.

3.2.3. Internal Consistency of the Border Approach. Although Shapiro (2018)'s border pair fixed effect likely absorbs potentially confounding unobservables (part of ξ_k that affects the ad decision), it highlights a new econometric tension. The motivation for the inclusion of border fixed effects is the existence of spatial correlation between counties across the same DMA border, captured by $\gamma_{b(k)}$, that can be differenced out. On the other hand, the zero spatial correlation assumption requires that residuals from the focal county do not affect the ad level through correlated demand shocks with off-border counties elsewhere in the DMA. The above two lines of arguments jointly suggest that the border approach relies on conditions difficult to justify; that is, spatial correlation is high or perfect across DMA borders but zero within DMA borders. Therefore, we suggest evaluating the plausibility by examining the relevant spatial correlations with observable data. In our data, spatial correlations exhibit the opposite pattern required to justify the border approach; they are higher within DMAs than between cross-border county pairs. Unemployment and income correlations are statistically lower for cross-border county pairs at 0.80 and 0.59, respectively, relative to the correlations reported above between a border county and its DMA's most populated county of 0.84 and 0.70 (p values for both differences are less than 0.001).[10]

To formalize this argument and illustrate the inconsistency in spatial correlation assumptions, consider two cross-border contiguous counties $b1$ and $b2$ lying in the same DMA border b (for example, Rio Arriba and Archuleta on the right panel of Figure 2). A simplified version of the conversion model is thus

$$y_{b1} = \alpha \cdot d_{b1} + \xi_{b1} = \alpha \cdot d_{b1} + \gamma_{b12} + e_{b1} \quad (11)$$

$$y_{b2} = \alpha \cdot d_{b2} + \xi_{b2} = \alpha \cdot d_{b2} + \gamma_{b12} + e_{b2}. \quad (12)$$

Correspondingly, their ad levels, without loss of generality, are determined as

$$d_{b1} = \rho_1 \gamma_{b12} + \rho_2 e_{b1} + \delta_b + \eta_{b1} \quad (13)$$

$$d_{b2} = \rho_1 \gamma_{b12} + \rho_2 e_{b2} + \delta_b + \eta_{b2}, \quad (14)$$

where they are endogenous to both common and idiosyncratic demand shocks ($\gamma_{b12}, e_{b1}, e_{b2}$) by different magnitudes (ρ_1, ρ_2). Other than the demand shocks in border counties, the ad level varies by exogenous shocks that are irrelevant to the conversion, decomposed as ($\delta_b, \eta_{b1}, \eta_{b2}$).

A third source of variation of the ad level is the demand shocks in other off-border counties within the same DMA, also captured in (η_{b1}, η_{b2}). For example, if $b1$ represents Rio Arriba on the right panel of Figure 2, its ad level d_{b1} is also affected by the demand shocks in the off-border county Sandoval ($b3$). Explicitly, if the conversion for Sandoval (b_3) is determined by

$$y_{b3} = \alpha \cdot d_{b1} + \xi_{b3},$$

then ξ_{b3} affects the ad level through η_{b1}, that is,

$$\mathrm{Cov}(\xi_{b3}, \eta_{b1}) \neq 0.$$

The unconfoundedness assumption in (10) requires ignorance and zero spatial correlation. Ignorance implies $\rho_2 = 0$. This suggests that the decision maker is not aware of shocks e that are specific to counties within the DMA but may be aware of shocks that are common across borders. Zero spatial correlation is also implied from (10) as $\text{Cov}(\eta_{b1}, e_{b1}) = 0$, and because ξ_{b3} affects ad level through η_{b1},

$$\text{Cov}(\xi_{b3}, e_{b1}) = 0. \tag{15}$$

Here is the econometric tension in this framework. The motivation for the cross-border comparison suggests $\text{Cov}(\xi_{b1}, \xi_{b2}) \neq 0$, with γ_{b12} capturing their common part. But the identification assumption in (15) suggests $\text{Cov}(\xi_{b3}, e_{b1}) = 0$, and thus $\text{Cov}(\xi_{b3}, \xi_{b1}) = 0$.[11] In other words, it requires that the cross-border county pairs ($b1$ and $b2$) are correlated, whereas the off-border county pairs ($b1$ and $b3$) are uncorrelated. See Appendix A for a more rigorous illustration of this tension.

Furthermore, to release the tension from above, we can either assume away the dependence of ad levels on unobserved demand shocks in off-border counties (i.e., $\text{Cov}(\xi_{b3}, \eta_{b1}) = 0$) or assume away the spatial correlation between contiguous borders (i.e., $\text{Cov}(\xi_{b1}, \xi_{b2}) = 0$). However, in both cases OLS is also consistent; the ad level is exogenous for off-border counties in the first case, whereas the cross-border comparison does not difference out any endogenous component in the second case.

3.2.4. When to Use the Border Approach and Potential Bias Reduction Versus OLS. Despite the discussion above, an empirical researcher may favor the border approach for its statistical power and empirical readiness relative to PEIVs or to reduce bias relative to OLS. The border approach leverages ad variation that comes from both the observed and unobserved parts of the preference determinants (X_{-k}, Ξ_{-k}). Therefore, it has more statistical power than PEIV estimators, which use variation only from X_{-k}. If instruments are not readily available or if the instruments do not deliver enough power, a researcher may elect to use the border approach while tolerating some statistical bias.

Yet given our discussion in the previous subsection, OLS may also be a viable estimation strategy, and it does not necessarily generate a more biased estimate than the border approach. In fact, the border approach may be more biased under certain conditions.

To see this, it is useful to ask, "What value can the cross-border observations provide?" The answer depends on the weight the unobservables shared across the border have in generating the endogeneity of the advertising. For example, if all endogenous variation is in common cross-border unobservables and all residual variation is exogenous, the border approach would eliminate all bias. But if all common cross-border unobservables were exogenous and all residual (county-specific) variation was endogenous, then the border fixed effect would leave us without any exogenous variation to identify the effect, whereas the endogenous local variation is unchecked in confounding the estimator.

To formalize and refine this intuition mathematically, we derive empirical conditions to compare the bias of OLS and the border approach based on the framework introduced from (11) to (14). We further assume that the error terms $(\gamma_{b12}, \delta_b, e_{bj}, \eta_{bj})$ are independent between each other, with variance $(\sigma_\gamma^2, \sigma_\delta^2, \sigma_e^2, \sigma_\eta^2)$.

The border approach estimator amounts to running a within-border-pair differenced regression between (11) and (12):

$$\Delta y_b = \alpha \Delta d_b + \Delta \xi_b,$$

where Δ represents the difference across borders such as $\Delta y_b = y_{b1} - y_{b2}$. We obtain expressions for the bias of estimators as

$$\text{bias}(\hat{\alpha}_{BA}) = \frac{\text{Cov}(\Delta d_b, \Delta \xi_b)}{\text{Var}(\Delta d_b)} = \frac{\rho_2 \sigma_e^2}{\rho_2^2 \sigma_e^2 + \sigma_\eta^2}.$$

On the other hand, the OLS estimator amounts to running OLS on Equation (11) on border counties. The bias term is

$$\text{bias}(\hat{\alpha}_{OLS}) = \frac{\text{Cov}(d_{b1}, \xi_{b1})}{\text{Var}(d_{b1})} = \frac{\rho_1 \sigma_\gamma^2 + \rho_2 \sigma_e^2}{\rho_1^2 \sigma_\gamma^2 + \rho_2^2 \sigma_e^2 + \sigma_\eta^2 + \sigma_\delta^2}.$$

From a direct comparison between the two bias terms, we have the following result.

Proposition 1. *The border approach is less biased than the OLS (i.e., bias($\hat{\alpha}_{BA}$) < bias($\hat{\alpha}_{OLS}$)) if any only if*

$$\rho_1 + \frac{\sigma_\delta^2}{\rho_1 \sigma_\gamma^2} < \rho_2 + \frac{\sigma_\eta^2}{\rho_2 \sigma_e^2}.$$

The condition proposed from above reveals situations where the border approach has a smaller bias than OLS. This occurs when the common component of exogenous ad variation (δ_b) is small and that of the endogenous demand shock (γ_{b12}) is large, both differenced out in the border approach. On the other hand, the implication from idiosyncratic shocks is opposite. The border approach has a smaller bias when the idiosyncratic part of exogenous ad variation (η_{bj}) is large and when that of the endogenous demand shock (e_{bj}) is small. In circumstances when these conditions hold, the border approach may reduce bias more than OLS and thus be preferred by researchers. Importantly, researchers should keep in mind that the border approach does not unambiguously reduce bias relative to OLS. The opposite patterns can lead to OLS being less biased. Therefore, comparisons of the estimators may be difficult to interpret.

4. Empirical Application

Having thus far discussed the identification conditions for the estimators, we now illustrate how they compare in an empirical application: political advertising. Although the purpose of the various identification strategies we have discussed is to resolve econometric endogeneity biases, we consider a case where there is no obvious unidirectional endogeneity bias (i.e., $\partial A / \partial \xi$ is not monotonic). This allows us to better compare the "local" weighting of advertising effects that can be produced by the border and IV approaches while also comparing their effects on statistical power.

4.1. Context and Data

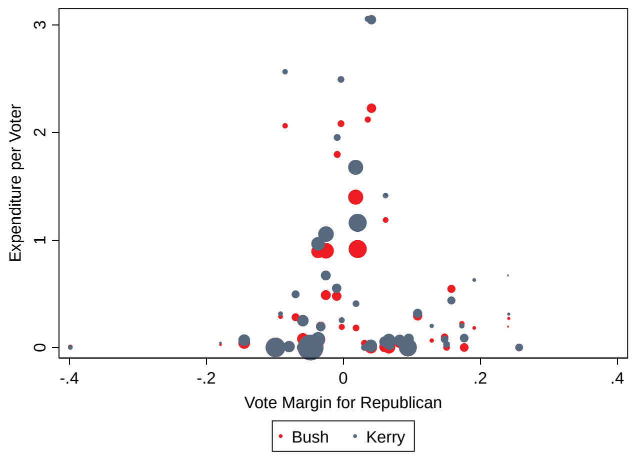
The outcomes of U.S. presidential elections have far-reaching implications. Policymakers, marketers, and political scientists have long studied whether the effects of presidential advertising can be large enough to affect electoral outcomes because they have implications for the debate on campaign funding and regulations on political advertising and communications. For our analysis, we use the same data set as Gordon and Hartmann (2013).

The main advertising variable is gross rating points (GRPs), a measure of the number of exposures per capita. Advertising variables are based on data from the Campaign Media Analysis Group (CMAG). Gordon and Hartmann (2013) pointed out and found that endogeneity biases are not a substantial concern in the case of political advertising. Because of the Electoral College system, a candidate's incentive to advertise diminishes if the unobserved (to the researcher) determinants of voters' preferences for the candidate are either too high or too low. In political language, candidates tend to advertise only in battlegrounds where voters like them neither too much nor too little. Gordon and Hartmann (2013) showed this pattern as evident in the distribution of advertising across the political leaning of advertising markets in Figure 5 and then documented that fixed effects estimates yield nearly identical estimates to those using supply-side instruments (ad prices from the year before the election). In the analysis below, we formally show a Hausman test that is consistent with the lack of endogeneity.

Data on (projected) advertising costs by market and time slot (day part) are obtained from SQAD, a market research firm. Gordon and Hartmann (2013) used lagged advertising costs as an instrument for advertising levels. Based on the observation that candidates can vary in the frequency at which they advertise during the day, Gordon and Hartmann (2013) used cost per mille (CPM), a per capita measure of advertising costs, for eight dayparts as their main IV.

As additional controls, Gordon and Hartmann (2013) used market-level party affiliation (a measure of local political preferences), the occurrence of a Senate election and local weather conditions (variables that affect voter turnout), population age, local unemployment, and

Figure 5. Cross-Market Variation in Advertising Expenditures: 2004 Presidential Election



Notes. GRPs by state-level voting margin in the 2004 election; horizontal axis is the state-level Republican vote share minus the Democratic vote share. Vertical axis is in hundreds of GRPs, such that one unit indicates one exposure per voter, on average. Bubbles are proportional to the state's voting-age population.

wages (demographics). The outcome variable, county-level vote share, is collected from www.polidata.org.

4.2. Gordon and Hartmann (2013) Fixed Effects and Supply-Side IV Specifications

We first replicate the results in Gordon and Hartmann (2013) to subsequently compare them to specifications that separately apply the border approach and preference externality-based IVs. The first two columns of Table 3 provide the estimates using the model and data from Gordon and Hartmann (2013). Column (1) is the fixed effect specification, which includes party-year and party-DMA fixed effects. It also includes a variety of demand-side observables measured at the county level (and more aggregate levels), which would be characterized as x_k in the model formulation.

We contrast this estimate with their supply-side IV specification in Column (2), which uses lagged ad prices as instruments. Consistent with the conjecture of no endogeneity biases in this context, the estimates are statistically indistinguishable, with the Hausman test having a p value of 0.89.[12] The partial R^2 in the IV specification is 0.34, suggesting that the ad prices do explain a reasonable amount of variation above and beyond the fixed effects and demand-side observables at the county level. However, the F statistic is quite low.[13] Therefore, the instruments are weak, explaining an increase in the standard errors reported, yet not so weak as to yield insignificant estimates.

4.3. Preference Externality Instrumental Variables

Political advertising highlights the unique and important value of the PEIV's inclusion of demand determinants of

Table 3. Comparison of Identification Strategies: Political Advertising

	Dependent variable is Ln(share) - Ln(share0)					
	All counties				Border counties	Small borders
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Ads)	0.053***	0.051**	0.060**	0.053***	0.041***	0.009
	(0.014)	(0.024)	(0.024)	(0.019)	(0.015)	(0.019)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Party-dma FE	Yes	Yes	Yes	Yes	Yes	Yes
Party-dma-border FE	No	No	No	No	Yes	Yes
Party-border-year FE	No	No	No	No	Yes	Yes
Lagged ads prices IVs	No	Yes	No	Yes	No	No
Preference Externality IVs	No	No	Yes	Yes	No	No
Observations	6,428	6,428	6,428	6,428	2,540	852
R^2	0.629	0.629	0.629	0.629	0.781	0.826
First-stage excluded F		2.601	48.705	56.885		
First-stage partial R^2		0.336	0.484	0.639		
Hausman test (p-val)		0.890	0.644	0.993		
Clustered SE	Party-DMA				Party-DMA, Party-Border	

Notes. Columns 1–4 use 1,607 counties for two parties in two years. Column 5 uses 635 border counties involving 545 unique counties in 93 DMA borders for two parties in two years, and column 6 uses only 213 border counties in 35 small borders, defined as borders with the population in the border-DMA smaller than 10% of the total population in that DMA on both sides in both years. Ads are measured as 1 + gross rating points (000). Controls include senate election, same incumbent, distance, population proportion in age buckets (25–44, 45–64, 65+), unemployment rate, average salary, rain(2000), rain(2004), snow(2000), and snow(2004). "Lagged ads price IVs" include lagged ad price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. "Preference externality IVs" include the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Clustered standard errors are in parentheses. The Hausman test compares column 1 to the IV estimates in columns 2–4. Because of the pairing of border counties that is necessary for the border approach, the data sets in columns 1–4 are not comparable to those in columns 5 and 6.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

advertising as sources of identification. The cross-market advertising variation in Figure 5 is extensive and systematic, but the ad prices that Gordon and Hartmann (2013) used as instruments, a supply-side determinant of advertising, likely explain a limited part of that variation. Most of the advertising incentives are likely driven by demand-side factors. To the extent that local unemployment, income, etc., or changes in these variables over time motivate voters to consider switching parties, and hence, advertisers to reach out to them, the PEIV approach relies on identifying variation that should drive much more of the observed distribution of advertising.

We consider the demand-side determinants of advertising that reflect preference externalities into the instruments. The PEIVs used for identification in this analysis are the DMA aggregations of the demand-side controls that Gordon and Hartmann (2013) included. Column (3) of Table 3 reports the results. The first stage partial R^2 for this strategy is higher at 0.48 relative to the 0.34 when using the lagged ad price instruments in column (2), suggesting that demand instruments do explain more variation in the advertising levels. The F statistic is much higher at 48.7, suggesting that the preference externalities are quite "strong" instruments for explaining the variation of ad exposure, mitigating concerns about "weak instruments." The point estimate with this strategy is about 20% higher than in the Gordon and Hartmann (2013) specifications, but the Hausman test does not reject the estimates to be different from the fixed effects estimates in column (1).

Column (4) uses both the preference externalities and lagged ad prices as IVs to include both demand- and supply-side determinants of advertising as instruments. In this case, the first-stage partial R^2 jumps to 0.64. The partial R^2 of column (4) is not much smaller than the sum of those from using the supply-side instruments (column (2)) and demand-side instruments (column (3)), suggesting that the two sets of instruments indeed capture meaningfully different sources of variation in advertising decisions in this empirical context. The F-statistic in column (4) reaches 56.9. These instruments are stronger, with standard errors falling midway between the previous IV specifications and the fixed effects estimates. Some precision is lost with these IVs, but not as much. Encouragingly, the point estimate is almost exactly the same as that found in the fixed effects and supply-side IV specifications that Gordon and Hartmann (2013) estimated. The p value for the Hausman test of the difference between these combined IVs and the fixed effects in column (1) is 0.99. Therefore, we conclude that neither weak instrument nor issues of representativeness of the estimates ("LATE" issue in the literature) arise in using preference externality IVs.

4.4. Border Approach in Political Advertising Data (All Border Counties)

We shift away from the IV approach that can be justified by the model of advertiser's behavior to the border approach based on Shapiro (2018) despite the econometric tension demonstrated in the last section. In column (5) of Table 3, we use data from 2,540 border-county-party-year observations, corresponding to 635 border counties (involving 545 out of 1,607 unique counties analyzed in Gordon and Hartmann 2013) for two parties across two election years. The advertising coefficient is nearly 30% lower compared with the fixed effect estimates, but based on a visual inspection of the standard errors, this is not a significant difference from the preceding estimates.[14]

Evaluating the potential loss in statistical power from dropping non-border observations, we find that the standard errors are only slightly larger than those from the fixed effects model, which includes 2.5 times as many observations. We believe this is because the primary identifying variation arises at the DMA level, and there are the same number of DMAs involved in the border approach. The border approach leverages all variation that remains after conditioning on FEs and may be less affected by issues regarding the lack of statistical power.

Yet, as we illustrate analytically in Section 3.2.4 in the context of bias, the border-pair fixed effects themselves may absorb useful variation. In Table 4, column (1), we report fixed effect estimates on the border data set. Compared with the border approach in column (5) of Table 3, there is a small decrease in the standard errors, suggesting the slightly more efficient nature of the OLS estimator in this context.

4.5. Border Approach in Political Advertising Data (Small Border Counties)

In light of the unconfoundedness assumption that underlies the border approach, another implementation of the border approach is to focus on small border counties. As described in the Section 3.2.1, some border counties represent a large share of the DMA such that the ignorance assumption, and hence, the unconfoundedness assumption, $E(e_k|d_{l(k)}, x_k, \gamma_{b(k)}) = 0$, is unlikely to hold. Therefore, we restrict our analysis to counties in small borders where this assumption is more plausible for borders that represent a small share of the DMA.

Although restricting to small borders may alleviate the identification challenge and make the ignorance assumption hold up empirically, it comes at a cost. A possible critique that the border approach faces is that the estimates may be "local" to only those counties in border regions and not representative of the effect size of the majority of counties in the country, as Shapiro (2018) noted. The focus on small border counties may raise particular concerns that the estimates may be "local." This can be difficult to assess in applications with endogeneity because the estimates may differ because of the locality or because the identification strategy is better (or worse) at resolving endogeneity problems. Because we believe we can rule out endogeneity biases broadly in this context, our analysis allows us to assess the potentially "local," spatially heterogeneous, and unrepresentative effects that might arise when using the more plausible small borders in border approach applications.

Table 4. Local Effects for Counties in Borders and Small Borders

	DV: Ln(share) − Ln(share0)			
	Borders	Small Borders		
	FE (1)	FE (2)	BA (3)	IV (4)
Ln(Ads)	0.056***	0.006	0.009	0.001
	(0.013)	(0.016)	(0.017)	(0.015)
Control	Yes	Yes	Yes	Yes
Party-DMA-border FE	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes
Party-border-year FE	No	No	Yes	No
PE IV	No	No	No	Yes
Observations	2,540	852	852	852
R^2	0.768	0.817	0.826	0.222
Hausman OLS (p value)			0.854	0.572
Hausman BA (p value)				0.207

Notes. Column 1 uses the identical sample of all border counties with column 5 in Table 3, and columns 2–4 use the identical sample of counties in small borders in column 6 of Table 3. Ads are measured as 1 + gross rating points (000). Controls include senate election, same incumbent, distance, population proportion in age buckets (25–44, 45–64, 65+), unemployment rate, average salary, rain(2000), rain(2004), snow(2000), and snow(2004). "PE IV" includes the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Standard errors in parentheses are clustered at the level of party-DMA and party-border. The Hausman OLS reports the Hausman test to compare columns 3 and 4 to the fixed-effect estimate in column 2. Hausman BA reports the Hausman test to compare column 4 with the border approach in column 3.

*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

We define small borders to be those with the population in the border-DMA smaller than 10% of the total population in that DMA on both sides in both years.[15] Among counties representing less than 10% of their DMAs' VAP, Figure 4 shows that the distribution of counties' share of the DMA is similar for border and non-border counties. We have identified 35 out of 93 borders to be small, which corresponds to 213 out of 635 border counties.

Table 3 column (6) reports the border approach on this subsample of small borders. We find that dropping additional counties does increase standard errors, but they are still no larger than any IV specification. In fact, the R^2 is larger than when analyzing all border counties. The notable difference is that the estimated ad coefficient drops to nearly zero. Once again, we are unaware of a test to formally compare this estimate to the fixed effects in (1) that is most efficient and consistent based on the above Hausman tests, but the loss of significance and nearly zero ad coefficient indicate a fundamentally different conclusion.

In order to further understand this pronounced drop in estimates, we discuss two sets of results. First, comparable near-zero ad effects are derived when using alternative specifications on small border regions (fixed-effects and PEIV in columns 1 and 4 of Table 4), so the small effect from the border approach (column 6 of Table 3, replicated in column 3 of Table 3) does not result from a better fit of the identification assumptions. It is driven by local effects; that is, advertising responsiveness in small border counties is different from other counties. In small border regions, the p value for testing the difference between fixed effects and border approach was 0.854. Similarly, that between fixed effects and PEIV was 0.572, and that between the border approach and PEIV was 0.207.

Another possible explanation of the difference in small-border-county estimates arises from potential measurement error, where only the DMA-level exposure is observed rather than the county-level exposure. We use simulations in Appendix C.1 to show that small counties are more subject to measurement error in ad exposure, which would bias the estimates downward, regardless of whether the county is at a DMA border or not. One counterargument to this is that the IV specification, which should remove measurement error, is also small. Separately, we have analyzed both fixed effects and IV specification on small counties that are not on borders, and the estimates are not dissimilar on analysis of all counties big and small. By elimination, this result suggests that local heterogeneous effects are the likely rationale for different effects within the small border regions.

Table 5. Controlling for Labor Market Spillovers

	DV: Ln(share) − Ln(share0)			
	(1)	(2)	(3)	(4)
Ln(Ads)	0.060**	0.055**	0.048*	0.046*
	(0.024)	(0.024)	(0.026)	(0.026)
Unemploy	−0.054***	−0.059***	−0.056***	−0.060***
	(0.011)	(0.010)	(0.011)	(0.010)
Income	0.008***	0.007***	0.008***	0.007***
	(0.002)	(0.001)	(0.002)	(0.001)
Unemp-contiguous		0.019		0.014
		(0.012)		(0.013)
Inc-contiguous		0.007		0.007
		(0.005)		(0.005)
Unemp-largest			0.035***	0.031**
			(0.011)	(0.013)
Inc-largest			−0.005	−0.005
			(0.005)	(0.005)
Other controls	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes
Party-dma FE	Yes	Yes	Yes	Yes
PE IV	Yes	Yes	Yes	Yes
Observations	6,428	6,428	6,428	6,428
R^2	0.629	0.631	0.630	0.632
First-stage excluded F	48.716	48.452	16.721	16.832
First-stage partial R^2	0.484	0.482	0.445	0.445
Hausman, p-val	0.644	0.763	0.945	0.987

Notes. All columns use 1,607 counties for two parties in two years. Ads are measured as 1 + gross rating points (000). Other controls include senate election, same incumbent, distance, population proportion in age buckets (25–44, 45–64, 65+), rain(2000), rain(2004), snow(2000), and snow(2004). "Lagged ads prices IV" include lagged ad price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. "Preference externality IV" includes the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Clustered standard errors are in parentheses. The Hausman test compares the IV estimates with corresponding FE estimates with the same set of control variables, respectively, in each columns.

$^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

4.6. Controlling for Spillovers in IV Implementations

The identifying assumption for using preference externalities as IVs is that, conditional on the demographics in the focal county, our identifying variation—aggregate level variation of demographics and other observables—is orthogonal to the local demand shocks. One primary source is the labor-market conditions that vary over time, that presidential candidates care about when campaigning. However, those labor market conditions, such as unemployment rate and wage level, have geographic spillovers and may affect voting preferences in nearby counties. For example, a negative shock in the labor market in nearby counties may affect the voting in a similar manner as a negative labor market shock in the focal county.

In order to deal with such concerns, we include the unemployment and income level in contiguous counties as additional controls. Table 5 reports the results. Column (1) is a replication of column (4) in Table 3, which explicitly reports the coefficient for the unemployment rate and annual income. We find that both labor market demographics affect voters' utility levels in a reasonable direction. The estimate for the ad effect is slightly decreased when we include the two labor market demographics in contiguous counties, which themselves are not significant (column 2). In column (3), we include another set of controls for geographic spillovers: the labor market condition in the largest county within the DMA and the estimate for the ad effect that is further driven down to 0.048. The coefficient for the unemployment rate in the largest county is, however, positive, which we suspect reflects the nonlinear dependence of the ad investment in our utility specification, because this variable affects the ad level greatly. In the last column, we include both sets of controls for spillovers. The estimate further drops to 0.046 but is still significant. There is a decrease in first-stage F-statistics for columns (3) and (4) because the inclusion of the labor market condition in the largest counties explains a great deal of ad variation at the DMA level. A Hausman test shows no evidence of a difference between the estimates in these tables and the fixed effects estimate.

It may be possible that there are other regions with confounds or spillovers, but as long as they do not perfectly overlap with the DMA, researchers can exclude them out when building the instruments, thus alleviating the bias concern. In our context, we can drop all but a selected set of strategic counties, for example, counties lying in the same DMA but different states, to form instruments as we suggest in Section 3.1. This strategy risks losing power because only a subset of DMAs will cross states, so we suggest it when approaches such as the above are unreasonable. In fact, we tried utilizing this narrow variation from DMAs crossing states in our application, but it reduces the statistical power and yields estimates at the margin of significance because we lose one-third of the sample, and the estimates reduce to the lower end of those reported in our primary specifications in Table 3. It may, however, be viable in other applications where estimates are larger or statistical power is not so vulnerable to this narrow variation.

4.7. Discussion

Overall, these analyses suggest robust findings across the fixed effects and instrumental variable specifications despite very different sources of identifying variation. Had an endogeneity bias actually existed in these data, we should at least expect the fixed effect estimate to be much different. An application of Shapiro (2018)'s border approach provides an effect that is smaller but not statistically different. However, this does not hold when borders that represent a nontrivial share of the DMA are excluded from the analysis. The former border approach inclusive of all borders would exhibit more endogeneity bias in most applications, unlike ours, where we rule out endogeneity bias based on advertising incentives and Hausman tests establishing the validity of fixed effects without IVs or the border approach. On the other hand, small borders yield more selectivity in included counties and less representative effects. The border approach thus may present a trade-off between representativeness bias when using small counties where the ignorance assumption is more plausible and endogeneity bias when using large counties where ignorance is implausible.

Instrumental variables are not immune from locality in estimates, but the robustness of the PEIV to the other two approaches based on fundamentally different sources of variation suggests limited concern in our application. We suspect that this is because the IVs are at least inclusive of all types of counties, even if some markets are weighted more heavily than others. On the other hand, we find evidence that the border approach recovers geographically local parameters when focusing on counties in which the unconfoundedness assumption is more likely to hold.

One may suggest that political advertising may be unique in that those living in small border localities are different and may not respond the same to advertising targeted toward the population centers of media markets and that the border approach could perform better in other applications. But it is also quite likely that these unique people may be influenced differently in other domains. Advertising for antidepressants may be focused on population centers of media markets (see, e.g., Shapiro 2018). Campaigns for CPG goods (Shapiro et al. 2021) may function differently in localities that are more likely to have Whole Foods, Trader Joe's, and other grocery outlets that do not carry or similarly emphasize traditional packaged goods. Moreover, the theoretical inconsistency with firms' optimizing behavior may also be more problematic when studying firms that are accustomed to making frequent advertising decisions.

We suspect that the lack of representativeness concerns for our IV estimators may be common to other advertising contexts. Noncompliance concerns are particularly strong in the case of the electoral college. In nonbattleground states where candidates lead ahead or lagged behind, neither party responds to exogenous shocks from the demand side, and both keep their advertising expenditures at a low level or zero. Other advertising applications do not have so many zero advertising observations to exhibit the pure noncompliance that can exacerbate LATE concerns for the IVs, so representativeness of the IVs elsewhere may be less problematic.

Finally, one can assess the likely plausibility of the ignorance and spatial correlation assumptions by evaluating observable variables that might exhibit similar concerns as unobservables. For example, Shapiro (2020) used a "balance test" to evaluate whether differences in observables at the border are predicted by advertising differences at the border. A lack of correlation suggests ignorance. We ran these tests for our context as well. Although the results suggest that ignorance at the border is possible, it can also suggest ignorance everywhere. This leaves no reason to favor the border approach and its representativeness concerns from so many dropped observations in place of OLS on the entire data set. As we show above, the spatial correlation assumption can also be assessed using observables, but the potential ignorance everywhere (conditional on fixed effects) would obviate the need for that as well. In this class of tests, it is important to keep in mind that observables are typically a small set of all potentially confounding variables. There are many other potential variables that may fail such tests, could we observe them. Because it only takes one such variable to bias the estimator, this also favors IVs where only one exogenous instrument for each endogenous variable is needed.

5. Conclusion

We provide a framework that unites the border approach and an IV approach, both of which rely on a "preference externality" for the identifying variation. The approaches differ in three dimensions: (i) the identifying assumptions, (ii) statistical power concerns, and (iii) the potential to produce unrepresentative effects. We have shown that the identifying assumptions in the border approach add a complicated and implausible structure of spatial correlation and assume the decision maker's ignorance of some local preferences. The IV approach is consistent with policymakers placing positive weight on the preferences of all groups of constituents or customers. Spatial correlations between IVs and unobservables could confound estimation, but conditioning on local realizations of the IVs should reduce that concern, and the estimator requires only the existence of one observable preference externality that lacks such spatial correlation, whereas the border approach requires all unobserved preference externalities to lack spatial correlation.

Neither estimator suffers from "power" concerns. Standard errors are only slightly affected in border applications because the focus on border counties does not reduce the number of observations of the more aggregate advertising decision. The addition of preference externalities into the set of potential IVs mitigates weak instrument concerns because it brings important demand-side variation into the first stage.

From a representativeness perspective, IV estimates could produce estimates that weight aggregate markets differently, but this does not appear to be a concern. The border approach can produce local effects based on the types of counties included. We find that ad effects are systematically different for all identification strategies when applied to the small border counties in which identifying assumptions in border approaches is most plausible.

Preference externality estimators rely on barriers to tailoring policies to unique (types of) individuals. Some of these barriers are being relaxed, especially in the context of advertising where microtargeting is increasingly feasible. Although this may shrink the aggregate level at which a decision is made, the effect on the viability of preference externality estimators is unclear. On the one hand, smaller decision-making regions actually increase the number of observations (clusters) at which decisions are made and can increase statistical power. On the other hand, decision makers such as advertisers do have incentives, if possible, to define clusters to be more homogeneous, which reduces the externality in explaining the advertising treatment. Nonetheless, privacy concerns are pushing against extreme microtargeting, especially in political contexts. Google states: "While we've never offered granular microtargeting of election ads, we believe there's more we can do to further promote increased visibility of election ads. That's why we're limiting election ads audience targeting to the following general categories: age, gender, and general location (postal code level)" (Spencer 2019). This limits, for example, targeting ads based on users browsing or search histories, which are a primary revealed preference metric about heterogeneity (Haggin 2019). As the marketing literature has documented (Rossi et al. 1996), such metrics reflect preference heterogeneity much more than demographics.

Outside of advertising, preference externalities are still commonly caused based on administrative regions. For example, COVID-related restrictions were set at the county level in 2020. One could posit that substantial heterogeneity among cities within a given county could lead to the kinds of preference externalities that we have explored in this paper. We leave it up to those with expertise in these areas to evaluate the applicability in health and other contexts.

Finally, the ignorance assumption defined herein likely applies more broadly. Ignorance assumptions may be necessary in other applications of unconfoundedness where the objective functions of those defining treatment are penalized by ignoring payoff-relevant information. In this paper, we used relative population size of the individuals whose preferences are ignored to consider a case where the assumption is more plausible. Population is a logical criteria in voting, but in other cases the logic for ignoring preferences may also relate to factors such as the margin attainable from advertising and converting a specific individual.

Acknowledgments

The authors thank Bradley T. Shapiro for sharing the data for the contiguous status for each county. The authors thank Latika Chaudhary, Yuxin Chen, Peter Rossi, and the seminar participants at Peking University, Rochester's Simon School of Business, RUC, Stanford GSB, University of Chile, and University of Houston for providing valuable feedback. All errors are our own. The authors are listed in reverse alphabetical order.

Appendix A. Econometric Tension of the Border Approach

To formalize the above argument, consider three contiguous counties in the right panel of Figure 2, Sandoval (k_1), Rio Arriba (k_2), and Archuleta (k_3), where k_1 is off the border, whereas k_2 and k_3 form the cross-border pair. More generally, the unobservable for our focal county, k_2, can be decomposed as

$$\xi_{k_2} = \gamma_{b23} + \gamma_{b12} + \epsilon_{k_2},$$

where γ_{b23} captures the correlation between ξ_{k2} and ξ_{k3}, and γ_{b12} captures the correlation between ξ_{k1} and ξ_{k2}. By comparing this assumption with the outcome equation as

$$\begin{aligned} y_k &= \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \xi_k \\ &= \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_k + \gamma_{b(k)} + e_k, \end{aligned} \tag{A.1}$$

$\gamma_{b(k_2)} = \gamma_{b23}$ and $e_{k_2} = \gamma_{b12} + \epsilon_{k_2}$. The unconfoundedness assumption is thus

$$\begin{aligned} E(\gamma_{b12} \mid d_{l(k_2)}, x_{k_2}, \gamma_{b23}) &= 0 \\ E(\epsilon_{k_2} \mid d_{l(k_2)}, x_{k_2}, \gamma_{b23}) &= 0. \end{aligned} \tag{A.2}$$

Applying the law of iterative expectation, we have

$$E(\epsilon_{k_2}|d_{l(k_2)}, x_{k_2}) = 0$$
$$E(\gamma_{b12}|d_{l(k_2)}, x_{k_2}) = 0.$$

However, we would expect γ_{b23} to satisfy similar statistical properties as γ_{b12}. By symmetry between γ_{b12} and γ_{b23}, we would expect the same property for γ_{b23},

$$E(\gamma_{b23}|d_{l(k_2)}, x_{k_2}) = 0,$$

which makes $\xi_{k2} = \gamma_{b12} + \gamma_{b23} + \epsilon_{k2}$ exogenous to the ad decision (conditional only on the observed demand determinants x_{k_2}). This means that there is no need to include border fixed effects in (A.1). The identification argument is degenerated to what is used in Huber and Arceneaux (2007), or even stronger,

$$E(\xi_k|d_{l(k)}, x_k) = 0,$$

that the local demand shocks are orthogonal to the ad decision.

Therefore, justifying the border pair fixed effects for the purpose of identification requires a qualification to the zero spatial correlation assumption introduced in Section 3.2.2. Specifically, the assumption is zero spatial correlation within DMAs but perfect spatial correlation across DMA borders. This modified assumption may be difficult to motivate empirically.

Appendix B. Hausman Test for Specifications

The Hausman test is widely used to evaluate two estimators (Hausman 1978). To perform the test, the null hypothesis is set so that one estimator is more efficient than the other under the null but inconsistent under the alternative, whereas the other is always consistent. We first show classical examples where the Hausman test can be used to evaluate the inclusion of additional control variables, as well as the necessity of the use of IV. We then demonstrate the regression-based execution of the Hausman test that can deal with nonclassical standard errors. We finalize this section by the pairwise test among the three estimators in our empirical contexts: the fixed-effect estimator, the border approach, and the PEIV.

B.1 Two Classical Examples of the Hausman Test

Consider first the usage of instrumental variable. Suppose we have the model

$$y_i = \alpha_0 + \alpha \cdot x_i + \varepsilon_i, \tag{B.1}$$

with a set of valid instrumental variables z_i. The job is to decide whether to use 2SLS or OLS to estimate the original equation. The null hypothesis is

$$\mathrm{Cov}(\varepsilon_i, x_i) = 0. \tag{B.2}$$

Similarly, consider two alternative estimators. One is the OLS estimator $\hat{\alpha}_1$, and the second is the 2SLS estimator $\hat{\alpha}_2$. Under the null, both estimators are consistent, whereas $\hat{\alpha}_1$ is more efficient, because $\hat{\alpha}_2$ uses only part of the variation of x_i that can be explained by z_i. Under the alternative, however, only 2SLS estimator $\hat{\alpha}_2$ is consistent. The Hausman test can be used to evaluate whether to use IV for model specification.

Consider second the inclusion of additional controls. Think about the model

$$y_i = \alpha_0 + \alpha \cdot x_i + \varepsilon_i, \tag{B.3}$$

and the job is to decide whether we should include additional control variables z_i such that

$$y_i = \alpha_0 + \alpha \cdot x_i + z_i'\delta + e_i.$$

The null hypothesis is

$$\delta = 0.$$

Now consider two alternative estimators. One is the OLS estimator $\hat{\alpha}_1$ without including z_i, and the other is the OLS estimator $(\hat{\alpha}_2, \hat{\delta})$ after including z_i as the control. Under the null hypothesis, both $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are consistent, but $\hat{\alpha}_1$ is more efficient because the inclusion of z_i may absorb some variation of x_i. Under the alternative, $\hat{\alpha}_1$ is inconsistent because of omitted variable bias, whereas $\hat{\alpha}_2$ is consistent.

It is worth noting that the second case of including additional control can be formulated in way that is identical to the first case, with the null hypothesis as $\mathrm{Cov}(\varepsilon_i, x_i) = 0$ in (B.3).

B.2. Regression-Based Implementation for Nonclassical Standard Errors

Classical implementation of the Hausman test calculates the difference between two estimators $(\hat{\alpha}_1 - \hat{\alpha}_2)$ and its variance to construct test statistics. However, this is applicable only in classical case, where the error terms are i.i.d distributed. When the standard errors are nonclassical, such as heteroskedastic robust or clustered, the computation is complicated. Fortunately, Davidson and MacKinnon (1990) and Adkins et al. (2012) introduced a regression-based implementation that can cope with the nonclassical standard errors.

Consider the test of IV and OLS described in (B.1) and (B.2). The Hausman test can be performed in the following two step procedure.

1. Regress x_i on z_i as

$$x_i = \gamma_0 + z_i'\gamma + \eta_i,$$

and get the predicted value $\hat{x}_i$.

2. Regress y_i on $(x_i, \hat{x}_i)$ as

$$y_i = \alpha_0 + \alpha \cdot x_i + \delta \cdot \hat{x}_i + e_i,$$

and report the test statistics for $\delta = 0$.

As described from above, the test for the inclusion of additional control can be formulated in the same manner, thus also implementable following the procedure. Any specifications of nonclassical standard errors can be imposed easily in the second step.

B.3. Testing Between Specifications in our Empirical Contexts

We employ three different specifications to test the ad effects: the fixed-effect estimators, the border approach, and the PEIV. Intuitively, their efficiency levels are in a decreasing order. The fixed-effect estimator uses all ad variations, the border approach uses all ad variations from the off-border parts, and the PEIV uses only part of *explained* ad variations from the off-border parts. On the other hand, the unconfoundedness assumptions for their consistency also decrease in restrictiveness. The fixed-effect estimator requires the decision maker to ignore all unobserved demand shocks, and the border approach only requires the decision maker to ignore idiosyncratic demand shifters in the border, where the PEIV does not impose any restrictions. The above rankings in efficiency and consistency requirements naturally motivate the usage of the Hausman test.

Implementation-wise, the test between fixed effects and the border approach is to test the inclusion of additional border-by-time fixed effects, thus implementable using the above two-step procedure. The test between fixed effects and the PEIV is to test the usage of instrumental variables, also implementable. To test between the border approach and the PEIV, we use the Hausman test for the usage of instrumental variables in the border approach specification.

Appendix C. Measurement Error

In most applications of preference externality estimators, local ad exposure is measured at an aggregate level, and local variation is assumed away. However, if this assumption is not true, that is, there exists local variation of viewership but researchers can observe only an aggregate measure, this will lead to a measurement error problem. In this section, we first demonstrate the existence of such a problem and then the relevance of it to the identification strategies discussed in this paper. Finally, we discuss potential candidate solutions.

C.1. Measurement Error

In most applications, we have outcome variable y_k (this can be either $\log q_k$ or $\log q_k - \log(1-q_k)$), depending on underlying demand models), expressed as

$$y_k = \alpha_0 + \alpha_1 A_k + \alpha_2 x_k + \xi_k \tag{C.1}$$

with local ads exposure A_k determined by

$$A_k(d_l, x_k, \tilde{v}_k)$$

In practice, we do not observe finer-level ad exposure A_k. Instead, we observe only the aggregate level A_l, which mechanically equals

$$\begin{aligned} A_l &= \frac{1}{n_l}\sum_{k\subset l} n_k A_k \\ &= \frac{1}{n_l}(n_k A_k + n_{-k}A_{-k}) \end{aligned}$$

where $n_{-k} = n_l - n_k$, and

$$A_{-k} = \frac{1}{n_{-k}}\sum_{j\neq k} n_j A_j = A_{-k}(d_l, x_{-k}, \tilde{v}_{-k}).$$

By simple algebra, we have

$$A_k = \frac{n_l}{n_k}A_l - \frac{n_{-k}}{n_k}A_{-k}.$$

Plugging this into (C.1), we have

$$\begin{aligned} y_k &= \alpha_0 + \alpha_1\left(\frac{n_l}{n_k}A_l\right) + \alpha_2 x_k - \alpha_1\frac{n_{-k}}{n_k}A_{-k} + \xi_k \\ &= \alpha_0 + \left(\alpha_1\frac{n_l}{n_k}\right)A_l + \alpha_2 x_k + (u_k + \xi_k) \end{aligned} \tag{C.2}$$

Other than the commonly discussed endogenous ad decisions that may bias the estimates, that is, $E(\xi_k A_{l(k)}) \neq 0$, there are two other sources of bias coming from the measurement issue in (C.2):

1. $\text{Cov}(A_l, u_k) \neq 0$, simply because u_k is a part of A_l.
2. $\text{Cov}(x_k, u_k) \neq 0$, because x_k is also considered in the ad decision, that is, $\text{Cov}(x_k, d_l) \neq 0$.

C.2. Two Identification Approaches Revisited

With the measurement error in mind, we examine the validity of the two identification approaches discussed in our study: the border approach and the preference externalities IV approach.

Both approaches suffer in the presence of measurement error. For the border approach, even if the ad decision d_l is unconfounded with respect to local demand shocks ξ_k in (C.2), we still have the measurement error problem. The estimated coefficient may even become negative as $\text{Cov}(A_l, u_k) < 0$. For the IV approach, $h(X_{1l})$ is not a valid instrument anymore, because X_{1l} affects the ad decision d_l, which is further correlated with measurement error u_k.

In general, it is quite difficult to find any valid instrument from observables $\{x_k\}_{k\subset l}$. These observables are observed by advertisers when making ad decisions, thus entering d_l. They are naturally correlated with measurement error u_k. In our framework, the only candidate instrument is some observed factors in $\tilde{v}_k$, which shifts the ad exposure in an exogenous manner (say, weather). It is unknown to the advertiser when making ad decisions and therefore orthogonal to u_k. Yet it is correlated with coarser-level ad exposure A_l through affecting the A_k. We talk about the validity of such an instrumental variable solution in the next subsection.

C.3. Quantifying the Bias

In this subsection, we try to quantify the bias from a naive OLS estimate of (C.1) by directly inputting the aggregate-level ad exposure A_l to proxy for local-level ad exposure A_k, which gives an OLS estimate of $\hat{\alpha}_{1,OLS}$. For simplicity, we ignore the term of $\alpha_2 x_k$ and thus the second source of endogeneity in Appendix C.1. By plugging into (C.2), we can characterize the bias in the estimates as

$$\hat{\alpha}_{1,OLS} - \alpha_1 \cdot \frac{n_l}{n_k} = \frac{\text{Cov}(A_l, u_k + \xi_k)}{\text{Var}(A_l)}.$$

After some calculation (detailed proof in Appendix C.4), we have

$$\begin{aligned} \hat{\alpha}_{1,OLS} &= \alpha_1\frac{\text{Cov}(A_l, A_k)}{\text{Var}(A_l)} + \frac{\text{Cov}(A_l, \xi_k)}{\text{Var}(A_l)} && \text{(C.3)} \\ &= \alpha_1 B_{1k} + B_{2k}, && \text{(C.4)} \end{aligned}$$

where the first term B_{1k} relates to the measurement error, and the second term B_{2k} relates to the endogenous ad decisions.

Classical measurement errors usually lead to attenuation bias. Although we do not directly observe A_k, A_l is one proxy (say, $A_l = A_k + e$). Therefore, $\text{Cov}(A_l, A_k) < \text{Var}(A_l)$ and $B_{1k} < 1$, which implies downward bias. However, in the current context of counties within the same DMA,

$$A_l = \frac{n_k}{n_l}A_k + \frac{n_{-k}}{n_l}A_{-k},$$

the first term implies that DMA-level ad exposure A_l is a weighted average of county-level A_k, thus with smaller variance, which will bias the estimates upward; the second term may drive the estimates downward, but it is not pure noise, which is different from the classical measurement error.

To further investigate the direction of the bias, we assume symmetry among different counties k other than the size, that is, $\mathrm{Var}(A_k) = v$ and $\mathrm{Corr}(A_j, A_k) = \rho$, and B_{1k} can be simplified as

$$B_{1k} = \frac{\rho + (1-\rho)\frac{n_k}{n_l}}{\sum_{j\subset l}\frac{n_j}{n_l}\left(\rho + (1-\rho)\frac{n_j}{n_l}\right)} \tag{C.5}$$

Obviously, $B_{1k} = 1$ when $\rho = 1$ or $n_k = n_j$ for all $j, k \subset l$. In other words, there is no measurement error problem if (1) there is no local variation of viewership, and A_l is the exact measure of A_k, or (2) there is no asymmetry across different counties. In this case, the two directions cancel out, which leads to unbiased estimates.

To further understand the direction of bias under asymmetry, we plot one numerical example in Figure C.1 for eight counties in one DMA. The size of the counties, n_k, is picked randomly from 3 to 9, and we plot the value of B_{1k} under different values of ρ. First, in contrast with the standard attenuation bias that always biases downward the estimates, in our setting, the estimates are biased upward for large counties. Second, the magnitude of bias is decreasing in ρ. In the extreme case with $\rho = 1$, there is no measurement error and $A_l = A_k$, $B_{1k} = 1$.

Although not our main focus in the current study, we also derive some analysis on the bias from an endogenous ad decision, which is captured by the second term of (C.3). To simplify B_{2k}, we write out the first order condition of (3) as

$$\sum_{k\subset l} n_k\left((p_k - c)\frac{\partial q_k}{\partial A_k} - w_l\right)\frac{\partial A_k}{\partial d_l} = 0$$

and let

$$r_k = (p_k - c)\cdot\frac{\partial q_k}{\partial A_k} - w_l,$$

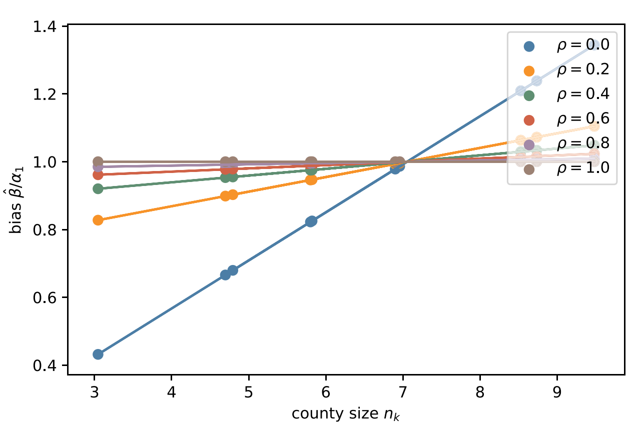
so we have

$$\sum_{k\subset l} n_k r_k \frac{\partial A_k}{\partial d_l} = 0, \tag{C.6}$$

And with some simplification shown in Appendix C.4, we have

$$B_{2k} = B_2\mathrm{Cov}(d_l, \xi_k), \tag{C.7}$$

Figure C.1. Bias from Measurement Error



Notes. This figure plots the ratio of OLS estimates and the true parameters, B_{1k}, for eight counties of different sizes within the same DMA, under different values of ρ. The county sizes n_k are picked randomly between 3 and 9.

where

$$B_2 = \left(\sum_{j\subset l}\frac{n_j}{n_l}\frac{\partial A_j}{\partial d_l}\right)\cdot\frac{1}{\mathrm{Var}(A_l)},$$

and if r_k is constant in (C.6), $B_2 = 0$.

C.4. Proof

The bias term is

$$\hat\alpha_1 - \alpha_1\frac{n_l}{n_k} = -\alpha_1\frac{n_{-k}}{n_k}\cdot\frac{\mathrm{Cov}(A_l, A_{-k})}{\mathrm{Var}(A_l)} + \frac{\mathrm{Cov}(A_l, \xi_l)}{\mathrm{Var}(A_l)},$$

and with

$$A_{-k} = \frac{n_l}{n_{-k}}A_l - \frac{n_k}{n_{-k}}A_k,$$

we have

$$\mathrm{Cov}(A_l, A_{-k}) = \frac{n_l}{n_{-k}}\mathrm{Var}(A_l) - \frac{n_k}{n_{-k}}\mathrm{Cov}(A_l, A_k)$$

plug in to get (C.3).

To derive (C.5), with our assumption, $\mathrm{Cov}(A_j, A_k) = \rho v$ for $j \neq k$, we have

$$\begin{aligned}\mathrm{Cov}(A_l, A_k) &= \sum_j\frac{n_j}{n_l}\mathrm{Cov}(A_j, A_k) = \frac{1}{n_l}\left(n_k v + \rho\sum_{j\neq k}n_j v\right)\\ &= \left(\rho + (1-\rho)\frac{n_k}{n_l}\right)v\end{aligned}$$

and

$$\mathrm{Var}(A_l) = \sum_k\frac{n_k}{n_l}\mathrm{Cov}(A_k, A_l) = \sum_k\frac{n_k}{n_l}\left(\rho + (1-\rho)\frac{n_k}{n_l}\right)v$$

plug in to get (C.5). To derive (C.7), observe that the numerator for B_{2k} is

$$\begin{aligned}\mathrm{Cov}(A_l, \xi_k) &= \sum_{j\subset l}\frac{n_j}{n_l}\mathrm{Cov}(A_j, \xi_k)\\ &= \frac{1}{n_l}\sum_{j\subset l}n_j\frac{\partial A_j}{\partial d_l}\mathrm{Cov}(d_l, \xi_k)\\ &= \frac{\mathrm{Cov}(d_l, \xi_k)}{n_l}\left(\sum_{j\subset l}n_j\frac{\partial A_j}{\partial d_l}\right),\end{aligned}$$

where the second line is derived using the delta method (first-order expansion of $A_j = E(A_j) + \frac{\partial A_j}{\partial d_l}(d_l - E(d_l))$).

Endnotes

[1] Please refer to Spenkuch and Toniatti (2018) and Dieterle et al. (2020) for more discussion.

[2] Whereas competition prevents a discontinuity analysis, the unconfoundedness assumption is broadly challenging when analyzing profit-maximizing firms. It assumes out the existence of any outcome determinants that are unobservable to econometricians but likely to be observable to the firm and therefore should be included when determining actions.

[3] We thank Brad Shapiro for suggesting this exercise of focusing on small borders.

[4] d_l can either be a dummy variable that equals to one if the advertiser is doing ads in group l, or it can also be an ad intensity.

[5] There is slight simplification of notation in the following equation. By $k \subset l$, we mean $k \subset l$ and $k \in K$.

[6] Externalities could also arise from preference parameters that may be captured by local random coefficients, but the current parsimonious specification with X and ξ suffices to characterize the border and IV implementations of preference externality estimators.

[7] Although the utility specification in (5) is simple, where the only heterogeneity comes from the idiosyncratic error term ϵ_{ik}, our framework can be easily modified to allow for a richer heterogeneity. For example, suppose we allow random coefficients as

$$u_{ik} = \alpha_0 + \alpha_{1i} g(A_k) + \alpha_2 x_k + \xi_k + \epsilon_{ik},$$

we can apply standard contraction mapping from Berry et al. (1995) and convert the conversion equation to

$$y_k = s(\{q_j\};\sigma) = \alpha_0 + \alpha_{1i} g(A_k) + \alpha_2 x_k + \xi_k,$$

where the function $s(\cdot;\sigma)$ is derived from contraction mapping, with an additional variance parameter σ for α_{1i}.

[8] A class of estimators leverage the political cycle (Sinkinson and Starc 2019, Moshary et al. 2021). These estimators leverage exogenous crowding out of commercial advertising because of the large influx of political advertisements leading up to elections, a form of externality caused by the capacity constraint of advertising. They vary from preference externality estimators in that it is not the focal advertiser's decision and layering structure of demand that cause the externality.

[9] By focusing on the national level, Thomas (2020)'s estimator becomes a time-series estimator with likely fewer advertising decisions to gain statistical power (i.e., standard errors would likely need to be clustered at the national level, of which there is one, or perhaps at the time level). We generalize this approach.

[10] Specifically, for each county lying on the DMA border i, we map with the most populated county in its DMA, i_1, and its cross-border contiguous counties, i_2. We collect the annual demographic data for five years between 2000 and 2004 and calculate two correlation coefficients, $\rho_{i_1} = \mathrm{corr}(y_{it}, y_{i_1 t})$ and $\rho_{i_2} = \mathrm{corr}(y_{it}, y_{i_2 t})$. We report their average $\rho_1 = \mathrm{Avg}(\rho_{i_1})$ and $\rho_2 = \mathrm{Avg}(\rho_{i_2})$ for unemployment $(\rho_1, \rho_2) = (0.80, 0.84)$ and for household income $(\rho_1, \rho_2) = (0.59, 0.70)$. We appreciate the review team for suggesting this exercise.

[11] We implicitly assume that $\mathrm{Cov}(\xi_{b3}, \gamma_{b12}) = 0$ and that the common component of the cross-border counties are distant away (than the idiosyncratic element e_{b1}) and thus uncorrelated with off-border county ξ_{b3}. In other words, if ξ_{b3} and ξ_{b1} are correlated, they are more likely to be correlated through e_{b1}.

[12] See Appendix B for more details.

[13] The original paper reported an F of 88.2, but it decreases to 2.6 after Kleibergen-Paap correction for cluster and robust standard errors.

[14] Unlike the IV specifications, where a Hausman test can statistically evaluate the difference in the estimates from the fixed effect specification, we are unaware of any comparable test for the border approach. The primary challenge is that the border approach is implemented in a different data set from the fixed effect or IV specification. The former is not even a subset of the latter because the same county may appear in multiple border pairs.

[15] We thank Brad Shapiro for suggesting this exercise.

References

Adkins LC, Campbell RC, Chmelarova V, Hill RC (2012) The hausman test, and some alternatives, with heteroskedastic data. Baltagi BH, Carter Hill R, Newey WK, White HL, eds. *Essays in Honor of Jerry Hausman* (Emerald Group Publishing Limited, Leeds, UK), 515–546.

Berry ST, Haile PA (2022) Nonparametric identification of differentiated products demand using micro data. Preprint, submitted April 13, https://arxiv.org/abs/2204.06637.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica.* 63(4):841–890.

Bound J, Jaeger D, Baker RM (1995) Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak. *J. Amer. Statist. Assoc.* 90(430):443–450.

Card D, Krueger A (1994) Minimum wages and employment: A case study of the New Jersey and Pennsylvania fast food industries. *Amer. Econom. Rev.* 84(4):772–793.

Davidson R, MacKinnon JG (1990) Specification tests based on artificial regressions. *J. Amer. Statist. Assoc.* 85(409):220–227.

Dieterle S, Bartalotti O, Brummet Q (2020) Revisiting the effects of unemployment insurance extensions on unemployment: A measurement-error-corrected regression discontinuity approach. *Amer. Econom. J. Econom. Policy* 12(2):84–114.

Dube A, Lester TW, Reich M (2010) Minimum wage effects across state borders: Estimates using contiguous counties. *Rev. Econom. Statist.* 92(4):945–964.

Fan Y (2013) Ownership consolidation and product characteristics: A study of the US daily newspaper market. *Amer. Econom. Rev.* 103(5):1598–1628.

Gentzkow M, Shapiro JM (2010) What drives media slant? Evidence from US daily newspapers. *Econometrica* 78(1):35–71.

Gordon BR, Hartmann WR (2013) Advertising effects in presidential elections. *Marketing Sci.* 32(1):19–35.

Gordon BR, Hartmann WR (2016) Advertising competition in presidential elections. *Quant. Marketing Econom.* 14(1):1–40.

Gordon BR, Lovett MJ, Luo B, Reeder JC III (2023) Disentangling the effects of ad tone on voter turnout and candidate choice in presidential elections. *Management Sci.* 69(1):220–243.

Haggin P (2019) Google to restrict political ad targeting on its platforms. *Wall Street Journal* (November 20), https://www.wsj.com/articles/google-to-restrict-political-ad-targeting-on-its-platforms-11574293253.

Hausman JA (1978) Specification tests in econometrics. *Econometrica* 46(6):1251–1271.

Huber GA, Arceneaux K (2007) Identifying the persuasive effects of presidential advertising. *Amer. J. Political Sci.* 51(4):957–977.

Imbens GW, Angrist JD (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2):467–475.

Masten MA, Torgovitsky A (2016) Identification of instrumental variable correlated random coefficients models. *Rev. Econom. Stat.* 98(5):1001–1005.

Moshary S, Shapiro BT, Song J (2021) How and when to use the political cycle to identify advertising effects. *Marketing Sci.* 40(2):283–304.

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

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Marketing Sci.* 15(4):321–340.

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

Shapiro BT (2020) Advertising in health insurance markets. *Marketing Sci.* 39(3):587–611.

Shapiro BT, Hitsch GJ, Tuchman AE (2021) Tv advertising effectiveness and profitability: Generalizable results from 288 brands. *Econometrica* 89(4):1855–1879.

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

Spencer S (2019) An update on our political ads policy.

Spenkuch JL, Toniatti D (2018) Political advertising and election results. *Quart. J. Econom.* 133(4):1981–2036.

Thomas M (2020) Spillovers from mass advertising: An identification strategy. *Marketing Sci.* 39(4):807–826.

Waldfogel J (2003) Preference externalities: An empirical study of who benefits whom in differentiated-product markets. *RAND J. Econom.* 34(3):557–569.