Preference Externality Estimators: A Comparison of Border Approaches and IVs

Xing Li, Wesley R. Hartmann, Tomomichi Amano*

January 17, 2020

Abstract

We document that identification strategies exploiting cross-border differences in treatment are a variant of the preference externality estimator more recently developed in the industrial organization literature. Waldfogel (1999) coined the term preference externality to describe how the aggregate tastes of heterogenous consumers can influence the products made available to one another within a common market. The externality forms the basis for an instrumental variable estimator where, after conditioning on observed preference determinants for a focal consumer type, the aggregate observables within the market, which vary by the preferences of other types, influence the focal type's "treatment" but are excluded from the focal type's outcome equation. Variation in treatment across geographic borders similarly arises from an externality where otherwise comparable individuals near a border face different policies because of different externalities from their respective aggregate regions. We use an advertising application to compare the border and IV implementations of preference externality estimators across three dimensions: i) identifying assumptions, ii) sacrifices in statistical power, and iii) local estimates of heterogenous effects.

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

1 Introduction

Two otherwise comparable individuals can face different policies, product offerings, prices or other treatments when the states, markets or other groups they belong to differ. When governments and firms incorporate group composition in their decision-making, the treatment any one individual experiences is affected by the rest of the group. This "preference externality", as termed by Waldfogel (1999), forms the basis of two causal inference approaches each predominant within different literatures. Instrumental variable implementations following George and Waldfogel (2003) have been applied to media slant (Gentzkow and Shapiro, 2010), mergers (Fan, 2013) and differentiated products more broadly (Berry and Haile, 2010) where the policy-makers are firms. Border implementations have studied minimum wage effects (Card and Krueger, 1994; Dube et al., 2010) and school choice (Black, 1999) where the policy-maker is a government entity such as a state or public school district. These border approaches have migrated into advertising with Huber and Arceneaux (2007)'s analysis of presidential elections and more recently to firm advertising by Shapiro (2017), Shapiro (2018),Tuchman (2017) and Tuchman et al. (2018).

The analysis of firms as policy-makers brings to light a sharp contrast between the plausibility of the identifying assumptions of these two preference externality estimators. Instrumental variable estimators are derived from a model of the policy-maker's decision process which, in the case of firms, involves profit maximization integrating over the entire distribution of preferences within a group. Border approaches however rely on an unconfoundedness assumption that policy-makers ignore preferences that are unique to each local region of analysis. Researchers increase the plausibility of this unconfoundedness assumption by using fixed effects and sometimes dropping observations where a local region makes up a large share of the group at which policy is set. Nevertheless, unconfoundedness assumptions are inconsistent with the optimizing behavior the policy-maker should engage in if placing positive weight on preferences for all constituents or customers.

Such estimators also suffer from statistical problems, namely statistical power and representativeness. In the case of border estimators, focusing on a subset of markets located on borders reduces the number of observations and hence statistical power. Moreover, borders may be unique thereby raising concerns about the representativeness of estimated effects. In the case of IV estimators, they also lack 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 as well. A question therefore arises as to whether border implementations' power and representativeness concerns are sufficiently small (relative to IVs) to justify their less appealing identifying assumptions. This is inherently a question unique to each empirical application, but representativeness concerns cannot be assessed and compared in most cases.

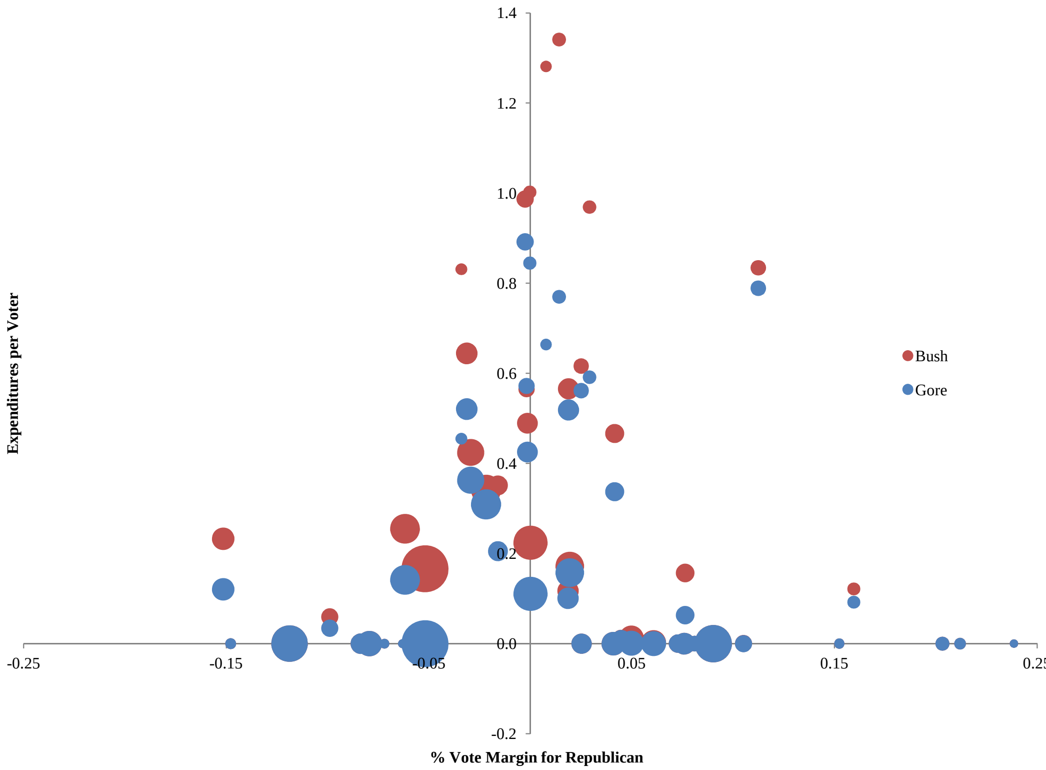
Our paper structures the discussion of these classes of estimators using a simple model of layered regions, and compares these estimators in the context of presidential advertising using the data and analysis from Gordon and Hartmann (2013). US presidential election advertising offers a unique context to study this comparison for three reasons. First, unlike many applications where representativeness concerns are confounded with whether a residual endogeneity bias exists, advertising in presidential elections has a mild, and potentially zero, endogeneity bias. Gordon and Hartmann (2013) argue this arises because a candidate's advertising is decreasing with unobserved demand shocks when leading its competitor in a market, but increasing when trailing. The non-monotonic relationship yields positive biases in some markets and negative biases in others; when an election is close there might be a nearly equal fraction of both markets. Their estimates support this in that a fixed effect estimate is nearly identical to an estimate using supply-side instrumental variables. Second, Huber and Arceneaux (2007) point out that the distinction of battleground states where advertising is concentrated creates unique cross-border variation in advertising, i.e. a portion of a non-battleground state might receive high advertising because it is located in a media market that also includes battleground states. Consumer preference variation being a primary driver of advertising would ordinarily leave the researchers with weak "supply-side" instruments, but the preference externality IV actually includes demand side variables because the preference variation can be treated as excluded from local demand after conditioning on local preferences. Third, the preference heterogeneity from heavily left to right leaning regions allows for a comparison of how the two estimators differ in terms of their representativeness concerns.

To the question of statistical power, we find that neither IVs nor the Border Approach increases standard errors enough to yield insignificant estimates. Supply-side instruments alone are weak with a low first-stage F after using the Kleibergen-Paap correction for clustered and robust standard errors. The first stage partial R squared is however 0.34. Standard errors increase by 71 percent over the fixed effects specification but the estimates are still significant with p-values less than 0.05. As expected, the preference externality (PE) IVs are much stronger with a first-stage excluded F of 48.7 and a partial R squared of 0.48. The important role of preference variables in explaining advertising variation is clear from inspecting Figure 1, reproduced from Gordon and Hartmann (2016). When the instrument strategies are combined to include both demand and supply-side determinants of advertising, the first stage F increases to 56.9 with a partial R squared of 0.64, yielding standard errors that are midway between the fixed effect and using either instrument strategy alone. When we turn to the Border Approach, the statistical power seems not to be a concern either. The largest standard errors in the Border Approach are at this same level and occur when we use only the small counties where the identifying assumptions are more likely to hold. When we apply the border approach to all border counties, standard errors are only 7 percent larger than the fixed effect estimate. We suspect the modest losses in statical power despite dropping so many observations result from the identifying variation existing at the aggregate level where no markets are dropped. The border fixed-effects may reduce variance as well.

The Border Approach and IV estimators differ substantially in the extent to which they can produce unrepresentative or "local" ad effects. The most severe threat of local average treatment effects (LATE) in IVs arise from non-compliance where, in our case, the presence of advertising may not be influenced by the instrument (Imbens and Angrist, 1994). Since advertising is a market level decision, the LATEs will disproportionately reflect some markets over others. Our point estimates for the supply-side IVs and PE IVs differ by 18 percent, but the difference is insignificant. When both sets of instruments are used, the point estimate is identical to the fixed effect. The robustness of the IVs and fixed effect estimate suggest LATE is not a concern here.

The Border Approach potentially introduces a geographically local ad effect at the disaggregate county level, by excluding all non-border counties across all (aggregate) markets analyzed. When analyzing all border counties, the point estimate of the Border Approach is 23 percent smaller, though not significantly different from the fixed effects and IVs. However, when we restrict the analysis to small border counties where the unconfoundedness assumption, that local demand shocks do not influence advertising, is more plausible, the point estimate drops by 83 percent to near zero and becomes insignificant, despite a standard error that is identical to the lowest IV specification. We run some robustness checks to validate that this is not an attenuation bias arising from measurement error problems in that county-level advertising is not actually observable and is instead proxied for by media market level advertising. This does not appear to be the case because small non-border counties do not produce such small estimates. We conclude that the difference arises because the estimated effects are local to small border counties and do not reflect the larger, more urban high population counties that were excluded here, but are included in all of the IV and fixed effect specifications. These counties which are unique for identification appear to also be

Figure 1: Cross-Market Variation in Advertising Expenditures: 2000 Presidential Election



Note: GRPs by state-level voting margin in 2000 election: horizontal axis is the state-level Republican vote share minus the Democrat 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 .

unique in their advertising effects.

In summary, the PE-IVs have more desirable identifying assumptions in that they are consistent with optimizing behavior and produce much more representative effects than border implementations of the preference externality estimator. Politics may be a case with extreme potential for unrepresentativeness, thereby overstating these concerns for the border implementation, but we see no reason why the IV would be worse in other advertising applications. For example, many advertising applications do not have so many zero advertising observations to exhibit the pure non-compliance that can exacerbate LATE concerns. As to whether border vs. non-border heterogeneity in ad effects might exist in other applications, the common case of branded packaged goods likely includes heterogeneity because of urban vs. rural differences in the availability of private label retailers such as Trader Joe's and Whole Foods.

We hope our discussion and analysis encourages work in literatures that would typically apply border approaches to also consider preference externality IVs. Alternatively, if observed determinants of preferences do not exist to implement the IV approach, border approaches may be an alternative way to exploit the preference externality for identification. Advertising, like many other applications, suffers from a scarcity of identification approaches and the preference externalities discussed here are a promising alternative that can be applied to applications as diverse as advertising, school choice and minimum wage analysis.

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 policy making 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, while the final section concludes.

2 A model of Advertising Decisions and Conversion in a Layering Structure

Preference externalities arise because a common policy affects heterogeneous individuals. Both border and PE-IV estimators exploit a structure whereby the unit of analysis (an individual, zip code, county etc.) is smaller than the policy making region (a state or Designated Market Area, DMA).

To compare the identifying assumptions of these estimators, we define an illustrative model of advertising allocation and demand response (analogous to a supply and demand side model) in which allocation decisions occur at a more aggregate level than demand (conversion). The allocation decisions are made for groupings of recipients (targets) 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 individuals types. For example, one group could be a time period when a given advertising intensity is held fixed. Or, the grouping could be based on search terms or any other targeting variable that is not specific to a single response observation.

We focus on the common practice of advertising to individuals grouped by geographic region. In addition to the preference externality estimators, this layering structure is also exploited by a different identification approach in advertising that we refer to as the Local Variation Approach. That approach is not our focus, but it raises potential concerns with advertising applications of preference externality estimators so we use its features to help define the model. The extant advertising identification papers that exploit this layering structure are summarized in the following table:

Approach	Paper	Level of Ad Decision	Level of Demand/Conversion Analysis
Border	Huber and Arceneaux (2007)	DMA	Individuals in state-DMA / sub-DMA
Border	Shapiro (2017)	DMA-Month/Week	(DMA-border / sub-DMA)-Month/Week
IV	Thomas (2018)	Nation-Week	DMA-Week
Local Variation	Hartmann and Klapper (2017)	Nation-Year	DMA-Year
Local Variation	Stephens-Davidowitz et al. (2017)	Nation-Year-Movie	DMA-Year-Movie

Huber and Arceneaux (2007) evaluates the effect of political advertising for individuals in non-battleground states and compares those in DMAs that do and do not overlap with a battleground state. Shapiro (2017) measures the effect of DMA level advertising on the counties within the DMA that are located on the border of another DMA. This strategy has also been applied in Shapiro (2018), Tuchman (2017) and Wang et al. (2018). Thomas (2018) develops a preference externality IV estimator where some ad decisions are made at the national level and conversion is measured at the DMA level. Hartmann and Klapper (2017) and Stephens-Davidowitz et al. (2017) evaluate the effect of nationally televised Super Bowl ads on sales measured at the DMA level. The Local Variation approach exploits exogenous variation in advertising viewership across the DMAs within the nation.

2.1 Conversion/Demand model

We begin by describing conversion at the more granular level k, then model the advertisers' decision at a more aggregate level l. To clarify notation for how individual decisions aggregate up to levels k and l, let $P = \{1, 2, ..., N\}$ be the entire population of consumers, and $i \in P$ denotes any consumer. Let $K = \{k_1, k_2, ..., k_C\}$ be a partition of P, representing the granular layer to analyze conversion. $L = \{l_1, l_2, ..., l_M\}$ is a coarser partition of K, representing the layer where advertisers are making advertising decisions. By coarser 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 gender, 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. Another example is DMA-based advertising decisions with county-level conversion available, where K is the set of counties, and L is the set of DMAs.

Let $q_k(A_k, x_{1k}, \xi_k)$ be the share of individuals in group k who choose the alternative. The choice is influenced by the level of advertising, A_k as well as x_{1k} and ξ_k which are exogenous determinants of the choice that are respectively observable and unobservable to the researcher. An example of such a model would be a simple logit discrete choice model for individual i, where:

$$\begin{aligned} q_k &= E(Y_{ki}) \\ \text{where } Y_{ki} &= 1 \text{ if} \\ u_{ki} &= \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_{1k} + \xi_k + \epsilon_{ki} > 0 \end{aligned}$$

where $g(\cdot)$ is a function that converts advertising levels to utility. This model implies

$$\ln(q_k) - \ln(1 - q_k) = \alpha_0 + \alpha_1 g(A_k) + \alpha_2 x_{1k} + \xi_k \tag{1}$$

The advertising realized in a given market, $A_k(d_l, x_{2k}, \tilde{z}_k, \tilde{\omega}_k)$ is a function of the variation in the coarser level l advertising decisions and/or variation in local realizations in viewership:

- Variation in advertising decisions
  - d_l: the advertising decision made at the more aggregate level l
- Variation in local realizations in viewership:[1]

[1] There exists a fourth component ψ_k, which are observable to the advertiser when advertising, but unobservable to the researcher. This may create extra bias if it is correlated with demand-side unobservable

- x_{2k}: determinants of advertising exposure which are observable both to the advertiser when the advertising decision is made and to the researcher
- $\tilde{z}_k$: determinants of advertising exposure which are unknown by the advertiser when the advertising decision is made but are observable to the researcher
- $\tilde{\omega}_k$: determinants of advertising exposure which are unknown by the advertiser when the advertising decision is made and are unobservable to the researcher
- $\tilde{v}_k = \{\tilde{z}_k, \tilde{\omega}_k\}$

We use tilde notation for variables that are unknown when advertisers make decisions and thus exogenous to demand shocks, such as weather, which might affect whether or not people are at home watching television when an ad airs. The logic of including each of these distinct factors will become clear when we discuss the assumptions and sources of variation in the identification strategies we consider.

2.2 Advertiser's problem

In making the advertising decision, d_l, to each group $l \in L$,[2] the advertiser solves the following problem[3]

$$\max_{d_{l_1},\ldots,d_{l_M}} \sum_{l \in L} \left(\sum_{k \subset l} n_k \left(\left(p_k - c\right) q_k \left(A_k, x_{1k}, \xi_k\right) - w_l \cdot A_k \right) \right), \tag{2}$$

where n_k is the population size in group k, p_k is the price for the product, q_k is the per-capita sales, c is the marginal cost, w_l is the per impression ad cost. The advertiser makes the advertising decisions d_l based on $(X_l, \xi_l) = \left(\{x_{1k}, x_{2k}\}_{k \subset l}, \{\xi_k\}_{k \subset l}\right)$ and w_l. Solving the problem in (2), we get

$$d_l \left(X_l, \xi_l, w_l\right). \tag{3}$$

We use a static model for illustration but similar characterizations would exist in a dynamic advertising model. Due to our focus on factors that might bias advertising effects, we omit marginal costs of production c from the ad decision, but they could be included if an

ξ_k. However, we do not explicitly list this component here, as the channel is implicitly captured by the effect of ξ_k on ads decision d_l.

[2] 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.

[3] There is slightly simplification of notation in the following equation. Literally, by $k \subset l$, we mean $k \subset l$ and $k \in K$.

identification strategy were based on them. Pricing p is another endogenous decision which also affects ad decisions but we can abstract away from it for the purposes of the analysis here. Gordon and Hartmann (2016) present an advertiser model for presidential candidates, but the Electoral College introduces unnecessary complexities to our analysis, so we use the more common firm problem here for illustration.

2.3 Preference Externalities

The preference externalities arise because local-level advertising, A_k, is a function of external determinants of preferences $X_{-k} \subset X_l$ and $\xi_{-k} \subset \xi_l$. Externalities could also arise from local random coefficients, but X and ξ are sufficient to characterize the border and IV implementations of preference externality estimators.

3 A Comparison of Identification Strategies

In this section, we use the above model to describe existing advertising identification strategies. The model also helps to clarify and evaluate the identifying assumptions. The parameter of interest is the ad effect α_1 in (1), reposted here as

$$y_k = \alpha_0 + \alpha_1 g\left(A_k\right) + \alpha_2 x_{1k} + \xi_k, \tag{4}$$

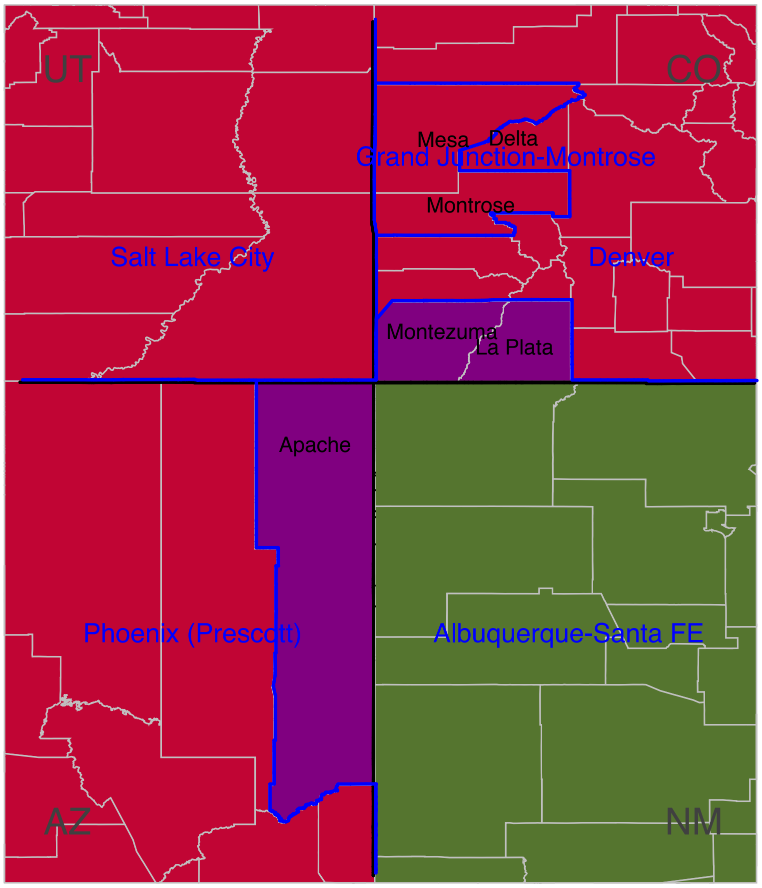
and the econometric endogeneity of A_k arises because it is a function of the demand shock ξ_k:

$$A_k\left(d_l\left(X_l, \xi_k, \xi_{-k}, w_l\right), x_{2k}, \tilde{z}_k, \tilde{\omega}_k\right). \tag{5}$$

3.1 Border Approach by Huber and Arceneaux (2007)

The border implementation of preference externality estimators was first introduced to advertising by Huber and Arceneaux (2007). Their idea was to identify ad effects by exploiting the fact that parts of some non-battleground states received more advertising than they should because their advertising region overlapped with a battleground state. One such example, depicted in Figure 2, is the Alburquerque DMA which covers the majority of New Mexico (a battleground state, shaded in green) and overlaps into small portions (counties) of Arizona and Colorado (Apache county in Arizona, Montezuma county and La Plata county in Colorado), both right-leaning non-battleground states. The overlapping region is shaded

Figure 2: Cross-State DMA Overlap Example & Huber and Arceneaux (2007) Identifying Assumptions



in purple here. The preference externality is the influence of political tastes in the green shaded Albuquerque DMA on the advertising levels in the purple regions. The three purple counties receive high levels of advertising from the Alburquerque DMA, whereas the remaining parts of the two states (shaded in red) receive low levels of advertising in their corresponding DMAs. If we think of advertising as binary, the purple region is a treatment group and the red regions are control groups.

Applying our model, the outcome equation is

$$y_k = \alpha_0 + \alpha_1 g\left(A_k\right) + \alpha_2 x_{1k} + \gamma_{s(k)} + \tilde{\xi}_k, \tag{6}$$

where $s\left(k\right)$ is the collection of counties defined by a state. The unobserved demand shock ξ_k is decomposed into two components: one that is common within the state and captured by the fixed-effect $\gamma_{s(k)}$, the other is the residual local shocks $\tilde{\xi}_k$. The unconfoundedness assumption that provides a consistent estimate of α_1 is therefore

$$E\left[\tilde{\xi}_k A_k | x_k, \gamma_{s(k)}\right] = 0. \tag{7}$$

There are two scenarios under which this assumption might hold. First, it could be that $\xi_k = \gamma_{s(k)}$ in which case $\tilde{\xi}_k = 0$. In words, the unconfoundedness assumption holds if the overlapping border county, k, has an error term that is perfectly correlated with that of the broader state in which it resides. If this assumption does not hold, then the residual unobservable, $\tilde{\xi}_k = \xi_k - \gamma_{s(k)}$, must be excluded from the advertising decision d_l which raises the concern that the advertiser is ignoring some local preferences. The battleground state justification for this seems compelling at first glance, but if the purple region represented a large populations share of the DMA, candidates might reduce DMA level advertising relative to a comparable DMA where the purple region is small. Exposures purchased and delivered to voters in the purple region effectively increase the cost of advertising to those in battleground state part of the DMA.

An irony about the identifying assumptions of the border approach relates to its spatial correlation properties. As described above, the unconfoundedness assumption holds if there is perfect correlation in unobservables across borders of a DMA. But, if that assumption does not hold, then unconfoundedness fails when there is any correlation within DMA borders. Specifically, if $\tilde{\xi}_k$ is correlated with $\xi_{k'}$ for any $k' \in l(k)$ and $k' \neq k$, then the unconfoundedness assumption is violated since $\xi_{k'}$ enters d_l and hence A_k. It could only be satisfied if the entire $\xi_l = \{\xi_k, \xi_{-k}\}$ is excluded from d_l, but then we would not have an endogeneity problem and we would have the implication that advertisers are ignoring information they should pay attention to (i.e. demand shocks in battleground states).

The state fixed effect, $\gamma_{s(k)}$, in Huber and Arceneaux (2007) is particularly valuable in the context of preference externality estimators because it creates a contrast between two different aggregations of counties (states and DMAs) and policy-making differs across these. They point out that campaigns focus most efforts at the state level, due to the Electoral College, except for TV advertising which uniquely occurs in the DMA aggregation. This helps separate the identification of advertising from other campaign activity. We apply this logic in a robustness check of our application of PE IVs below.

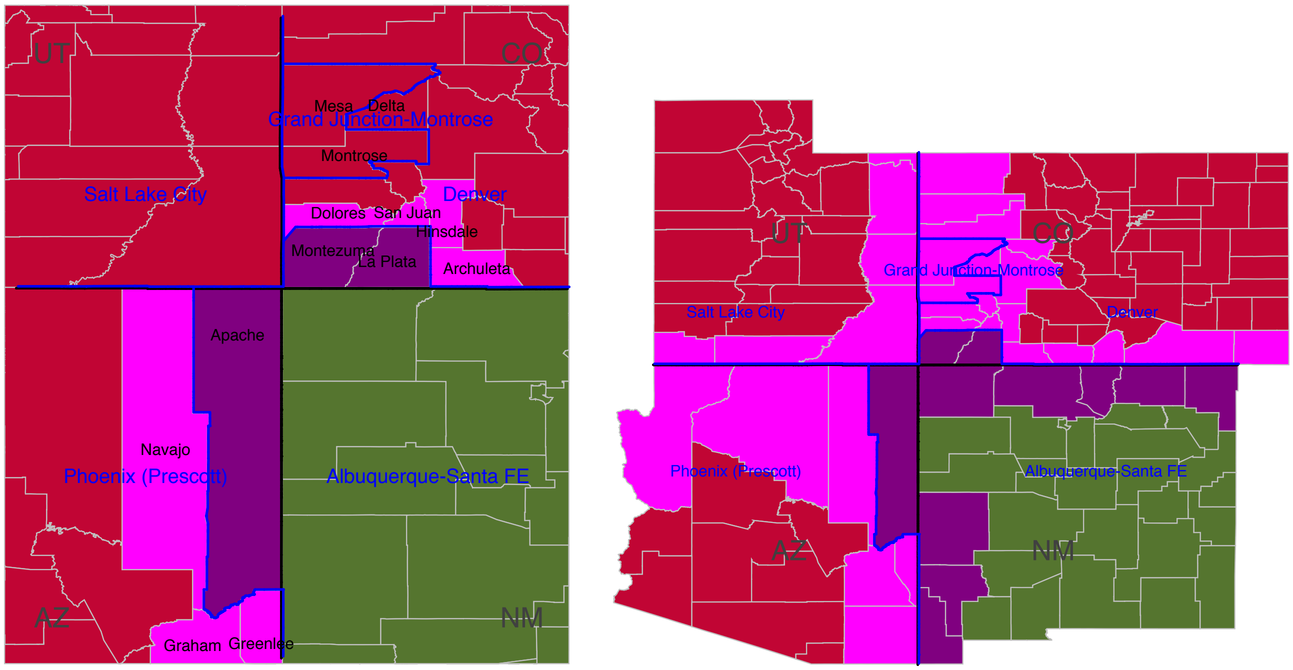
3.2 Border Approach Implementations Following Shapiro (2017)

Huber and Arceneaux (2007)'s border approach was attractive due to the Electoral College's ability to create these sharp border discontinuities in advertising, but that source of variation did not not generalize well to other advertising applications. Shapiro (2017) extended the approach to broader advertising applications by changing the fixed effect in Equation 7 from $\gamma_{s(k)}$ to $\gamma_{b(k)}$, where $b(k)$ is a collection of counties lying on both sides of the same DMA border. This follows the same logic as Dube et al. (2010)'s extension of Card and Krueger (1994)'s approach of cross-border comparisons to evaluate minimum wage effects. If we use the same example in the previous subsection for illustration in the first panel in Figure 3, the more lightly shaded pink counties (Navajo, Graham, and Greenlee counties in Arizona, and Dolores, San Juan, Hinsdale, and Archuleta counties in Colorado) take the place of the larger red areas (Phoenix DMA and Denver DMA, respectively) to be the effective "control" counties for the darker purple counties in the Alburquerque DMA that formed the treatment in Huber and Arceneaux (2007). The smaller control region increases the plausibility of the unconfoundedness assumption. Further, there is no explicit asymmetry between treatment and control counties other than the fact that pink and dark purple counties are experiencing various levels of treatment A_k. The preference externality for the pink counties comes from the other parts of their DMAs, shaded in red. This also greatly expands the number of treatment counties to include all DMA borders, shaded in purple along the 4 corners states in the second panel of Figure 3; each with their respective externality region.

The identifying assumptions, depicted in the right panel of the figure, are exactly the same as in Huber and Arceneaux (2007), but for the fixed effect change, i.e. $E\left[\tilde{\xi}_k A_k | x_k, \gamma_{b(k)}\right] = 0$, i.e., $E\left[\tilde{\xi}_k \xi_{k'} | x_k, \gamma_{b(k)}\right] = 0$ for any $k' \in l(k)$ and $k' \neq k$.

The same spatial correlation irony exists, though it becomes a little more ironic as we consider some DMA border counties that do not cross state boundaries. One example is the three counties in Colorado lying on DMA borders: Montrose and Mesa in Grand Junction-Montrose DMA, and Delta in Denver DMA. The perfect correlation argument for unconfoundedness would argue that Mesa and Delta counties which cross DMA borders are perfectly correlated. But if they are not perfectly correlated, then there must be no spatial correlation within DMA between Mesa and Montrose. The irony comes from the relative positions of the three counties, as they seem to be symmetric. The perfect correlation is also unlikely to exist across DMA borders for many potential pairs, e.g. in pairs of counties in which the extent of urbanization is asymmetric, such as Los Angeles and Bakersfield, or Santa Clara (Silicon Valley) and Merced counties in California. But its also not plausible

Figure 3: Cross-State DMA Overlap Example & Shapiro (2017) Identifying Assumptions



that demand is uncorrelated within DMA for many of these examples such as Los Angeles and Orange or Ventura counties in the Los Angeles DMA, or Santa Clara and San Mateo counties (where Stanford sits along the border in the San Francisco-Oakland-San Jose DMA), i.e. pairs of counties that are relatively comparable in the extent to which they are urbanized and populated. This broader southwest depiction of DMAs is shown in Figure 4.

Further, the exclusion of the local demand shocks from the advertising decision, d_l is implausible for border counties that represent a large share of their DMA, such as Los Angeles, Las Vegas, Phoenix, Tucson, El Paso, Santa Clara, Alameda etc. It is certainly not plausible when the DMA and county are one in the same, as in Bakersfield and San Diego as shaded in green. Practical implementations of the border approach will drop such counties.

More generally. in Table 1, we use 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. Two border counties represent more than 90% of their DMA, and two more between 70–80 percent of the DMA share. In light of this concern, in our empirical application below, we consider estimators with both all border counties, and only those that represent a small share of the DMA, such that the unconfoundedness assumption is plausible. Among counties representing less than 10 percent of their DMAs' VAP, Figure 5 shows that the distribution of counties' share of

Figure 4: Southwest DMAs

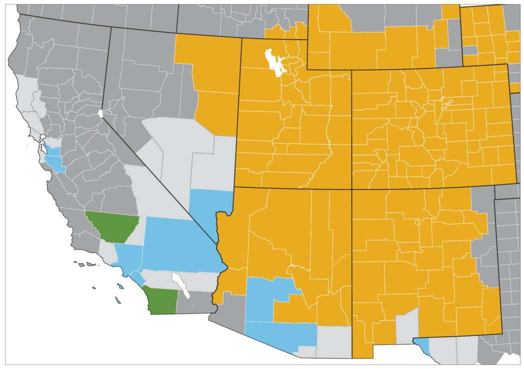


Table 1: Border vs. Non-Border Counties' Percent of DMA Population: Political Ads

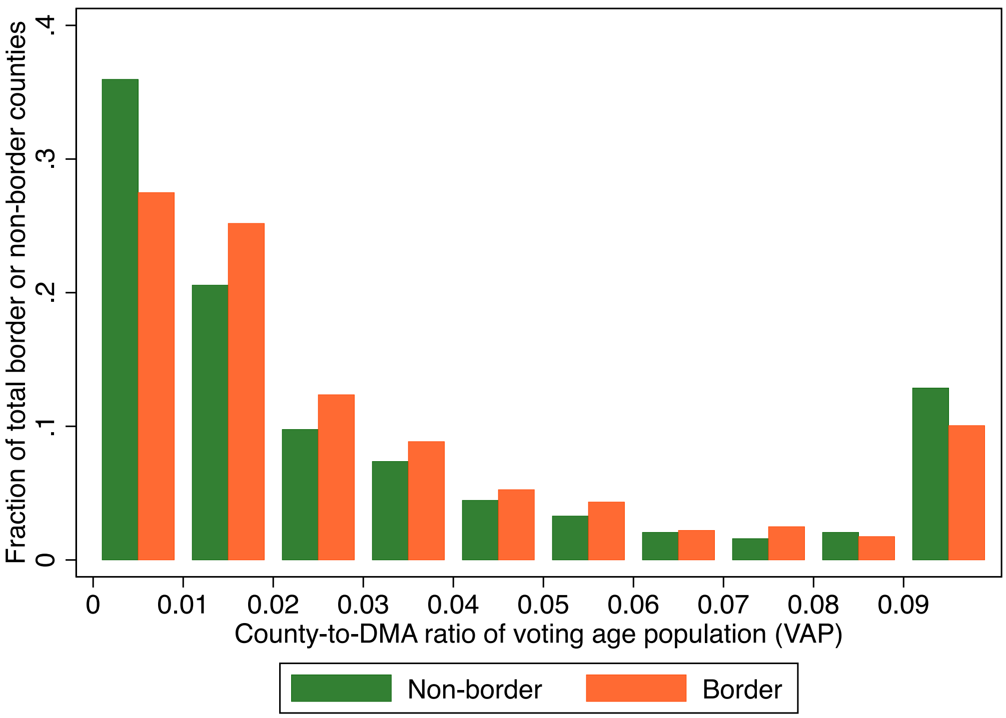
County VAP / DMA VAP	% of Non-Border Counties	% of Border Counties
0-0.1	88.4	91.7
0.1-0.2	5.9	4.4
0.2-0.3	2.4	1.5
0.3-0.4	1.4	0.9
0.4-0.5	0.9	0.7
0.5-0.6	0.6	0.0
0.6-0.7	0.3	0.0
0.7-0.8	0.1	0.4
0.8-0.9	0.0	0.0
0.9-1	0.1	0.4
N	1,065	542

the DMA is similar for border and non-border counties.

3.3 Preference Externality Instrumental Variables

The border approach had the appeal that it could be applied broadly which is particularly valuable with a scarcity of identification approaches for advertising, but the above limitations led us to look elsewhere. Trying to apply the same intuitive preference externality variation, we began to explore what we term the preference externality IV. Not realized by us initially, this is the same estimator applied by Gentzkow and Shapiro (2010) and Fan (2013) and generalized in the context of differentiated products by Berry and Haile (2010). Also unknown to us at the time or perhaps subsequently, Thomas (2018) developed a preference externality IV that focused on national advertising and DMA level measurement as

Figure 5: Distribution of Counties' Share of DMA Population by DMA Border or Not



Notes: 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.

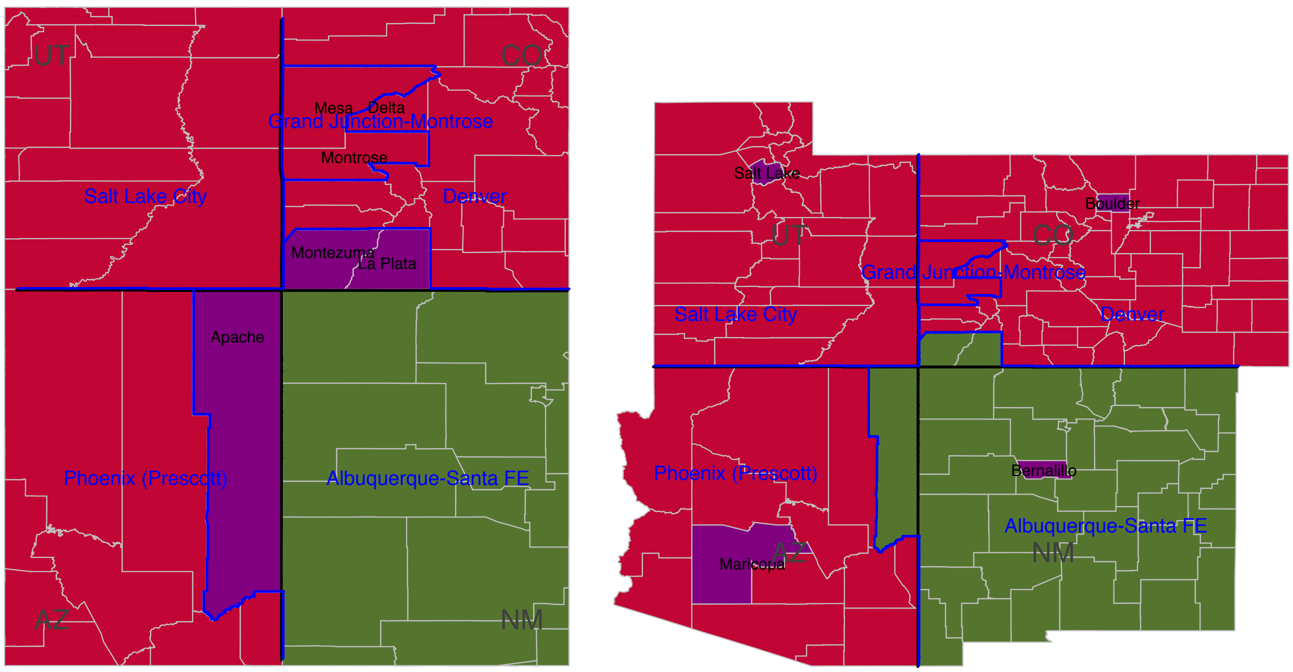
opposed to DMA advertising and county level measurement. By focusing on the national level, the estimator becomes a time-series estimator with likely fewer advertising decisions to gain statistical power (i.e. standard errors would likely need to be clustered at the national level, of which there is one, or perhaps at the time level).

The idea of the preference externality IV is that essentially every county, big or small, at a border or not, realizes an externality as long as there is preference heterogeneity across counties in the DMA and the county does not constitute its own DMA. The left panel of Figure 6 depicts the simple case without any fixed effects where we can begin the discussion of the identifying assumptions. Like the border approach, the preference externality comes from the green region to determine the advertising for a focal purple county. The identifying assumption here is that, $E\left[\xi_k Z_l | X_k\right] = 0$, or written explicitly to note the externality but including some redundancy in the conditioning arguments, $E\left[\xi_k Z_{-k} | Z_k, X_k\right] = 0$. The intuition being that for some observable determinant of preferences, $Z_l \in X_l$, Z_{-k} can be treated as excluded and exogenous upon conditioning on that same variable for the region of analysis, i.e. Z_k. For example, if a hybrid automobile manufacturer did not find it profitable to advertise to right leaning customers, some right leaning customers in the purple region of the red state as indicated by Z_k might realize advertising because there are more left leaning customers elsewhere in their DMA, Z_{-k} than in the other DMAs in their home state.

The appeal of the preference IV estimator over the border approach is to also include analysis of larger metropolitan areas that might not lie on borders or are too large to satisfy the unconfoundedness assumptions of the border approach. For example, the right panel of Figure 6 depicts counties containing Salt Lake City, Phoenix (Maricopa county), Albuquerque (Bernalillo county) and Boulder in purple, recognizing that even if some of those urban counties are comparable, they might receive different advertising because of the externalities they realize from other parts of their DMAs as depicted in shades of green or red.

The outcome equation in this case follows Equation 4, with the instruments, Z_l, including market level variables that explain advertising and are excluded from Equation 4 after conditioning on the local realization of those variables, i.e. a $Z_k \in x_k$. The primary concern for identification is that Z_{-k} might influence local outcomes through a route other than advertising. Consider the hybrid vehicle example again. The left or right leaning preferences in other counties might also influence whether a state offers carpool stickers for hybrid vehicles. The key to resolving this goes back to Huber and Arceneaux (2007)'s insight that the difference between state and DMA groupings of counties helps separate advertising from other

Figure 6: DMAs and the Preference Externality IV



policies. We can simply add a state fixed effect by similarly decomposing $\xi_k = \gamma_{s(k)} + \tilde{\xi}_k$. Unique to the PE IV approach we can also include other aggregations of Z into the right hand side of equation 4. For instance, $Z_{-k\in s(k)}$ is in the right-hand side of the outcome equation, but $Z_{-k\in l(k)}$ is still excluded from the outcome equation but included in the first stage. We demonstrate a simple robustness check along these lines below.

We could similarly account for spillovers from neighboring counties or the center of DMAs. We only require that the county Zs relevant to non-advertising decisions are not exactly the same as included in the DMA. In principle, we could also follow Shapiro (2017)'s conditioning approach and include fixed effects for contiguous regions.

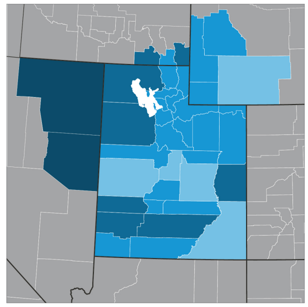
3.4 Other Advertising Identification Approaches

3.4.1 Supply-side instrumental variables

We briefly introduce the supply-side instrumental variables strategy applied by Gordon and Hartmann (2013). They use the ad cost w_l as an instrument for ad viewership A_k. Although this identification strategy does not make use of the layering structure,[4] it is also consistent

[4]The identifying variation exists at the aggregate level l, i.e., $E\left[w_l \xi_k\right] = 0, \forall k \subset l$, but analysis could be conducted at the granular level relying on $E\left[w_{l(k)} \xi_k\right] = 0$.

Figure 7: A Hypothetical Example of Local Advertising in Salt Lake



with our derived model in (4) and (5). Typical concerns with regard to instrumental variables arise, as we discuss below. They actually lag the ad cost to non-election years to avoid advertising demand shocks in election years that might influence ad prices.

3.5 Local Variation Approach

The Local Variation Approach of Hartmann and Klapper (2017) and Stephens-Davidowitz et al. (2017) is not a preference externality estimator but exploits the layering structure described in our model. The reason we cannot apply it to our data is also a potential concern for advertising implementations of preference externality estimators and some other identification approaches. Namely, the local variation approach relies on local, k-level, variation in the viewership of advertising, v_k. In the Local Variation applications, Super Bowl ads are national decisions, such that l is the nation. Conversion is measured at a k-level of DMAs. The same estimator could be applicable to DMA level ad decisions, where the k level is a county just as in the previous discussions. Such advertising variation is hypothetically depicted for the Salt Lake DMA in Figure 7 where different shades of blue reflect different levels of viewership of a DMA-level ad. If the variation in viewership, $\tilde{v}_k$, were exogenous, the local variation approach could be applied.

The challenge is that none of the advertising papers cited herein actually observe advertising at the county level. They assume that local realizations of advertising are uniform across the entire DMA. This can create a measurement error problem that we explore in the appendix and rule out below in our empirical application.

4 Empirical Application

While 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 instrumental variables or border strategies.

Gordon and Hartmann (2013) point out and find that endogeneity biases are not a substantial concern in the case of political advertising. A candidate's incentive to advertise diminishes if the unobserved (to the researcher) determinants of voters preferences for the candidate is either too high or too low. In political language, candidates only tend to advertise in battlegrounds where voters neither like them too much nor too little. Gordon and Hartmann (2013) show this pattern as evident in the distribution of advertising across the political leaning of advertising markets in Figure 1, and then document that fixed effects estimates yield nearly identical estimates to those using supply-side instruments (ad prices from the year before the election). We replicate these results and then compare them to specifications which separately apply the Border Approach and PE IVs.

Political advertising also highlights the unique and important value of the PE IVs inclusion of demand determinants of advertising as sources of identification. The variation in Figure 1 is extensive and systematic, but the ad prices Gordon and Hartmann (2013) use as instruments likely explains a limited part of that variation. Most of the advertising incentives are driven by demand side factors. To the extent 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 PE IV approach relies on identifying variation that should drive much more of the observed distribution of advertising. We document this in the following estimates.

4.1 Gordon and Hartmann (2013) Fixed Effects and Supply-side IV Specifications

Table 2 provides the estimates using the model and data from Gordon and Hartmann (2013). The first two columns are the fixed effects and supply-side IV specifications from their paper. Notably, the estimates are incredibly close, despite the supply-side IV estimate relying on variation in how candidates separately incorporate ad prices across different day parts into their advertising decisions. The partial R^2 in the IV specification is 0.34 suggesting the ad prices do explain a reasonable amount of variation above and beyond the fixed effects and demand side observables at the county level. The F statistic is however quite low.[5] The instruments are therefore weak, explaining an increase in the standard errors reported, yet not so weak as to yield the estimates insignificant.

4.2 Preference Externaltiy Instruments

Next, we consider the PE IVs that introduce demand side determinants of advertising into the instruments. Gordon and Hartmann (2013) include a variety of demand side observables measured at the county (and more aggregate levels) which would be characterized as x_{1k} in the model formulation above. The excluded instruments used for identification in this analysis are their DMA-aggregation. Column (3) reports the results. The first stage partial R^2 for this strategy is much higher at 0.49 relative to the 0.34 when using the lagged ad price instruments, suggesting demand instruments do explain more variation in the advertising levels. The F statistic is much higher at 44.9, suggesting the PE IVs are quite "strong" in explaining the variation of ad exposure, mitigating concerns about "weak instruments". The estimates with this strategy are about 20 percent higher than in the Gordon and Hartmann (2013) specifications, but that difference is not statistically significant.

Column (4) uses both the PE IVs and lagged ad price IVs to include both demand and supply-side determinants of advertising as instruments. In this case the first stage partial R-squared jumps to 0.64 and the F-statistic reaches 61.2. 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 found in the fixed effects and IV specifications Gordon and Hartmann (2013) estimated. From this we conclude that neither weak instrument nor LATE concerns arise in the IV specifications.

4.3 Border Approach in Political Advertising Data (all border counties)

Finally, we apply the Border Approach based on Shapiro (2017). In these specifications, we shift away from instrumental variables approaches that can be justified by the model of advertiser's behavior, to an "unconfoundedness approach" that restricts analysis to a

[5]The original paper reported an F of 88.2, but it becomes low as 2.6 after Kleibergen-Paap correction for cluster and robust standard errors.

Table 2: 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 as IV	No	Yes	No	Yes	No	No
PE IVs	No	No	Yes	Yes	No	No
Observations	6428	6428	6428	6428	2540	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 R2		0.336	0.484	0.639		
Clustered SE	party-dma				party-dma party-border	

Notes: Columns 1–4 use 1,607 counties for two parties in two years, columns 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 is 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), snow(2004). "Lagged ads prices IV" include lagged ads price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. "PE IV" include the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Clustered standard errors in parenthesis.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

subset of regions where local differences in demand are assumed to be excluded from the advertiser decision. As Shapiro (2017) notes, a critique this approach has faced is that the estimates may be "local" to only those counties in border regions and not generalizable to the majority of counties in the country. 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. However, by focusing on the current political case where endogeneity biases are less likely to be significant, these should be separable. These estimates in column (5) are conducted on 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 in two years. The advertising coefficient using the border strategy is about 20 percent lower but based on the standard errors this is not a significant difference from the preceding estimates. Standard errors are only slightly larger than the fixed effects which include 2.5 times as many observations. We believe this is because our primary identifying variation arises at the DMA level, and there are the same number of DMAs involved in the Border Approach.

4.4 Border Approach in Political Advertising Data (small border counties)

As described at the end of section 3.2, some border counties represent a large share of the DMA such that the identifying assumption, $E\left[A_l\left(\xi_k - \gamma_{b(k)}\right)\right] = 0$, is unlikely to hold. We therefore restrict our analysis to counties in small borders where this assumption is more plausible for borders that represent a small share of the DMA. While the lack of a clear endogeneity bias suggests this restricted sample will not resolve a bias, it is possible that small borders may be more likely to have "local", spatially heterogeneous effects.

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[6]. We have identified 35 out of 93 borders to be small, which corresponds to 213 out of 635 border-counties. 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-squared is larger than when analyzing all border counties. The notable difference is that the estimated ad coefficient drops to nearly zero.

In order to further understand the difference in estimates from all border counties and

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

Table 3: Local Effects for Counties in Small Borders

	DV: Ln(share) - Ln(share0)			
	(1)	(2)	(3)	(4)
Ln(Ads)	0.015	-0.000	0.008	0.011
	(0.018)	(0.024)	(0.025)	(0.018)
Controls	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes
Party-dma FE	Yes	Yes	Yes	Yes
Lagged ads prices as IV	No	Yes	No	Yes
PE IV	No	No	Yes	Yes
Observations	784	784	784	784
R^2	0.772	0.772	0.772	0.772
First-stage excluded F		4.307	49.534	416.529
First-stage partial R2		0.453	0.619	0.903

Notes: All columns use 196 counties for two parties in two years; these counties belong to 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 is 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), snow(2004). "Lagged ads prices IV" include lagged ads price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. "PE IV" include the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Standard errors in parenthesis are clustered at the level of party-DMA.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

small border counties (columns 5 and 6 of Table 2), we complement with the following two sets of results.

For one, we further show that such difference does not result from a better fit of the identification assumptions in the Border Approach because the same distortions occur when applying other estimators to these same counties. Replicating the fixed-effects and IV specifications (columns 1–4 in Table 2) to the same set of counties in small borders in Table 3, we find all fixed effects and IV specifications also have small insignificant results when estimated on that sample of counties in small borders. This result suggests that differences in the fit for identification assumptions, or the extent to which local demand shocks are unconfounded, are unlikely to account for the varying estimates in small border counties. Rather, it is likely that estimates in small border regions are driven by local effects, i.e., advertising responsiveness in small border counties are different from other counties.

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 A to show that small

Table 4: Comparison of Effects in Small and Large Counties

	DV: Ln(share) - Ln(share0)			
	Fixed effects		Supply-side IV	
County size	small	large	small	large
	(1)	(2)	(3)	(4)
Ln(Ads)	0.045*	0.060***	0.064**	0.041*
	(0.023)	(0.012)	(0.030)	(0.022)
Controls	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes
Party-dma FE	Yes	Yes	Yes	Yes
Lagged ads prices as IV	No	No	Yes	Yes
Observations	2080	4348	2080	4348
R^2	0.681	0.654	0.680	0.654
First-stage excluded F			10.655	2.100
First-stage partial R2			0.670	0.216
Clustered SE	party-dma			

Notes: Columns 1 and 3 use 520 small counties, defined as counties whose population is below 1% of their DMA for both years, columns 2 and 4 uses the remaining 1,087 counties. Ads is 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), snow(2004). "Lagged ads prices IV" include lagged ads price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. Clustered standard errors in parenthesis.
$^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

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. To assess the measurement error, we run Gordon and Hartmann (2013)'s analyses on small vs large counties to see whether a small population share would reduce their estimated effects. Table 4 demonstrates that small counties do not generally tend to exhibit lower estimated effects. By elimination, this result suggests local effects are the likely rationale for different effects within the small border regions.

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. But, it is also quite likely that these unique people may be influenced differently for antidepressant campaigns focused on population centers of media markets (e.g. Shapiro, 2017) and for consumer packaged campaigns (Tuchman et al., 2018) than those in the localities that are more likely to have Whole Foods, Trader Joes and other grocery outlets that do not carry or similarly emphasize traditional packaged goods. Unfortunately, it is not possible to test whether "local effects" exist in these applications

In Appendix B, we explore alternative border implementations where we pair counties (as opposed to Shapiro, 2017's groups of counties) on a border. This alternative implementation allows us to leverage more observations to test for the locality of small border counties. The general conclusion holds that small border counties appear to yield systematically different inference that is common to the various estimators we consider.

4.5 Controlling Spillovers in PE IVs

The identifying assumptions for the PE IV is that, conditional on the demographics in the focal county, our identifying variation – aggregate level of demographical variations – 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 geographical 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 concern, we simply 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 2, which explicitly reports the coefficient for the unemployment rate and annual income. We find that both labor-market demographics affect voter's utility level 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 is further driven down to 0.046. The coefficient for the unemployment rate in the largest county is, however, positive, which we suspect reflects the non-linear dependence of the ad investment in our utility specification, as this variable affects the ad level greatly. In the last column, we include both sets of controls for spillovers and the estimate for ad effect is further dropped to 0.044, but still significant. We suspect this is because the PE IV is now more selective about which types of counties produce identifying variation.

Table 5: Controlling for Labor Market Spillovers

	DV: Ln(share) - Ln(share0)			
	(1)	(2)	(3)	(4)
Ln(Ads)	0.053***	0.050***	0.046**	0.044**
	(0.019)	(0.019)	(0.020)	(0.020)
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			0.036***	0.032**
			(0.011)	(0.012)
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
Lagged ads prices as IV	Yes	Yes	Yes	Yes
Observations	6428	6428	6428	6428
R^2	0.629	0.631	0.630	0.632
First-stage excluded F	57.742	58.642	20.853	20.927
First-stage partial R2	0.639	0.638	0.607	0.607

Notes: All columns use 1,607 counties for two parties in two years. Ads is 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), snow(2004). "Lagged ads prices IV" include lagged ads price (CPM) in different day time, interacted with party and year dummy, as instrumental variables. "PE IV" include the DMA-aggregate controls, interacted with party and year dummy, as instrumental variables. Clustered standard errors in parenthesis.
$^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

4.6 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 this data, we should at least expect the fixed effect estimate to be much different. An application of Shapiro (2017)'s Border Approach provides an effect which is smaller but not statistically different. However, this does not hold when borders that represent a non-trivial share of the DMA are excluded from the analysis.

Instrumental variables are not immune from locality in estimates, but the robustness of the PE IVs to two other, fundamentally different, sources of variation suggests no problem in our application. We suspect 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, the Border Approach systematically discards all high population, more urban, counties.

5 Conclusion

We have provided a framework that unites border approaches and instrumental variable approaches that both 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 local effects. We have shown the identifying assumptions in the Border Approach add a complicated and implausible structure of spatial correlation and assume ignorance of some unique local preferences, while the IV implementation is consistent with policy-makers placing positive weight on the preferences of all consituents or customers. 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 externality instruments to supply-side instruments 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. Border strategies can produce local effects based on the types of counties included. We find ad effects are systematically different for all identification strategies when applied to the small border counties where identifying assumptions in border approaches are most plausible.

References

Berry, S. T. and P. A. Haile (2010). Nonparametric identification of multinomial choice demand models with heterogeneous consumers. Technical report, National Bureau of Economic Research.

Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. *Quarterly Journal of Economics 114*(2), 577–599.

Bound, J., D. Jaeger, and R. M. Baker (1995). Problems with instrumental variables estimation when the correlation between instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association 90*(430), 443–450.

Card, D. and A. Krueger (1994). Minimum wages and employment: A case study of the new jersey and pennsylvania fast food industries. *American Economic Review 84*(4), 772–793.

Dube, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *Review of Economics and Statistics 92*(4), 945–964.

Fan, Y. (2013). Ownership consolidation and product characteristics: A study of the us daily newspaper market. *American Economic Review 103*(5), 1598–1628.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

George, L. and J. Waldfogel (2003). Who affects whome in daily newspaper markets? *Journal of Political Economy 111*(4), 765–784.

Gordon, B. R. and W. R. Hartmann (2013). Advertising effects in presidential elections. *Marketing Science 32*(1), 19–35.

Gordon, B. R. and W. R. Hartmann (2016). Advertising competition in presidential elections. *Quantitative Marketing and Economics 14*(1).

Hartmann, W. R. and D. Klapper (2017). Super bowl ads. *Marketing Science, forthcoming*.

Huber, G. A. and K. Arceneaux (2007). Identifying the persuasive effects of presidential advertising. *American Journal of Political Science 51*(4), 957–977.

Imbens, G. W. and J. D. Angrist (1994). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Masten, M. A. and A. Torgovitsky (2016). Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics 98*(5), 1001–1005.

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

Shapiro, B. T. (2017). Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants. *Journal of Political Economy, forthcoming*.

Shapiro, B. T. (2018). Advertising in health insurance markets. *Marketing Science, forthcoming*.

Stephens-Davidowitz, S., H. Varian, and M. Smith (2017). Super returns to super bowl ads. *Quantitative Marketing and Economics 15*(1), 1–28.

Thomas, M. (2018). Mass advertising as a natural experiment. *Quantitative Marketing and Economics Conference*.

Tuchman, A. (2017). Advertising and demand for addictive goods: The effects of e-cigarette advertising. *working paper*.

Tuchman, A., G. J. Hitsch, and B. T. Shapiro (2018). Heterogeneity in own and competitive advertising effects across categories. *working paper*.

Waldfogel, J. (1999). Preference externalities: An empirical study of who benefits from whom in differentiated products. *NBER Working Paper No 7391*.

Wang, Y., M. Lewis, and D. A. Schweidel (2018). A border strategy analysis of ad source and message tone in senatorial campaigns. *Marketing Science*.

A Measurement Error

We mentioned previously, that in contrast with the Local Variation Approach that relies on local variation of viewership to identify ad effects, in most applications of the Border Approach, this variation is assumed away and the local ad exposure is measured at an aggregate level. However, if this assumption is not true, i.e., there exists local variation of viewership but researchers can only observe an aggregate measure, this will lead to a measurement error problem. In this section, we first demonstrate the existence of such problem, then the relevance of it to the identification strategies discussed in this paper. Finally we discuss the potential candidate solutions.

A.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:

$$y_k = \alpha_0 + \alpha_1 A_k + \alpha_2 x_{1k} + \xi_k \tag{8}$$

with local ads exposure A_k determined by

$$A_k\left(d_l, x_{2k}, \tilde{z}_k, \tilde{w}_k\right)$$

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

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

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

$$A_{-k} = \frac{1}{n_{-k}} \sum_{j \neq k} n_j A_j = A_{-k}\left(d_l, x_{2,-k}, \tilde{z}_{-k}, \tilde{w}_{-k}\right).$$

By simple algebra, we have

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

Plugging this into (8), we have

$$\begin{aligned} y_k &= \alpha_0 + \alpha_1 \left(\frac{n_l}{n_k} A_l \right) + \alpha_2 x_{1k} - \alpha_1 \frac{n_{-k}}{n_k} A_{-k} + \xi_k \\ &= \alpha_0 + \left(\alpha_1 \frac{n_l}{n_k} \right) A_l + \alpha_2 x_{1k} + (u_k + \xi_k) \end{aligned} \tag{9}$$

There are three source of bias in (9):

1. $\text{Cov}\,(A_l, u_k) \neq 0$, simply because u_k is a part of A_l.

2. $\text{Cov}\,(A_l, \xi_k) \neq 0$, since ad decisions are endogenously made.

3. $\text{Cov}\,(x_{1k}, u_k) \neq 0$, as x_{1k} is also considered in the ad decision, i.e., $\text{Cov}\,(x_{1k}, d_l) \neq 0$.

A.2 Three Identification Approaches Revisited

With the measurement error in mind, we examine the validity of three identification approaches discussed.

For the Local Variation Approach, there is no measurement error problem, as the local level ad exposure A_k is observed, which is at the same level as conversion y_k.

For the Border Approach, even if the ad decision d_l is unconfounded with respect to local demand shocks ξ_k in (9), we still have the measurement error problem, so that the first and the third source of bias still exist. The estimated coefficient may even become negative as $\text{Cov}\,(A_l, u_k) < 0$.

For the PE IV Approach, $h\,(X_{1l})$ is not a valid instrument anymore, as X_{1l} affects the ad decision d_l, which is further correlated with measurement error u_k.

In a 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, thus correlating with measurement error u_k naturally. In our framework, the only candidate instrument is $\tilde{z}_k$, which shifts the ad exposure in an exogenous manner (say, weather). It is unknown to the advertiser when making ad decisions, thus orthogonal to u_k. And it is correlated with coarser level ad exposure A_l through affecting the A_k. We talk about the the validity of such instrumental variable solution in the next subsection.

A.3 Instrumental variable solutions

For simplicity, we re-write (9) as

$$y_k = \alpha_0 + \alpha_1 A_l + \alpha_2 x_{1k} + \nu_k \tag{10}$$

where

$$\nu_k = u_k + \xi_k = -\alpha_1 \frac{n_{-k}}{n_k} A_{-k} + \xi_k$$

As we showed before, both A_l and x_{1k} in (10) are correlated with the ad decision, and thus endogenous to the error term u_k. We seek instruments for both endogenous variables, if possible.

For ad exposure A_l, we have candidate instrument $\tilde{z}_k$ which shifts the local ad variation A_k exogenously. To guarantee the validity of this instrumental variable, we need to assume out the spatial correlation between $\tilde{z}_k$ across different sub-region k within the same coarser region l, or else $\tilde{z}_k$ is spatially correlated with $\tilde{z}_{-k}$, which determines A_{-k} and is thus correlated with the error u_k.

Even if we assume the existence of the spatially-uncorrelated viewership shocks $\tilde{z}_k$, we still need some instruments for x_{1k}. This is quite difficult, for the same reasons as before, that we need to find some variables that are correlated with demographics x_{1k} but excluded from the ad decision d_l. Fortunately, we don't need to find such instruments if the only parameter of interest is α_1 instead of α_2. We can put the $\alpha_2 x_{1k}$ into the error term. The exclusion restriction of z_k is still valid as $\mathrm{Cov}\,(x_{1k}, \tilde{z}_k) = 0$.

Alternatively, we can use a "control function" and decompose the error term ν_k as

$$\nu_k = \rho x_{1k} + \omega_k$$

with

$$\mathrm{Cov}\,(x_{1k}, \omega_k) = 0$$

and plug in (10) to get

$$y_k = \alpha_0 + \alpha_1 A_l + (\alpha_2 + \rho)\, x_{1k} + \omega_k$$

and use $\tilde{z}_k$ to instrument for A_l. The exclusion restriction is still valid for $\mathrm{Cov}\,(\tilde{z}_k, \omega_k) = \mathrm{Cov}\,(\tilde{z}_k, \nu_k) - \rho \mathrm{Cov}\,(\tilde{z}_k, x_{1k}) = 0$. There is no advantage of a control function approach over a put-in-error approach, as neither provides consistent estimates of α_2.

A.4 Quantifying the Bias

The instrumental variable approach described in the previous subsection is quite restrictive. In this subsection, we quantify the bias (relative to α_1) in applying OLS on (10), instead of seeking valid instruments. For simplicity, we ignore the term of $\alpha_2 x_{1k}$, and thus the third source of endogeneity in A.1. The formula for the bias is thus

$$\hat{\alpha}_{1,OLS} - \alpha_1 = \frac{\text{Cov}\left(A_l, \nu_k\right)}{\text{Var}\left(A_l\right)}.$$

After some calculation (detailed proof in A.5) we have

$$\hat{\alpha}_{1,OLS} = \alpha_1 \frac{\text{Cov}\left(A_l, A_k\right)}{\text{Var}\left(A_l\right)} + \frac{\text{Cov}\left(A_l, \xi_k\right)}{\text{Var}\left(A_l\right)} \tag{11}$$

$$= \alpha_1 B_{1k} + B_{2k}, \tag{12}$$

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

Classical measurement error is usually caused by attenuation bias. If we look at the term in B_{1k}, since we do not directly observe A_k, and A_l is an errored measure of A_k (say, $A_l = A_k + e$), $\text{Cov}\left(A_l, A_k\right) < \text{Var}\left(A_l\right)$ and $B_{1k} < 1$, implying downward bias. However, in the current context of counties within the same DMA,

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

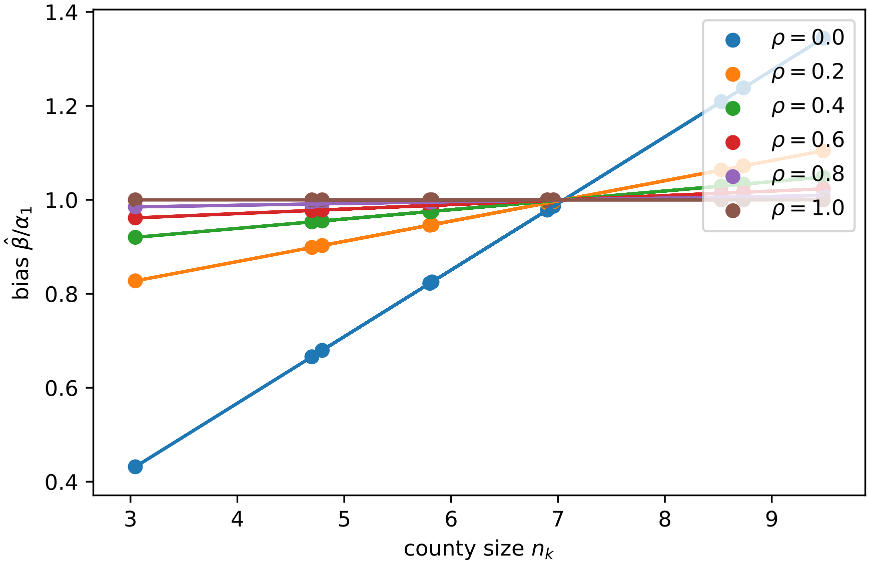
the first term implies that DMA-level ads exposure A_l is a weighted average of county-level A_k thus with smaller variance, which will bias the estimates upwards; the second term may drive the estimates downward, but it is not pure noise, which is different from the classical measurement error.

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

$$B_{1k} = \frac{\rho + (1-\rho)\frac{n_k}{n_l}}{\sum_{j \subset l} \frac{n_j}{n_l} \left(\rho + (1-\rho)\frac{n_j}{n_l}\right)} \tag{13}$$

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

Figure 8: 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.

exact measure of A_k, or (2) there is no asymmetry across different counties. In this case, the two directions cancel out which lead to unbiased estimates.

To further understand the direction of bias under asymmetry, we plot one numerical example in Figure 8 for eight counties in one DMA. The size of the counties, n_k, are 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 which always bias 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 $A_l = A_k$, $B_{1k} = 1$.

A second source of bias relates to the endogenous ad decision, which is captured by the second term of (11). To simplify B_{2k}, we write out the first order condition of (2) 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

$$\sum_{k \subset l} n_k r_k \frac{\partial A_k}{\partial d_l} = 0 \tag{14}$$

with some simplification shown in A.5, we have

$$B_{2k} = B_2 \mathrm{Cov}\left(d_l, \xi_k\right) \tag{15}$$

where

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

and if r_k is constant in (14), $B_2 = 0$.

A.5 Proof

The bias term is

$$\hat{\alpha}_1 - \alpha_1 \frac{n_l}{n_k} = -\alpha_1 \frac{n_{-k}}{n_k} \cdot \frac{\mathrm{Cov}\left(A_l, A_{-k}\right)}{\mathrm{Var}\left(A_l\right)} + \frac{\mathrm{Cov}\left(A_l, \xi_l\right)}{\mathrm{Var}\left(A_l\right)}$$

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

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

plug in to get (11)

To derive (13), with our assumption, $\mathrm{Cov}\left(A_j, A_k\right) = \rho v$ for $j \neq k$, we have

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

and

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

plug in to get (13).To derive (15), observe that the numerator for B_{2k} is

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

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

B County Pairs in the Border Strategy

The border approach used by Shapiro (2017) differs from others such as Dube et al. (2010), in that the unit of analysis that is "matched" to form a cross-border pair is an aggregation of the counties on a given border as opposed to individual counties. This is more conservative in that all counties on a given side of the border are typically assumed to be treated the same, and Shapiro (2017) effectively has fewer observations than if analysis was conducted at the county level. In our analysis above, we used county-level observations for comparison to Gordon and Hartmann (2013), but kept the pairing at the border-region level (thus the inclusion of border-by-time fixed effects). This difference becomes particularly relevant in our small-border specification (column 6 of Table 2), when we exclude border regions that represent at least 10 percent of the DMA population.

Pairing at the border-region level allows small counties on two sides of a DMA border to be included in the analysis even if a neighboring county in the border region might have been large enough to put the whole border region at a population share greater than 10%. This pairing gives more statistical power to the Border Approach while still limiting the influence of larger counties that may drive advertising decisions.[7]

Table 6 reports fixed effects, instrumental variables and border strategy estimates for county-pair matches. There are 541 counties on the DMA border that form 545 matched

[7] This approach may be questionable in that it acknowledges that an adjacent border-region may be driving advertising incentives but still assumes that local small county factors are irrelevant given their counterpart on the other side. In other words, it assumes that an adjacent county on the opposite side of a border is more similar than an adjacent county on the same side of the border.

pairs, where matching is done with replacement. Counties on average appear just more than once since they match with multiple counties on the opposite side of their DMA border. Columns (1) and (2) document fixed effects and IV estimates on this sample, with both the supply-side and PE instruments. The estimates are similar to those found in Table 2, even though non-border counties are excluded from the analysis. Column (3) applies the Border Approach by including the party-pair-year fixed effect which is a common time-specific effect for the counties on each side of the border and within the focal pair. The estimate here is smaller, as in Table 2. Columns (4) to (6) replicate this analysis for the sample of those county pairs that are each less than 3% of their DMA population. This reduces the number of observations by about half. The estimates are remarkably close across all specifications despite using very different sources of variation for identification. The estimates are, however, all substantially smaller than the fixed effects and IV specifications that were not restricted to these small border regions. This indicates clearly different inference in these regions and is once again suggestive that narrowing the focus of counties identifies effects that may not be generalizable to non-border counties and larger border counties that drive advertising decisions.

Table 6: Comparison of Identification Strategies: Political Advertising

	Dependent variable is Ln(share) - Ln(share0)					
	all pairs			small pairs		
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(Ads)	0.048***	0.052***	0.030**	0.032**	0.034**	0.031*
	(0.011)	(0.014)	(0.012)	(0.013)	(0.016)	(0.018)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Party-county-pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Party-year FE	Yes	Yes	Yes	Yes	Yes	Yes
Lag price, PE IV	No	Yes	No	No	Yes	No
Party-pair-year FE	No	No	Yes	No	No	Yes
Observations	4360	4360	4360	2136	2136	2136
R^2	0.972	0.972	0.991	0.967	0.967	0.989

Notes: Columns 1–3 use 545 cross-border county-pairs, involving 541 unique counties. Columns 4–6 restrict to small county-pairs, defined as county pairs with county-to-DMA population ratio smaller than 3% for both counties in the pair and in both years. Ads is 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), snow(2004). Standard errors in parenthesis are clustered at party-DMA and party-border simultaneously.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$