# Spillover Presidential Ads and Campaign Contributions in a Polarized System

Word Count: 8243

## I. Introduction

The link between campaign advertisements and individual campaign contributions is not well understood, and challenging to study because political parties choose ad content, timing and location for only partly observed reasons. Urban and Niebler, "Dollars on the Sidewalk: Should US Presidential Candidates Advertise in Uncontested States?" (*AJPS* 2014) examine how *unintentional* campaign ads—television ads targeted at competitive states, which also reach viewers in non-competitive states due to media market spillover—affect campaign contributions. They study the 2008 Presidential campaign, and hypothesize that these spillover ads prompt higher contributions in uncontested states (where no one buys campaign ads). Their research design is clever: They determine which zipcodes (their unit of analysis) in uncontested states receive spillover ads. These zipcodes are "treated"; the other zipcodes in noncompetitive states which do not receive spillover ads are the controls. They combine Wisconsin Advertising Project data (WiscAds) on TV advertisements and Federal Election Commission (FEC) data on individual campaign contributions. They exploit variation in advertisements at the media-market level and in contributions at the zipcode level, and report evidence for higher contributions in treated zipcodes. They use propensity score matching (PSM) to match treated and control zipcodes.

In this paper, we first seek to replicate the Urban-Niebler results and then extend their project using a different research design.[1] Urban and Niebler sought to use state fixed effects when

[1] Researchers are paying greater attention to whether published results can be replicated (e.g., King 1995; Imai 2005; Lupia and Elman 2014; Franco, Malhotra, and Simonovits 2015). We are engaged in a broad project using papers drawn principally from the *American Journal of Political Science*, in which we are studying the differences between balancing approaches, applied to real-world datasets. As part of that project, we sought to replicate Urban and Niebler. We are grateful to the AJPS for their replication policy, covering both data and code, which made possible our overall

estimating propensity scores (a sensible idea). However, they did so using conditional logit, likely without understanding the strong assumptions underlying conditional logit.[2] We explain why conditional logit cannot be used to estimate propensity scores, instead use ordinary logit with state dummies, and reassess their results, using an array of balancing methods, including PSM. With balancing methods alone, we find results similar to theirs -- a positive, often statistically significant coefficient on a treatment dummy, but large heterogeneity in treatment effects between states. However, if we combine balancing with regression applied to the balanced sample, these results weaken and become statistically insignificant.

We then extend their analysis in two ways. First, in the data that many zip codes have zero contributions, and when contributions are made, the amounts are highly skewed. See Figure 1 and summary statistics table in Appendix. To address the frequent zeros and the skewness, we implement a zero-inflated model (an approach to campaign contributions also used in Aldrich et al., 2017 and Barber et al., 2017), which models contributions as involving two separate but linked decisions: whether to contribute, and how much to contribute (conditional on contributing something). We find this to be an intuitively appealing model for studying campaign contributions. With this approach, we find evidence that spillover ads increase the amount contributed, but not the likelihood of a contribution.

We also use their data to address partisan response to campaign ads and contributions. We study the differential impact of Republican and Democratic ads on Republican and Democratic campaign contributions. We are not aware of prior research on this issue. We find strong evidence

project and this replication; and to Urban and Niebler for answering our questions about their paper. With only their data, but not their code, we would have struggled to understand how their results were achieved.

2 Their results technically replicate: using their data and Stata code, we obtain the same results.

that ads by one party prime both same party and opposite party contributors (known as "negative partisanship"). We also fine evidence for a stronger Republican response to Democratic ads than vice-versa.

The paper proceeds as follows. We first discuss the Urban and Niebler research design and their clever identification strategy. Second, we discuss where they went wrong in implementing PSM, and correct their error. We then provide evidence on the effects of spillover ads using both balancing methods alone, and combining balancing with regression. Third, we implement a zero-inflated negative binomial (ZINB) model, and study separately the effect of spillover ads on the likelihood of contribution and the amount contributed. Fourth, we study the separate effects of Democratic and Republican ads on own-party and opposite-party contributions.

## II. The Urban-Niebler Research Design; Original and Corrected

### A. Urban-Niebler Design and Problems with Conditional Logit

Urban and Niebler study whether spillover TV ads in presidential campaigns prompt campaign contributions in non-competitive states. They study the 2008 Presidential election campaign. Among the 51 US states (including the District of Columbia), 15 were contested and received the vast bulk of TV ads. The Urban-Niebler research design seeks to compare treated (with spillover) and control (no spillover) zipcodes within the same state. Among the 36 uncontested states, 24 have partial spillover; the others do not. For example, Alaska and Hawaii are not usable for their study because they have no spillover zipcodes; DC is not usable because it has 100% spillover; but California is usable because has some spillover along the Nevada border. States with zero or 100% spillover should drop out of their analysis. However, their estimation strategy resulted in estimating effects for the 12 states with zero or complete spillover, and comparing treated zipcodes to "matched" zipcodes that have very different propensities to be treated.

The authors use multiple variants of propensity-score matching (PSM), but all use the same propensity score estimation model.[3] They estimate propensity scores using "logit estimation with state fixed effects," using as covariates: median household income, Percent Hispanic, Percent African American, Percent of College Graduates, and population density.[4] The treatment is receipt in a zipcode of 1,000 or more spillover ads during the 2008 campaign. The authors use state fixed effects (FE) as covariates in estimating propensities, and run a single propensity score estimation model across all states. They use the Stata command "xtlogit, fe".

For linear regression, a standard result in undergraduate econometrics is that linear regression with unit dummies (here, the units are states) (a "state dummies estimator") produces the same results as a state fixed effects (FE) estimator, which first removes unit means from all variables.[5] Is this also true for logit with state fixed effects, as implemented in standard statistical packages, as one might hope?[6] No, unfortunately. What Stata misleading calls logit with unit fixed effects is actually conditional logit (the xtlogit, fe command calls the clogit command). To estimate unconditional fixed effects logit, one must use ordinary logit with unit dummies as regressors. Unconditional fixed effects logit can be problematic when the number of groups is very large and the number of observations per group is small—the oft-noted "incidental parameters problem" (Katz 2001; Croupe 2005; Woolridge 2015; Beck 2018). But this is not a concern for the Urban-Niebler research design, in which the number of groups is fixed (24 states), and there are many observations per state. When incidental parameters are not a concern, the state dummies

3 The authors also consider genetic matching and coarsened exact matching in their appendix as robustness checks.

4 Urban and Niebler at 328 and eqn. (1). Logit is a common specification for estimating propensity scores.

5 See, e.g., Wooldridge (2015), ch. 14. There can be small differences in standard errors between the two approaches, depending on details of how a statistics package estimates standard errors for fixed effects models.

6 We discuss Stata in text, but R has a similar problem. R has a bife (for binary fixed effects) package, which is mimics Stata xtlogit.

estimator is unbiased. Arpino and Mealli (2011) recommend using unconditional fixed-effects logit to estimate propensity scores with grouped data, based on simulation results.

In contrast, because of how conditional logit treats unit effects, it cannot be used to estimate marginal effects (Beck 2018). This means there is no way to extract propensity scores from the conditional logit estimation. Stata uses a fudge, which does not solve this issue. The Stata default for extracting predicted values following clogit is called pc1 (the post-estimation command is simply "predict p"). Stata explains that this will produce the predicted value (here the predicted propensity score) "conditional on one positive outcome within group," with no further explanation. What this means initially puzzled us, but it turns out to mean the predicted probability of treatment if there were exactly one positive outcome (one treated zipcode) in each state. This produces weirdly low propensity scores, with a mean of only 0.00148 for the 24 usable states, far below the true mean of 0.321.[7] See Table 1.

Urban and Niebler, perhaps after trying this default, chose a different clogit option called pu0 (the post-estimation command is "predict, pu0"). Stata explains that this estimates the "probability of a positive outcome assuming the fixed effect is zero." This assumes that state effects can be ignored, which is contrary to the data and to the point of their state FE design. Actual state-level mean propensities vary widely, from 0.014 (Mississippi) to 0.948 (Montana). Using clogit with the pu0 option generates very different but again weird outcomes. One problem is that Urban and Niebler ran clogit estimation using data for all 36 noncompetitive states, including the 12 states with zero or complete spillover. For these states, the propensities are driven to 0 or 1 by the state fixed effects. Indeed, the initial clogit estimation drops these 12 states, and reports results

[7] These 24 states have 16,265 zipcodes. The mean **propensity** for these states is therefore, assuming one treated zipcode per state is then: 24/16,250 = 0.00148! The per-state mean values vary from 1/1,510 = .0007 in California (with 1,510 zipcodes) to 1/25 = .04 in D.C. (with 25 zipcodes). Odd indeed.

only for the remaining 24 states. But these 12 states magically reappear if one runs the predict command (with either the default pc1 or the pu0 option), following the clogit regression. Stata reports propensities for 36 states (20,215 zipcodes), with no warning that it has done so. All reported propensities, in all states, are influenced only by the non-fixed covariates, which are weak predictors of treatment, so most propensities are near 50%.[8]

Table 1 reports summary statistics for the propensity score, computed various ways: Using the pu0 option with 36 states; the pu0 option with 24 states (manually dropping the other 12 states); the pc1 option with 36 and 24 states, the state dummies estimator (ordinary logit with individual state dummies) for the 24 states with partial spillover, and ordinary logit combined with trimming to common support (dropping treated units that are outside the support of the control units and vice versa).

Manifestly, estimates using propensities for all 36 states, using clogit with the pu0 option, is not what one should do. Yet this are the only propensity scores that Urban and Niebler used. In their Appendix, they ran multiple variations for how to match on the propensity score, but never varied the model that they used to compute propensities. Conditional logit led them astray.

A second problem with their research design: One should cluster standard errors on state, which they did not do for their main results. We find that standard errors clustered on state are substantially larger than either ordinary or heteroskedasticity robust standard errors.

[8] In R, the predict command for the bife package drops the 12 states with zero or complete overlap. This is better than keeping them, but does not address the inability of conditional logit to estimate marginal effects.

**Table 1. Summary statistics with different propensity score estimation approaches**

Table shows summary statistics for propensity to be treated (have at least 1,000 spillover ads) for the Urban-Niebler data: (1) using conditional logit in Stata (either clogit or xtlogit, fe) with data for all 36 noncompetitive states, followed by "predict p"; (2) same, but manually dropping 12 states with zero or complete spillover; (3) using 36 states and "predict p, pu0," (4) same as (3) but manually dropping 12 states; (5) using ordinary logit with 24 states and dummy variables for each state; and (6) same as (5), but trimming propensities to common support within each state.

| Model | No. of states | No. of zipcodes | method | predict option | Mean | Min | Max |
|---|---|---|---|---|---|---|---|
| (1) | 36 | 20,215 | clogit | pu0 | .5821 | .3713 | .9150 |
| (2) | 24 | 16,265 | clogit | pu0 | .5823 | .3759 | .9150 |
| (3) | 36 | 20,215 | clogit | default (pc1) | .0019 | .0002 | .0620 |
| (4) | 24 | 16,265 | clogit | default (pc1) | .0015 | .0002 | .0431 |
| (5) | 24 | 16,265 | logit, state dummies | default | .3215 | .0086 | .9723 |
| (6) | 24 | 13,869 | logit, state dummies, trim to common support | default | .3264 | .0099 | .9476 |

## III. Trying Again with Better Methods

We next considered how one might study the Urban-Niebler research question, using their data. We study the 24 noncompetitive states with partial spillover, using three main approaches. First, we use PSM (with corrected propensity scores, using either the state dummies estimator or, perhaps closer to the spirit of their approach, requiring in-state matches). We also use a variety of other balancing methods balancing methods that we consider attractive based on the matching literature, plus their "robustness check" method, from their Appendix, of coarsened exact matching (CEM). We use their covariates plus zipcode population.

We were concerned, however, that these results might be sensitive to specification. To assess sensitivity to specification, we adopted three approaches. The first is to use a more normally distributed dependent variable, as a robustness check. Contributions (in $'000's) has both many zero (or small) values and a large positive skew. We therefore use ln(contributions/capita) as an alternative dependent variable, adding a small amount ($200) to each observation to avoid dropping zero-contribution observations.

The second is to combine balancing with regression. The third is to use a zero-inflated negative binomial model, which directly addresses the zero-or-small values, and posits a two stage contribution decision: first, whether to contribute and then, how much to contribute.

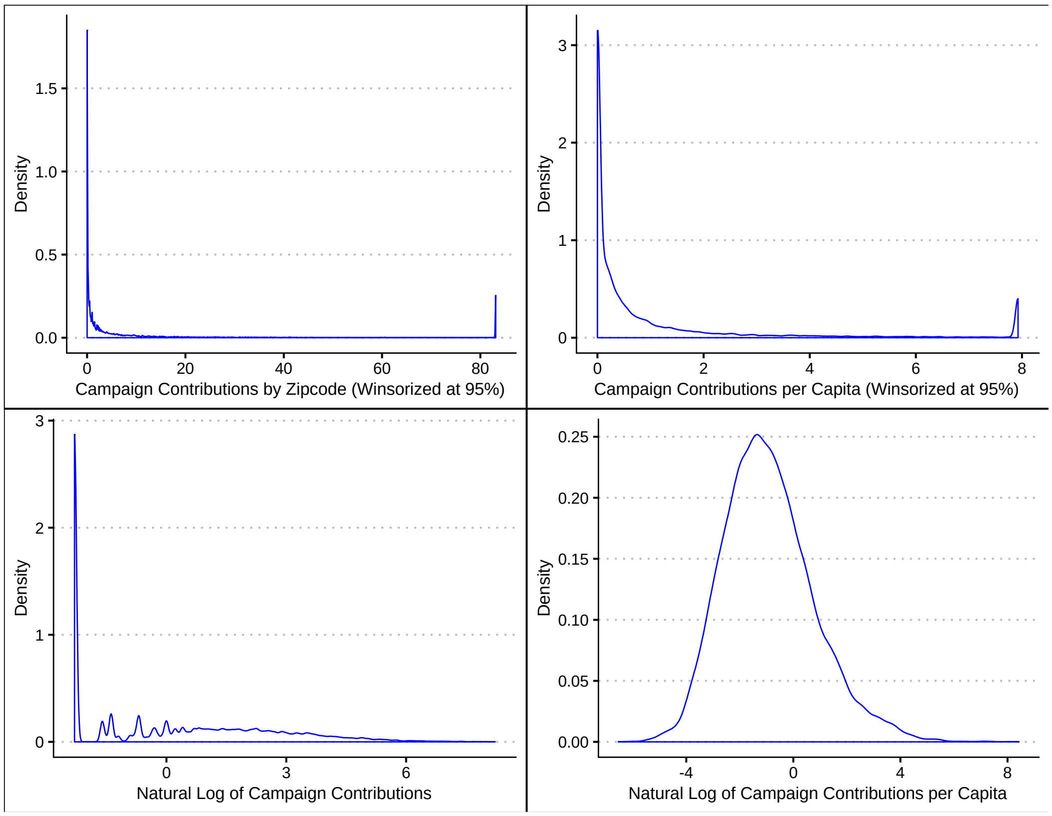
Urban and Niebler chose to consider a zipcode as treated if it received 1,000 or more spillover ads, and control otherwise. In the Appendix, we assess the sensitivity of the results presented below to varying the minimum number of ads needed for a zipcode to be considered to be treated. Coefficient estimates vary as we vary the threshold, but the qualitative results are not sensitive to variations on the threshold.

**A. Addressing Zero Values and Skewness in Contributions**

Campaign contributions are highly skewed (Bonica, 2014). As Urban and Niebler note, this skew remains at the zipcode level. Figure 1 provides a histogram of the frequency of contribution levels (in $'000s); we convert small positive amounts, below the $200 federal reporting threshold, to $200 (this affects only 17 zipcodes). The top left panel shows contributions in dollars, winsorized at the 95% level, the winsorized contributions are shown in the bar at the far right. The top right figure shows contributions per capita, which have a similar skew. The bottom left figure shows *ln*(contributions). Before taking logs, we add $100 to the 5,904 zero contributions to avoid losing these observations. Taking logs reduces the right skew, but the distribution remains highly skewed. Finally, the bottom right figure (1-4) shows *ln*(contributions per capita), using the same rules for constructing logged values. This measure is roughly normally distributed; we use it as an alternative dependent variable.

**Figure 1. Histograms using Different Dependent Variables**

Histogram of contribution amounts (in $'000s) for 16,265 zipcodes in 24 noncompetitive states with partial spillover of campaign ads. Top panels: contributions are winsorized at 95%. Bottom panels: Before taking logs we add $100 to so as not to drop the 5,904 zipcodes with zero contributions and winsorize positive but small contributions in 17 zipcodes to $200 (the legal threshold for reporting contributions).




## B. One-Step Balancing Methods

In Table 2, we apply PSM and a variety of other balancing methods. We summarize here the perhaps less familiar methods. For all approaches, we estimate ATT (average treatment effect on the treated). For PSM, we use 1:1 matching with replacement. Nnmatch is a routine for matching on covariates, with bias correction (Abadie and Imbens, 2011), and is implemented in Stata. We use 1:1 matching with replacement and bias correction. Entropy balancing (eBalance) (Hainmueller, 2012) provides weights that ensure exact balance on covariate means between treated and control groups, and is implemented in Stata and R.[9] Optimal matching Hansen (2004) creates groups of treated and control units with similar covariate values, and is implemented in R. Covariate balance propensity scores (CBPS) (Imai and Ratkovic, 2014) is similar to eBalance in that it provides balance on covariate means, but also estimates propensity scores, which can be used in a PSM approach. CEM (Iacus, Kong, and Porro, 2011) imposes exact matching on a limited set of covariates, with continuous covariates divided into bins to make exact matching feasible.[10]

We use the same balancing variables as Urban-Niebler, plus population (they use only population density).[11] We use two PSM approaches. In the first, we estimate a single propensity score equation across all states, using logit with state dummies:

$$\mathbf{1})\ p(D=1)_{i,s} = \alpha + \delta_s + \boldsymbol{\lambda X_{i,s}} + \epsilon_{i,s}$$

Here *i* indexes zipcodes, s indexes states, **X** is a vector of covariates (the Urban-Niebler covariates plus population),[12] $\delta_s$ are state dummies, and $\varepsilon_{i,s}$ is the regression error. In this approach, we allow across-state matches. In the second approach, we estimate propensity scores and conduct matching separately within each state. For methods which use the propensity score (PSM, IPW, and CBPS + PSM), we trim treated units to fall within the support of the controls (the propensity score range for control zipcodes). Trimming does not affect the number of treated zipcodes when we match across all states, but with state-specific matching reduces the number of treated zipcodes from

[9] eBalance can also be set to provide balance on higher moments. We set it to balance only on means.

[10] CEM is intended to be used with regression on the matched data (done below, but not done in this section).

[11] The Urban-Niebler choice to use population density as a covariate, but not population, struck us as unusual. Population is a stronger predictor of contributions than population density, and thus a better choice if one wants to use only one of the two variables, but we saw no strong reason not to use both.

[12] For references, the covariates are Median Household Income, Percent Hispanic, Percent African American, Percent of College Graduates, and population density

5,230 to 4,797. For the other matching approaches (nnmatch and optmatch), we use in-state matches. For weighting approaches (IPW and eBalance), weights are measured across all states.[13]

Table 2, Panel A shows results with raw contributions as the dependent variable. For comparison, Urban and Niebler report coefficient estimates for their main PSM variations of 5.3 to 7.2, with a similar, but somewhat broader coefficient range for the additional PSM variations in their Appendix. The exception is CEM, which drops many observations, for which they report a coefficient of 0.29 (insignificant). We put the CEM results aside, but note that in work in progress we find that CEM can produce results that are far away from other balancing approaches. Our coefficient estimates range from 4.3 to 10.25, most are significant.

Across approaches, we find reasonably consistent coefficients, ranging from 4.3 for eBalance to 10.4 for nnmatch. At the same time, the eBalance results are noticeably weaker than with the other approaches. Standard errors (s.e.'s) vary substantially; a principal reason is that s.e.'s are substantially larger if one either conducts within-state matching, or uses standard errors clustered on state; we would therefore distrust the much smaller s.e.'s for PSM in col. (1). But all estimates are statistically significant (5% level or better) or marginally significant (10% level).

In Panel B, we switch to *ln*(contributions/capita) as the dependent variable. All coefficients are positive, with a range from .063 to .249, implying a mean percentage increase in contributions of 6.5-28.3%.[14] However, the results weaken somewhat, and lose significance with the two exact-balance approaches (eBalance and CBPS). Still, overall, using balancing methods alone, *without regression*, we find reasonable support for the Urban-Niebler original hypothesis.

[13] In unreported robustness checks, we (i) limited the maximum times a single control unit could be matched to different treated units to 10 (keeping the 10 best matches), and limited the weights on control units for IPW and eBalance to 10. Results were not meaningfully different from those we report.

[14] Using the usual formula for converting coefficients from a log-linear specification to percentages: $(e^{0.063} - 1) = 0.065$; $(e^{0.249} - 1) = 0.283$.

**Table 2: Results for Contributions with One-Step Balancing Methods**

Estimates of average treatment effect on the treated (ATT), for contributions in spillover zipcodes using indicated balancing approaches. **Panel A**. Dependent variable is contributions in $'000's). **Panel B**: Dependent variable is ln(contributions/capita). We add $100 to zero contributions to avoid dropping zero-contribution zipcodes. **Both panels**: Contributions of less than $200 (minimum reporting threshold) in 17 zipcodes are set to $200. All balancing approaches use the following covariates: median household income, percent Black, Hispanic and college graduate; population, and population density. Propensity score matching (PSM) is 1:1 with replacement. Nnmatch is 1:1 with exact match on state and bias adjustment. Inverse propensity weights (IPW) and entropy balancing (eBalance) use ATT weights. Optmatch is Ben Hansen's optimal matching, with exact match on state. CBPS is covariate balance propensity score. CEM uses default cutpoints. Standard errors are clustered on state when available. * indicates significance at 5%.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Balancing Method | PSM | PSM | nnmatch | IPW | eBalance | optmatch | CBPS + PSM | CBPS + PSM | CEM |
| In-State Matching | No | Yes | Yes | No | No | Yes | No | Yes | Yes |
| Trim to common support | No drops | yes | n.a. | No drops | n.a. | No | No | No | no |
| **Panel A. Dep. Variable: Contributions ($'000's)** | | | | | | | | | |
| Treatment | 9.59* | 9.39* | 9.54* | 10.25* | 4.30 | 6.06* | 7.96* | 8.33* | 0.23 |
| | (1.96) | (4.03) | (3.58) | (3.91) | (2.57) | (1.59) | (3.08) | (4.48) | (0.29) |
| **Panel B. Dep. Variable: ln(contributions/capita)** | | | | | | | | | |
| Treatment | 0.209* | 0.190* | 0.249* | 0.212* | 0.063 | 0.084* | 0.081 | 0.128 | 0.021 |
| | (0.071) | (0.075) | (0.058) | (0.049) | (0.046) | (0.040) | (0.080) | (0.083) | (0.030) |
| No. of zipcodes | 7,309 | 6,753 | 8,460 | 16,265 | 16,265 | 5,792 | 7,228 | 6,588 | 7,744 |
| Treated zipcodes | 5,230 | 4,797 | 5,230 | 5,230 | 5,230 | 2,896 | 5,180 | 4,612 | 2,892 |

### C. Two-Stage Estimation: Balancing Plus Regression

In many situations, it can be appropriate to combine balancing with regression using the balanced data. IPW is formally "doubly robust" – it should produce unbiased estimates if either the propensity score estimation model or the regression model is correctly specified. So is eBalance (Zhao and Percival 2017).[15] The other approaches are not formally doubly robust, but matching plus regression on the matched sample, is an often recommended practice (Ho et al. 2007; Stuart, 2010). As Kang and Schaffer (2007) stress, if important covariates are not observed, both steps will often be misspecified. In that common situation, a doubly-robust or matching-plus-regression

15 CBPS is doubly robust if used for reweighting (Imai and Ratkovic 2014). Here, we use CPBS for matching

approach may not help much. Still, if one avoids extreme weights on particular observations, there is little reason to prefer a balancing-alone approach over balancing plus regression. We use eqn. (2) in estimating balancing-plus-regression results using (changing the dependent variable to *ln*(Cont/capita), and/or adding IPW or eBalance weights, as appropriate). We use the same covariates in balancing and regression.

$$\mathbf{2})\ Cont_{\mathrm{i},s} = \alpha + \delta_s + \beta D_{i,s} + \boldsymbol{\lambda X_{i,s}} + \epsilon_{i,s}$$

We present results in Table 3. The principal changes relative to Table 2, are that we: (i) report simple regression without balancing (col. (1)); and (ii) report results for PSM and CBPS + PSM only with in-state matching, which we view as a stronger approach because this ensures balance on unobserved state-specific factors. Similar to Table 2, results with Contributions as the dependent variable are in Panel A, and with ln(contributions/capita) as the dependent variable are in Panel B. Coefficients on covariates are suppressed. The results with balancing-plus-regression are much weaker than with balancing alone, as to both coefficients and statistical significance, although all coefficients remain positive. Only with nnmatch is there evidence for a treatment effect (marginally significant in Panel A; significant in Panel B). The large differences between balancing alone, and balancing-plus-regression imply that balancing alone is misspecified, and is not a reliable basis for inference. With the two-stage, balancing-plus-regression approach, we find limited evidence for a treatment effect.

**Table 3: Results with Balancing plus Regression**

Estimates of average treatment effect on the treated (ATT), for contributions in spillover zipcodes using indicated balancing-plus regression approaches. Dependent variables, balancing methods, and trimming are same as in Table 2, except (i) we run nnmatch without bias-adjustment; and (ii) for PSM, and CBPS + PSM, we report results only with in-state matching. Regressions use state FE and the same covariates used for balancing: median household income, percent Black, Hispanic, and college graduate; population, and population density; coefficients are suppressed. Standard errors are clustered on state* indicates significance at 5%.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| Balancing Method | None | PSM | nnmatch | IPW | eBalance | optmatch | CBPS + PSM | CEM |
| In-state matching | -- | Yes | Yes | No | No | Yes | Yes | Yes |
| **Panel A. Dep. variable: Contributions ($'000's)** | | | | | | | | |
| **Treatment dummy** | 1.97 | 2.33 | 4.74 | 1.05 | 3.69 | 1.43 | 1.07 | 0.22 |
| | (3.89) | (1.74) | (2.27) | (3.00) | (4.01) | (1.53) | (1.21) | (0.24) |
| **Panel B. Dep. Variable: ln(contributions/capita)** | | | | | | | | |
| **Treatment dummy** | 0.028 | 0.085 | 0.125* | 0.054 | 0.063 | 0.062 | 0.058 | 0.014 |
| | (0.044) | (0.079) | (0.052) | (0.031) | (0.033) | (0.038) | (0.023) | (0.049) |
| Number of zipcodes | 16,265 | 6,753 | 8,460 | 16,265 | 16,265 | 5,792 | 6,588 | 7,744 |
| Number of treated zipcodes | 5,230 | 4,797 | 5,230 | 5,230 | 5,230 | 2,896 | 4,612 | 2,892 |

### D. Zero-Inflated Model

We can address the large number of zero contribution values (5,904 zipcodes out of 16,265 reported zero contribution dollars), and the strong right skew in contributions, using a zero-inflated negative binomial (ZINB) model. This model assumes that contributors make two separate but related decisions: whether a contribution is made at all (a binary choice, usually estimated with logit) and how much to contribute, estimated using a negative binomial count model. The idea is that the factors predicting donation are likely to depend on exposure to ads and the specifics of the local political environment, while the decision on how much to give is will often relate to contributor-specific factors, such as income. The two-stage process seems to us to be a reasonable way to model campaign contributions, and has been used before to model contributions (Aldrich et al. 2018; Barber et al. 2018). However, there is no useful test for whether the ZINB model is appropriate.

A zero-inflated model can permit more intuitive modeling of the zeros than conventional count models and can also account for overdispersion (e.g., Zeileis et al., 2011; Ward and Ahlquist, 2018). The model requires the dependent variable to take integer values (so that a count model is feasible in the second-stage); this forces us to use contributions (in dollars rather than $'000s) as the sole dependent variable. The usual count model choices are Poisson (if appropriate, which is rare in practice) and negative binomial. We present negative binomial results in Table 4, after confirming that the Poisson model is not suitable due to strong overdispersion.[16]

$$\mathbf{3)}\ \ Cont_{i,s} = \begin{cases} \alpha + \delta_s + \beta D_{i,s} + \boldsymbol{\lambda X_{i,s}} + \boldsymbol{\eta \delta_s X_{i,s}} + \epsilon_{i,s} & if\ Cont_{i,s} = 0 \\ \alpha + \delta_s + \beta D_{i,s} + \boldsymbol{\lambda X_{i,s}} + \boldsymbol{\eta \delta_s X_{i,s}} + \epsilon_{i,s} & if\ Cont_{i,s} \geq 0 \end{cases}$$

We model both equations simultaneously, using a logit model for the top equation and a negative binomial model for the bottom equation.

In Table 4, column (1) shows results from the ZINB model without balancing; remaining columns use the same balancing approaches as in Table 3. There is no evidence that spillover contributions affect the likelihood of contributing; indeed most point estimates are negative, although insignificant. However, across balancing approaches (still putting CEM aside), there is reasonably consistent evidence that spillover ads lead to larger contributions. The overall coefficient estimate using PSM implies that treatment induces an average of $14,200 in additional campaign contributions;[17] whereas using optmatch, the ATT is $6,247. All of the other estimates are between these two values.

The dispersion across estimates using different balancing methods is reasonably large, though consistent in significance and direction – for example, the PSM and optmatch estimates

[16] The likelihood-ratio test for alpha for each model has $\chi^2 = 4.4 \times 10^8$ or more, indicating severe over-dispersion. The negative binomial model begins with the standard Poisson model, but accounts for over dispersion by adding a random error term to the outcome distribution (Greene, 2003; Ward and Ahlquist, 2018).

[17] We used the predicted total contributions from our model and subtracted the treated from the control.

are more than three standard deviations apart (using optmatch standard errors). For us, the differences reinforce the value of not relying on a single balancing method until more is known about which approach is likely to work better in which situations.

**Table 4: Zero-Inflated Negative Binomial Results**

Zero-inflated negative binomial models of contributions in dollars. Top set of results are from the binary model of whether to contribute; bottom set are from the count model of how much to contribute. All regressions include state fixed effects and standard errors clustered on state. Balancing approaches and covariates are same as in Table 3. Standard errors are clustered on state. * indicates significance at 5%.

| **Dependent Variable** | **Contributions ($)** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Regression | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Balancing Strategy | None | PSM | nnmatch | IPW | eBalance | optmatch | CBPS+PSM | CEM |
| State-specific matching | -- | Yes | Yes | No | No[18] | Yes | Yes | Yes |
| **Binary Model (Logit)** | | | | | | | | |
| Treatment | -0.024 | 0.029 | -0.041 | -0.089 | -0.89 | -0.030 | -0.096 | -0.042 |
| | (0.089) | (0.165) | (0.057) | (0.147) | (0.155) | (0.074) | (0.121) | (0.087) |
| **Count Model (Negative Binomial)** | | | | | | | | |
| Treatment | 0.128 | 0.276* | 0.274* | 0.148* | 0.159* | 0.112* | 0.192* | -0.001 |
| | (0.077) | (0.084) | (0.056) | (0.056) | (0.053) | (0.055) | (0.075) | (0.087) |
| Number of zipcodes | 16,265 | 6,753 | 8,460 | 16,265 | 16,265 | 5,792 | 6,588 | 7,744 |
| Number of treated zipcodes | 5,230 | 4,797 | 5,230 | 5,230 | 5,230 | 2,896 | 4,612 | 2,892 |

**F. Varying the Threshold Number of Spillover Ads**

We have thus far used the Urban-Niebler choice of 1,000 spillover ads s the threshold number for a zipcode to be considered to be treated. We next assess whether our findings are sensitive to that choice. Figure 2 shows the effects of altering the threshold, from 0 to 5,000 in increments of 100, for the PSM-plus-regression results we report in Table 4, col. 2, for the count stage of the ZINB model.

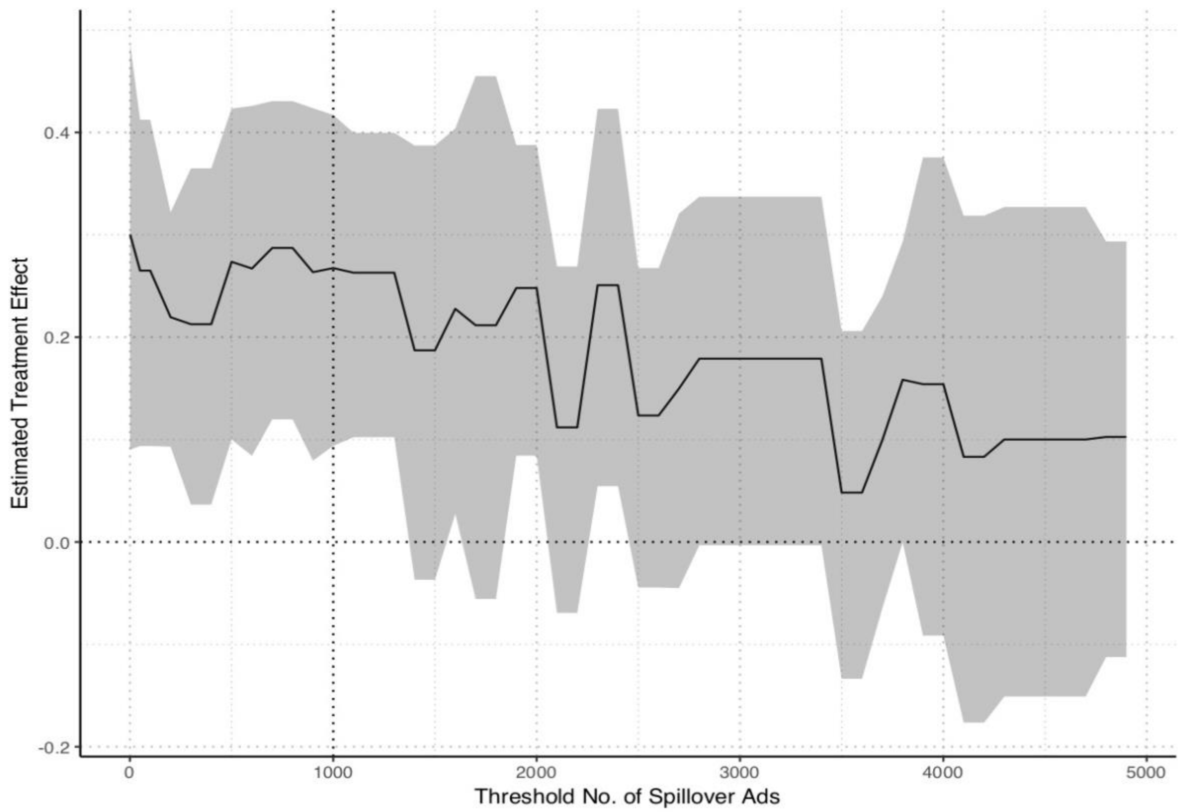
In Figure 2, coefficient estimates are always positive and, for thresholds up to 1,300, always significant. As one moves to higher thresholds, the estimates gradually decline and

[18] For IPW and eBalance, we cannot match on state. However, we include state dummies in estimating weghts and in the second-stage regression, so we are still addressing state level heterogeneity in both stages.

standard errors increase. The loss in precision is expected, because as the threshold increases, there will be fewer and fewer treated observations.[19] Overall, the results in Table 4 are reasonably robust to choosing a different threshold, including a zero or low threshold.

**Figure 2. Results with PSM Plus ZINB Regression, Using Different Treatment Thresholds**

Estimated treatment effects for amount contributed, conditional on contribution being made, from ZINB model, similar to Table 4, but varying the minimum number of spillover ads needed for a zipcode to be considered treated rather than control. Treatment threshold is varied in increments of 100. Dotted vertical line shows 1,000 ads (threshold used in Tables 2-4). Shaded area shows 95% confidence interval.




## IV. Partisan Treatments, Partisan Outcomes, and Negative Partisanship

Urban and Niebler have data on ads and contributions by political party, but did not exploit this data in their paper. We use this data to study the extent to which one party's ads generate

[19] See Appendix Table A4 and Figure A3 for details on the distribution of number of spillover ads by zipcode.

contributions for that party, versus the other party. This exploration is related to the motivations for political ads and campaign contributions generally, as well as the literature on polarization.

Political ads are intended to generate voting support, often among both committed and uncommitted voters. A secondary goal can be to induce stronger support, including contributions, among already committed voters. We would expect ads to generate contributions if either the individual hopes to profit from the donation through special privileges from the winner (Ansolabehere, de Figueiredo, and Snyder 2003) or the individual believes the candidate matches their prior political views (or that the opposition represents a significant a threat to those positions). For spillover ads, the first motive (personal gain from contributing) is unlikely. For the second, partisan ads can reinforce partisan behavior (including contributions), by reminding sympathetic viewers of the match between their views and those of the candidate or the party more generally (Huddy, Mason and Aarøe, 2015).

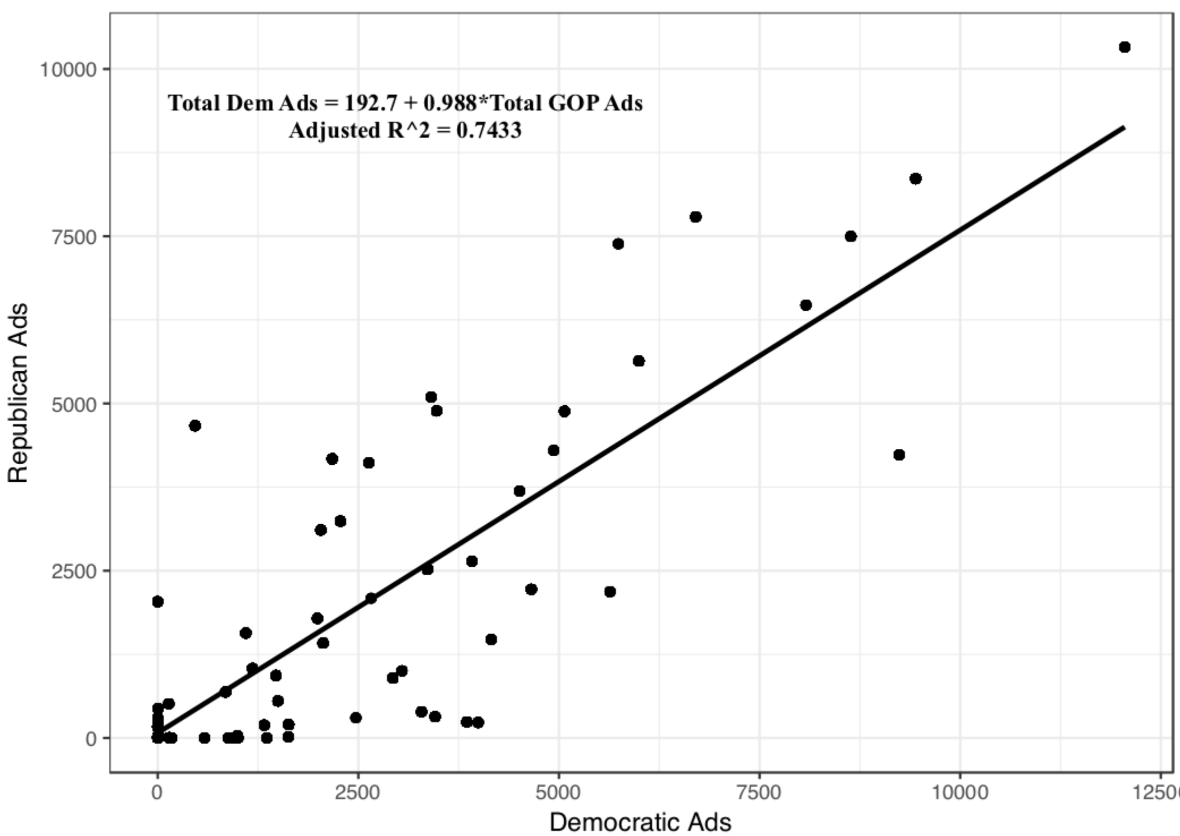
However, ads can also induce opposite-party contributions, by reminding unsympathetic viewers of the mismatch between their views and those of the candidate or party. Campaign ads can intensify partisan sentiment among those with strong priors (e.g., Iyengar, Sood, and Lelkes 2010; Druckman, Peterson, and Slothuus 2013; Druckman, Levendusky and McLain 2018).
The Urban-Niebler data allows us to explore whether Democratic ads induce net (Democratic minus Republican) contributions, and vice-versa, and in which zipcodes. In particular, if the effect of ads in inducing opposite party contributions is strong enough, a Democratic ad in a Republican zipcode could induce net Republican contributions, and vice-versa.

The Urban-Niebler data covers the 2008 election. We should also expect asymmetric polarization, with Republicans having been more polarized than Democrats at that time (McCarty, Poole, and Rosenthal 2006; Williamson, Skocpol, and Coggin, 2011; Mann and Ornstein 2016).

Asymmetric and negative partisanship implies that Democratic ads may induce a Republican response that is stronger than the Democratic response to Republican ads.

Figure 3 shows a scatterplot of the number of Democratic and Republican spillover ads, together with a best-fit regression line. The strong correlation is apparent, although with large variation in number of ads, and some variation in the relative number of Democratic vs. Republican ads. Our ability to measure the separate effects of Democratic and Republican ads depends on the zipcode-level departures from the regression line.

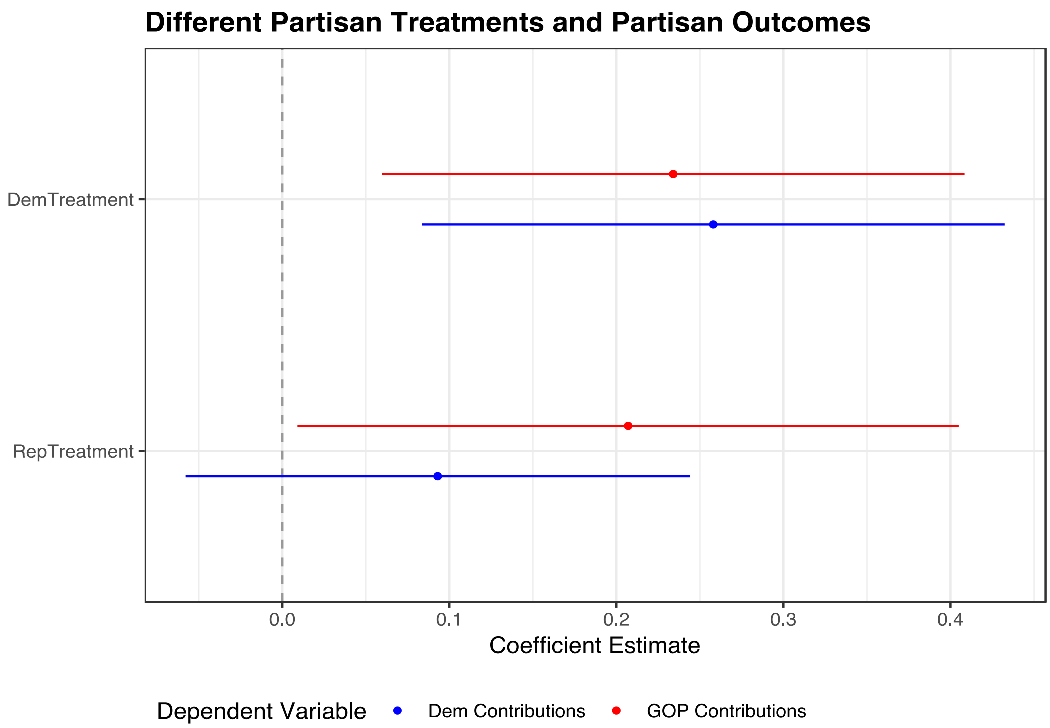
**Figure 3: Comparison of Democratic and Republican Ads**




In Figure 4, we estimate the separate effects of Democratic and Republican ads on Democratic and Republican contributions, using a ZINB model, with the within-state PSM as the balancing approach (similar to results from Table 4, column 2, but with different treatments and dependent variables). We present only the effect of spillover ads on amount contributed; effects on likelihood of contributing are always insignificant (see regression results in Appendix). We consider a zipcode to be treated with Democratic ads if it receives 500+ Democratic spillover ads and Republican ads if it receives 500+ Republican spillover ads.

**Figure 4: Coefficient Estimates for Contribution Amount, for ZINB Model of Democratic and Republican Spillover Ads on Democratic and Republican Contributions.**

Zero-inflated negative binomial models (ZINB) of contributions in dollars. Graph shows coefficient estimates and 95% confidence intervals, with standard errors clustered on state, for amount contributed (second stage of ZINB model) separately for Democratic and Republican spillover ads, and separately for Democratic and Republican contributions. ZINB model is otherwise same as in Table 4. Democratic treatment is 500+ Democratic spillover ads. Republican treatment is 500+ Republican spillover ads.




Converting regression coefficients to marginal effects), we find that, on average, Republicans gain $4,597 over Democrats from 500+ spillover ads (RepTreatment), while Democrats only $1,873 over Republicans. The confidence intervals overlap, but there is still some evidence that

Republicans gain more from a marginal spillover ad than Democrats, since Democratic ads induce more Republican contributions than Republican ads induce Democratic contributions.[20]

We next explore an implication from the similar Democratic and Republican responses to Democratic ads, and the greater Republican than Democratic response to Republican ads. Consider *net* own party contributions as an outcome of interest (Democratic minus Republican contributions for Democratic ads; Republican minus Democratic contributions for Republican ads). Putting aside the effects of covariates and state FE, in a close district, with 50% Republican voters, Republican ads will substantially help Republicans, while Democratic ads will barely help Democrats. In a Republican district with significantly more than 50% Republican voters, Democratic ads will, on balance, generate net Republican contributions – hardly the outcome that the Democrats would hope for. In contrast, in a Democratic district, Republican ads will still generate net Republican contributions, unless the district is strongly Democratic.

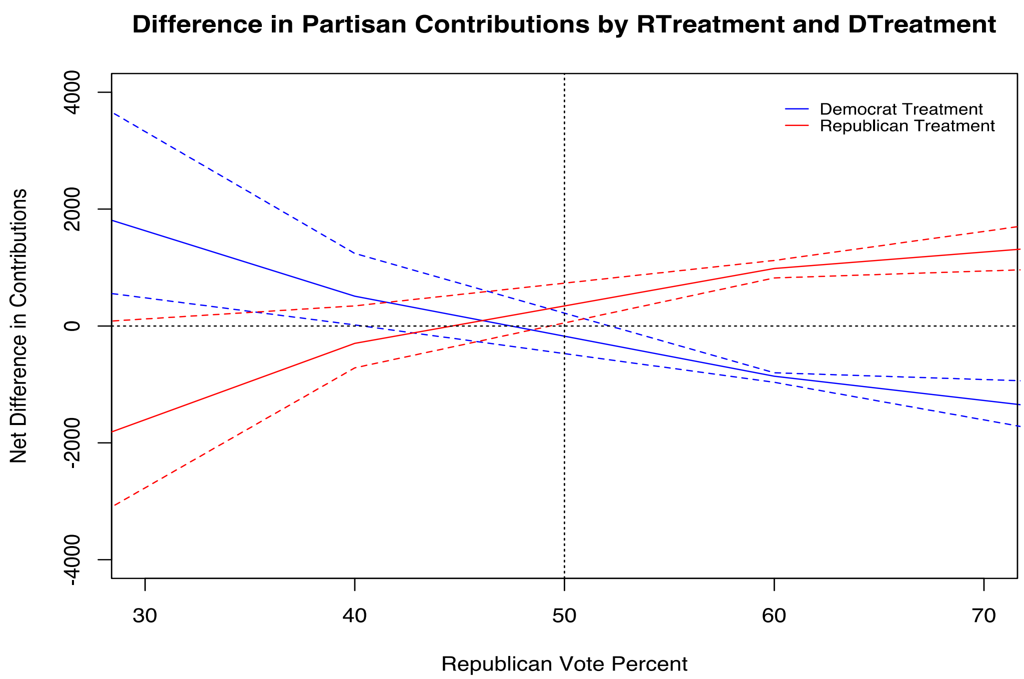
Figure 5 illustrates this by plotting the marginal effects for an example state, Illinois, for net Democratic contributions from Democratic treatment (500+ Democratic spillover ads), and net Republican contributions from Republican treatment (500+ Republican spillover ads, holding all other covariates at their Illinois medians, as a function of Republican vote share. As expected, the Democratic line slopes down. The point estimate turns negative at 46% Republican vote share. The Republican line slopes up, and turns positive at 44% Republican vote share. As expected, Republicans gain more from spillover ads in heavily Republican areas, and Democrats gain more in heavily Democratic areas. The difference in effects between the Democratic treatment on Republican contributions (which, from figure 4 is positive and statistically significant) compared

[20] Due to the high colinearity between Republican and Democratic ads (Figure 3), the difference between the 0.93 coefficient for Republican ads to generate larger Democratic contributions, and the 0.23 coefficient for Democratic ads to generate Republican ads is not statistically significant.

with the Republican treatment on Democrats (which is smaller but still positive, and no longer statistically significant) is reflected in positive net Republican contributions for Republican spillover ads in leaning-Democrat districts, but negative net Democratic contributions in leaning-Republican districts.[21]

**Figure 5: Predicted Net Marginal Effects from Zero-Inflated Negative Binomial Model.**

Blue lines are the marginal effects in Illinois (chosen as an example state because it has the most spillover zipcodes) for the gain to Democrats (Democratic minus Republican contributions) in treated zipcodes, holding covariates at their Illinois medians, as a function of Republican vote share. Red lines are similar but for Republicans: the marginal effects in Illinois for the gain to Republicans (Republican minus Democratic contributions) in treated zipcodes.




By many measures, partisanship has increased in both parties since 2008. Thus, Figures 4 and 5 would surely look different today. Also, campaign ads are intended to attract both votes and contributions, but we study only contributions. Still, two major takeaways from these figures

[21] The specific estimated Republican vote percentage at which each party will obtain positive net contributions varies to some extent by state. The graph for Illinois is only illustrative.

likely remain. The net gain in contributions from campaign ads, to the party buying the ads, are smaller than one might think, because of negative partisanship. And the opportunity to achieve net gains in marginal districts can be asymmetric between the two parties.

## IV. Discussion and Conclusion

Our analysis leads to five main takeaways, two substantive and three methodological.

### A. Effects of Spillover Ads on Campaign Contributions

Our first substantive result involves the effect of spillover campaign ads on contributions. Urban and Niebler studies the effect of spillover ads on campaign contributions at the zipcode level, but with an incorrect approach for estimating propensity scores. We use their data, correct their technical error, and then apply a variety of balancing methods, and use two-stage estimation (balancing plus regression). We find positive but insignificant coefficients for their outcome measure (dollar contributions). We find somewhat stronger, but still mostly insignificant coefficients with ln(contributions) as the outcome measure. Overall, their hypothesis has at most mild support.

However, when we use a zero-inflated negative binomial model, which models separately the decision whether to contribute and the amount contributed we find a positive relationship between spillover ads and amount contributed. In contrast, we find no significant relationship between spillover ads and the likelihood of contributing; indeed, across our array of balancing methods, most point estimates are negative.

### B. Positive versus Negative Partisanship and Asymmetric Polarization

Our second and third substantive results involve how partisanship mediates the response to spillover ads. We find strong evidence for both positive and negative partisanship: Ads by one party induce contributions to both parties. We also find evidence for asymmetric polarization. At least in 2008, Democratic ads induce nearly equal contributions to both parties. In contrast, we find some evidence that Republican ads induce Republican contributions more strongly than Democratic contributions. This implies that Democratic ads will induce net Democratic contributions only in districts which are majority (or close) Democratic already, while Republican ads can induce net Republican contributions in both Republican and marginally Democratic districts. If these results carry over to today, and if they apply to targeted as well as spillover ads, they have important implications for the use of campaign ads. They imply a stronger tradeoff for Democrats than for Republicans when targeting additional ads in marginal districts – which is where most ads are targeted.

We do not study the effect of ads on voting, but similar partisan effects are plausible – ads by one party could induce greater voting turnout by partisans from both parties. If so, this would further complicate party decisions on how to advertise – and where. Further research in this area is needed to estimate the effects of ads on both contributions and voting, and ideally how partisan effects have evolved over time. Indeed, a substantial contribution would be to model this trade-off for inducing campaign contributions with one for inducing votes and determine (even theoretically) inflection points where the parties should locate their efforts.

**C. Estimating Propensity Scores with Grouped Data**

We explain why conditional logit (which Stata confusingly implements using both the clogit and xtlogit commands) cannot be used to estimate propensity scores, and how Urban and Niebler went wrong in using xtlogit. This will not surprise those familiar with conditional logit. But in our experience, conditional logit is often misunderstood and misused.

**D. Value of Using Two Stage Methods, Rather than Balancing Alone**

Many papers report regression results without balancing, some use propensity score matching or another balancing approach alone; and some use balancing plus regression.[22] Urban and Niebler apply propensity score matching alone, without a second stage regression. We find large differences between estimates with balancing alone, and estimates that combine balancing with regression. These differences suggest the importance of using two stage balancing approaches. With single-stage estimation, using the Urban-Niebler outcome measure (zipcode level contributions), we found strongly significant positive coefficients across balancing approaches. However, the coefficients shrank, and significance disappeared, with two-stage estimation. This implies that balancing alone was misspecified in some way (we do not seek here to understand how).

Using two stage methods also allowed us to pursue the ZINB approach – which better accounts for the distribution characteristics of the dependent variable (many zero or small values) than OLS alone, or than balancing alone. By using balancing plus a ZINB model, we find strong evidence for a specific treatment effect on amount contributed (given that a contribution is made), and evidence against any effect on likelihood of contributing.

**E. Value of Using Multiple Balancing Approaches**

We apply multiple balancing approaches, and find substantial differences across those approaches. The implication is that researchers should have limited faith in results with a single balancing method. It is not technically hard to apply a number of balancing approaches. Doing so can enhance the credibility of the results that survive across methods. Using a variety of approaches

[22] Of the 28 papers from the AJPS we reviewed for our underlying matching project, 22 use two-stage estimation and 6 use balancing alone. None provide both single stage and two-stage results.

also limits the extent to which researcher discretion can influence reported results. Between the results presented here, and the various robustness checks in our appendix, we hope to have demonstrated a reasonable approach to using a variety of balancing methods.

## References


Abadie, Alberto, and Guido W. Imbens (2011), Bias-Corrected Matching Estimators for Average Treatment Effects, 29 *Journal of Business & Economic Statistics* 1-11.

Aldrich, John H., Andrew O. Ballard, Joshua Y. Lerner and David W. Rohde. (2017), Does the Gift Keep on Giving? House Leadership PAC Donations before and after Majority Status, 79(4) *Journal of Politics* 1449-1453.

Angrist, Joshua D. (1998), Estimating the Labor Market Impact of Voluntary Military Service Using Social Security Data on Military Applicants, 66 *Econometrica* 249-288.

Angrist, Joshua D., and Jorn-Steffen Pischke (2008), *Mostly Harmless Econometrics*.

Ansolabehere, Stephen, John M. De Figueiredo, and James M. Snyder Jr. "Why is there so little money in US politics?." *Journal of Economic perspectives* 17, no. 1 (2003): 105-130.

Arpino, Bruno, and Fabrizia Mealli. (2011) "The specification of the propensity score in multilevel observational studies." *Computational Statistics & Data Analysis* 55, no. 4: 1770-1780.

Barber, Michael J., Brandice Canes-Wrone, and Sharece Thrower. (2017) "Ideologically sophisticated donors: Which candidates do individual contributors finance?." *American Journal of Political Science* 61, no. 2: 271-288.

Beck, Nathaniel. (2018) Working paper "Estimating grouped data models with a binary dependent variable and fixed effects: What are the issues." *arXiv preprint arXiv:1809.06505* .

Bonica, Adam. (2014) "Mapping the ideological marketplace." *American Journal of Political Science* 58, no. 2: 367-386.

Coupé, Tom. (2005) "Bias in conditional and unconditional fixed effects logit estimation: A correction." *Political Analysis* 13, no. 3: 292-295.

Crump, Richard K., V. Joseph Hotz, Guido W. Imbens, and Oscar A. Mitnik (2009), Dealing with Limited Overlap in Estimation of Average Treatment Effects, 96 *Biometrika* 187-199.

Druckman, James N., Erik Peterson, and Rune Slothuus. "How elite partisan polarization affects public opinion formation." *American Political Science Review* 107, no. 1 (2013): 57-79.

Druckman, James N., Matthew S. Levendusky, and Audrey McLain. "No need to watch: How the effects of partisan media can spread via interpersonal discussions." *American Journal of Political Science* 62, no. 1 (2018): 99-112.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. (2015), Underreporting in political science survey experiments: Comparing questionnaires to published results. *Political Analysis* 23, no. 2: 306-312.

Greene, William H. (2003), *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall.

Hainmueller (2012), Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies, 20 *Political Analysis* 25-46.

Hansen, Ben B. (2004), Full Matching in an Observational Study of Coaching for the SAT, 99 *Journal of the American Statistical Association* 609-618.

Ho, Daniel W., Kosuke Imai, Gary King, and Elizabeth A. Stuart (2007), Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference, 15 *Political Analysis* 199-236.

Huddy, Leonie, Lilliana Mason, and Lene Aarøe. "Expressive partisanship: Campaign involvement, political emotion, and partisan identity." *American Political Science Review* 109, no. 1 (2015): 1-17.

Iacus, Stefano M., Gary King, and Giuseppe Porro (2012), Causal Inference without Balance Checking: Coarsened Exact Matching, 20 *Political Analysis* 1-24.

Imai, Kosuke. (2005). Do get-out-the-vote calls reduce turnout? The importance of statistical methods for field experiments. *American Political Science Review*, *99*(2), 283-300.

Imai, Kosuke, and Marc Ratkovic. (2014) "Covariate balancing propensity score." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76, no. 1: 243-263.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. "Affect, not ideology a social identity perspective on polarization." *Public opinion quarterly* 76, no. 3 (2012): 405-431.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. "The origins and consequences of affective polarization in the United States." *Annual Review of Political Science* 22 (2019): 129-146.

Katz, Ethan. (2001) "Bias in conditional and unconditional fixed effects logit estimation." *Political Analysis* 9, no. 4: 379-384.

King, Gary. (1995) Replication, Replication. *PS: Political Science & Politics* 28, no. 3: 444-452.

Lupia, Arthur., & Elman, Colin. (2014). Openness in political science: Data access and research transparency. *PS: Political Science & Politics*, 47(01), 19-42.

Mann, Thomas E., and Norman J. Ornstein. *It's even worse than it looks: How the American constitutional system collided with the new politics of extremism*. Basic Books, 2016.

Mason, Lilliana. "“I disrespectfully agree”: The differential effects of partisan sorting on social and issue polarization." *American Journal of Political Science* 59, no. 1 (2015): 128-145.

McCarty, Nolan, Keith T. Poole, and Howard Rosenthal. *Polarized America: The dance of ideology and unequal riches*. MIT Press, 2016.

Stuart, Elizabeth A. (2010), Matching Methods for Causal Inference: A Review and a Look Forward, *Statistical Science* 25(1): 1-21.

Urban, Carly., & Niebler, Sarah. (2014). Dollars on the Sidewalk: Should US Presidential Candidates Advertise in Uncontested States?. *American Journal of Political Science*, *58*(2), 322-336.

Ward, Michael D., and John S. Ahlquist, *Maximum Likelihood for Social Science: Strategies for Analysis*. Cambridge University Press, 2018.

Wooldridge, Jeffrey M. (2010) *Econometric analysis of cross section and panel data*. MIT press,

Wooldridge, Jeffrey M. (2015), *Introductory econometrics: A modern approach*. Nelson Education.

Zeileis, Achim, Christian Kleiber, and Simon Jackman (2008), Regression models for count data in R. *Journal of Statistical Software* 27, no. 8: 1-25.

Zhao, Qingyuan, and Daniel Percival. "Entropy balancing is doubly robust." *Journal of Causal Inference* 5, no. 1 (2017).