What Was the Persuasive Effect of Televised Campaign Advertising in the 2016 Presidential Election?

Arne Holverscheid[1,*], Brian Libgober[1], and Guillaume Pouliot[2]

[1] *Department of Political Science, Northwestern University, Evanston, IL, USA*

[2] *Harris School of Public Policy, University of Chicago, Chicago, IL, USA*

[*] *Corresponding author: Arne Holverscheid (arne.holverscheid@northwestern.edu)*

Abstract

Presidential campaigns raise billions of dollars. They do so by promising to use these donations to achieve their donors' goals of persuading fellow voters. Paid television advertising remains, by far, the largest line-item in Presidential campaigns budgets. But what is the persuasive return for that enormous public investment? We seek to answer this question through the analysis of a high frequency, seventeen-week survey asking a large sample of more than 4,500 voters about their voting intentions every week of the 2016 Presidential election. While before-and-after surveys are not new to political science, the extensiveness of repeated observations is the survey we use is unusual, and has not been used to examine questions about campaign spending efficacy. We expect a small, positive persuasive effect of advertising with diminishing returns. Null effects, large effects, or increasing returns would disconfirm our hypotheses.

During the 2016 election, Hillary for America paid a single media consultant, GMMB, 308 million dollars for the purpose of buying broadcast media advertisements.[1] Even in US Presidential politics, that is a lot of money. It amounted to well over half of the campaign's total spending. Spending, one might add, that depends on donations from ordinary citizens, who trust campaigns will not squander these funds on lavish rewards for politicians and well-connected political consultants, but instead use the money to influence the future course of politics and policy in the donor's preferred direction. To that end, campaigns make overwhelmingly large investments in advertising, especially television advertising. Indeed, television ads are almost always the biggest single line item for a US Presidential campaign. That fact remains true even despite the increasing prominence of social media in politics. Given the important role of campaign spending on the conduct of elections, a substantial literature has sought to evaluate what impact, if any, this expense has on whether citizens will vote and for which candidate [12, 7].

Considering the scale of investment in television advertising by Presidential campaigns, and also the common notion that these ads are persuasive [16], one might be forgiven for thinking that findings of efficacy would be a foregone conclusion. Indeed, in the wake of the 2016 election, the Clinton campaign was heavily criticized for misplacing ads in the waning days of the campaign [19], as if diverting a few thousand dollars from non-competitive Nebraska (whose media-market actually overlaps with a swing state, Iowa) to Wisconsin or Michigan would have changed the course of history. Seldom considered are the decidedly mixed findings in the social science literature. On the one hand, most of the recent experimental evidence finds effects of campaign ad spending that are quite limited [5, 1, 14]. On the other hand, credible observational studies may find small but statistically significant impacts [13, 17]. Given the competitiveness of US Presidential elections, these small effects are still substantively significant because they would make television advertising a pivotal explanation for who wins and why. Given the departure of the social scientific literature from general expectations, theory needs to account for the relatively limited efficacy found in empirical research. Psychology research suggests the possibility of a saturation effect: the more information available to citizens about candidates via news and other sources, the less marginal effect one should expect from campaign advertising [15]. If the saturation hypothesis is accepted, then the effects of television advertising in Presidential campaigns should be smaller than in other sorts of elections. Indeed, for primary and lower-level election contests, there is more agreement that ad spending matters between experimental and observational approaches [11, 17]. And yet, the disagreement between credible observational and experimental evidence in the Presidential general election domain raises questions, especially given the stakes of Presidential contests and the scale of the spending observed. Reconciling the apparent disagreement between observational and

[1] https://www.fec.gov/data/disbursements/?data_type=processed&committee_id=C00431569&committee_id=C00575795&recipient_name=GMMB&two_year_transaction_period=2016

experimental approaches is an important task for the literature on campaign effects going forward.

There are a number of possible explanations for why experimental and observational approaches seem to disagree. Social scientists are naturally tempted to credit the RCTs and discount observational studies, even those with plausible identification strategies. Admittedly, it is hard to fully discount the possibility that even credible observational studies are influenced in important ways by subtle lurking variables, reverse causality, and other issues. On the other hand, the findings of the RCTs run counter to common wisdom and lay expertise, which encourages a deeper investigation of the trade offs in methods. Indeed, a careful analysis of prior experimental studies in this domain lead to serious concerns about external validity. In particular, most experiments may be criticized on the grounds of lacking a *dosage* that can be expected to generate a substantive effect, and in this sense are potentially under-powered. If advertising needs a certain scale to have meaningful effects, then that could explain why RCTs do not find expected effects. To use a medical analogy, a typical dose of aspirin is 100 mg. In this space, it is as if we had good observational studies showing small, but pivotal efficacy at 100 mg, and RCTs that use 1 mg doses which show little efficacy. Moreover, while some observational studies are fairly credible because of their clever use of geographical discontinuities, they are limited by observing outcomes at a geographical rather than individual level. They also collapse all spending over time. Aggregating in both ways, one finds that votes in places respond to expenditure in the same places, which supports the inference that someone must have been persuaded somewhere at sometime. Of course, if campaigns target places where they expect to do well, such studies can get causality backwards. In our view, a stronger approach would directly examine whether ads persuade any individuals at any point in time. That said, the costs of larger surveys have until this point made such an exercise difficult.

In this report, we repurpose a panel survey that tracked 4509 voters weekly for the duration of the 2016 US Presidential general election campaign (n = 50,163). Other studies have used survey data to identify the persuasive effects of actual as opposed to randomized campaign spending [8], however these studies have not used surveys that recontact the same individual over time, certainly not as frequently, and not for as long, as the survey we use. The frequency of contact with likely voters allows us to examine how over time variation in ad spending relates to over time variation in attitudes. Moreover, these repeated same-sample polls are receiving increasing use to study Presidential as well as more local elections, which gives the analysis we develop value as a template for future work in a similar vein in other contexts. Although our study is observational because we do not randomly assign treatment, we are able to use person fixed-effects that control implicitly for demographic variables that observational studies usually control for (i.e. race, gender) and also implicitly for those that are hard to measure (i.e. cognitive ability, personality traits, media consumption habits, and so forth).

With the basic research question and empirical approach in mind, we proceed to formulate our precise statistical hypotheses. These details are recapitulated in our design table and further elaborated after we formulate our research hypothesis and statistical null hypotheses.

- Hypothesis 1: Increased intensity of spending on television ads by a U.S. Presidential candidate is associated with increased likelihood voters will intend to vote for that candidate. This hypothesis is associated with the following three, related null hypotheses.
  - H_{1A}: An increase in observable spending by Clinton has no relationship with vote intentions toward Clinton.
  - H_{1B}: An increase in observable spending by Trump has no relationship with vote intentions toward Trump.
  - H_{1C}: An increase in net observable spending by Clinton over Trump has no relationship with net intentions to vote for Clinton over Trump.

Rejecting any of these null hypotheses would provide statistical evidence consistent with the hypothesis that Presidential campaign spending has persuasive power. Given prior experimental work, one must anticipate the possibility of failing to reject the null with sufficient confidence. Therefore, a preregistered analysis would prudently consider how to define the minimal effect that is substantively meaningful. As explained below, the literature is silent about the smallest campaign effect that is meaningfully different from zero. Part of the reason is that a 1% difference in the outcome of any particular election is potentially outcome determinative, and if it were truly the case that Presidential campaigns could influence their electoral margin by 1% by spending twice as much on advertising as they currently do, that effect would still probably be considered large enough to be meaningful, even though that is indeed quite a small effect of campaign spending. Logically, it is hard to see any obvious limiting principle in this direction.

On the other hand, budget constraints on campaigns are real even if they are not precisely stated. It is hard for these campaigns to raise and spend more given the limits of donor interest and support. If they could raise more easily, they would. Additional spending in one place typically comes at a cost to some other priority.

Given these competing considerations around small but potentially electorally pivotal effects, we intend to consider three thresholds that would appear meaningful to a Presidential campaign and conduct frequentist equivalence testing for our estimates against these effects. We argue the smallest meaningful effect of spending would be whether a 10%, 100%, or 1,000% increase in campaign spending more would produce a 1% shift in support for the candidate. Formally, we shall call these critical values β_{10}, β_{100}, and $\beta_{1,000}$. The exact numeric values of these critical criteria depend on the functional form of the regression equation (i.e. OLS and logistic regression imply differing values for these coefficients), but in our preferred specification $\beta_{10}^{OLS} \approx 0.24$, $\beta_{100}^{OLS} \approx 0.03$, and $\beta_{1,000}^{OLS} \approx 0.01$. If a Presidential campaign must spend more than ten times what they do today to shift aggregate attitudes 1%, then from a campaign's perspective these advertisements are basically useless. Distinguishing a nearly zero elasticity of attitudes with respect to money (such as $\beta_{1,000}$) could prove challenging. Our prior simulation studies suggest we have more than adequate power even for these stringent tests (assuming mis-specification similar to the scenarios we can test). However, the larger cutoffs of β_{10} and β_{100} are still plausibly regarded as the minimum effect of substantive interest and therefore merit consideration as well. FEC disbursement data indicates that a 1% increase in spending would have cost the Clinton campaign alone around 3 million USD.

Besides the possibility of positive effects and null effects, another possible outcome of analysis is a mixed set of positive and null findings. If Clinton or Trump's ads were differently effective, it would not be hugely surprising from a theoretical or literature perspective [5]. That possibility would heighten questions about how and why particular ads and ad buyers are chosen by campaign, rather than whether campaigns are right to be engaging so extensively in this form of communication. For the purposes of equivalence testing, we shall assume that our estimate $\hat{\beta}$ is strictly smaller than the critical value we test against. If we can reject the statistical hypothesis that $\hat{\beta} = \beta_{1,000}$, it would present evidence that the effect of campaign spending on attitudes is vanishingly small. If $\hat{\beta}$ is not statistically distinguishable from $\beta_{1,000}$ but is statistically distinguishable from β_{100}, that still provides compelling evidence that campaigns get an extremely low return on investment for these funds, but a return that is at least conceivably reachable for campaigns in their present mode of organization. It implies that campaigns would need to more than double their current spending to decide a close election in their favor, but how much more would be hard to say. Rejecting the equivalence of $\hat{\beta}$ and β_{10} but not the other two, smaller critical values, it would also provide evidence of a low return on investment, but not so low as to meaningfully differ from prior observational estimates using alternative methods [17]. If we cannot statistically distinguish $\hat{\beta}$ from β_{10} or from 0, then the study design will have proved inconclusive, likely due to sample size concerns and other estimation issues not anticipated prior to analysis. The possibility that ads would have a significant, counter-productive effect on persuasion by causing voters to prefer the opposite position is theoretically imaginable, however for simplicity we do not test against this alternative.

1 Methods

1.1 Ethics information

This research protocol has been reviewed by Northwestern University's Institutional Review Board (Study No. STU00218619) and complies with all relevant ethical regulations. As the research involves secondary research on previously collected survey data, no additional informed consent or compensation for survey participants is required for this research. More details on the underlying study are available on the Understanding America Study 2016 Election Page [22].

1.2 Design

Our study seeks to understand how variation in levels of spending by each campaign individually and in combination relates to shifts in individual-level attitudes. In this subsection, we describe our data sources, how we connect them, and some strengths and weaknesses of the observational design.

Our survey data leverages extensively (and exclusively) the USC Dornsife/LA Times survey poll from the 2016 election cycle, also called the "Daybreak" poll. This survey draws from an underlying group of respondents that are continuously enrolled by USC's Dornsife Center under the auspices of the Understanding America Study, which examines a wide-variety of topics on economics, health, and other sorts of behaviors using this same pool of respondents. Members of the Understanding America Study are selected via probability sampling [20]. The panel uses a multi-step incentivized recruitment scheme (details available at Understanding America Study [21])[2] which involves an upfront payment of no less than $20 for joining the panel, the potential of an internet-connected tablet for those who do not have internet access, and compensation that works out to $40 per hour for providing survey responses. These are considerably stronger incentives than typical for most opinion surveys. The Daybreak Poll commissioned by the LA Times involved asking, on a daily basis, one-seventh of the UAS panelists who had opted-in to the election poll to answer a few questions about their voting intentions and perceptions of the race. On July 4, 2016, when the poll began, some 3,200 individuals had opted in to participate. Over time, the number of individuals opting in to the panel increased and eventually exceeded 4,500. Respondents were invited to participate on a daily basis, however not every panelist participated on the day they were invited to do so. If the respondent took more than one week to answer, their response was excluded. The polling response data thus constitutes an unbalanced panel with occasionally missing responses.

Although the Daybreak Poll has many unusually strong markers of a high-quality poll (i.e. probability sampling, strong participation incentives, large sample size, transparent methodology and extensive docu-

[2] `https://uasdata.usc.edu/index.php`

mentation), it is also the case that it was an "experimental" survey design, in the words of its designers,[3] and the poll was relatively (in)famous in the 2016 election for the extent to which it suggested an over-performance by Trump. For example, its final prediction was Trump winning the popular vote by 3% with a margin of error of 4.5%. Despite Trump's eventual win, the poll's overall prediction was erroneous because it sought to estimate the popular vote, which Clinton won by 2.1% (beyond the poll's margin of error). Reporting from data journalists at the time suggested the basic explanation for why the Daybreak Poll was such an outlier had to do with the upweighting of small subgroups, which every post-stratified survey does, and the fact the panel happened to draw a 19 year old black man from Illinois who was strongly supportive of Donald Trump.[4] A preference for Trump over Clinton was atypical for both young voters and black voters in the election, so the suggestion of that man's responsibility was tempting for the discrepancy from other polls. To their credit, the analysts at USC Dornsife did not change their methods midstream or in any way attempt to "fix" that particular issue. In the post-study post-mortem, the survey designers realized that their probability sampling method had over-included rural as opposed to urban respondents, and they had not weighted their study based on rurality.[4] Using weights to adjust for this over-representation so that the poll matches the US census, and even including the Illinois voter just mentioned, the Daybreak Poll would have predicted Clinton beating Trump by 1%. The latter was an accurate prediction within the poll's margin of error. None of these issues are particularly relevant for our purposes, as we are interested in individual-level treatment effects and not aggregate election predictions. That said, we consider the poll's notoriety and provenance worth mentioning.

Typical for studies of campaign advertising, we furthermore leverage data by the Wesleyan Media Project (WMP), which provides comprehensive coverage of political advertisements in the United States. The WMP grew out of an earlier project at the University of Wisconsin, named the Wisconsin Advertising Project (WiscAds). The WMP data is sourced from Kantar/CMAG and tracks ad airings on local broadcast, national broadcast, and national cable television going back to 2010, covering all 210 media markets in the country.

The granularity of the WMP data is a significant asset for our study. Each observation corresponds to an ad airing, capturing detailed information including the day, time, cost, and location (media market) alongside content variables such as tone or topic. In addition to the high level of detail in terms of time and place, we are also able to identify exactly who paid for an ad, as sponsors are classified into categories and identified by name. The data also provides us with the type of election an ad was intended to influence (i.e. "Presidential" or "Gubernatorial").

[3] https://election.usc.edu/2016//
[4] https://election.usc.edu/2016/

By matching the time and location of campaign ad spending captured by the WMP data to responses in the Daybreak poll, we will be able to correlate ad spending with changes in voter intentions. This linkage of individuals to campaign spending is key to our study design. Generally, campaigns purchase advertising from particular stations for specific times in individual media markets. A media market is defined as a collection of counties roughly belonging to the same geographic area. Media markets often cross-state lines, for example Lake County in Northwest Indiana is part of the Chicago media market, most of which is in Illinois. Public replication files from the UAS election studies do not contain geographic identifiers more disaggregated than the State, therefore in order to link spending to individuals we require access to lower level geographic information to make these links. The UAS allows researchers to apply for access to these data and to conduct analysis on their data on their secure platform ("Enclave"), where they are able to monitor data and code researchers import and verify there is nothing individually sensitive that they bring out. We obtained access to the Enclave and conducted some preliminary analyses about the suitability of our design for our goals, in particular calculating the number of respondents per media market by week. We discuss these analyses further below under the sections entitled Pilot study and Power Analysis. Note that the data we have collected to support our pre-registration are not on their own sufficient to estimate any effect of campaign spending on attitudes, because such an estimate requires some approach to determine what spending was observed. The Enclave is a third-party platform that intermediates all analyses and transfer of findings derived from analysis, which further bolsters the integrity of our pre-registered analysis.

Observational studies such as ours are typically subject to concerns about omitted variables confounding regressions. The UAS data includes much information on individual demographic traits that plausibly could have a relationship with the effectiveness of advertising in general (for example education level) or the effectiveness of specific campaign appeals (for example gender was plausibly implicated by many Clinton ads emphasizing her status as the first US female Presidential candidate). That said, these demographic traits do not vary and an individual fixed effect implicitly controls for these. One of the great strengths of the individual fixed effect design is that it implicitly controls for immutable latent traits that are measured and also the ones that are unmeasured, in some cases traits that one would measure with difficulty (e.g. baseline interest or knowledge in politics, cognitive ability). Because most tracking polls are unlike the Daybreak Poll in so far as they draw *new samples* every week and *do not recontact the same person* repeatedly, they do not permit the use of individual fixed effects. Instead, most other studies that would attempt a similar sort of design would necessarily rely on covariate adjustment rather than individual fixed effects.

Although political attitudes and behaviors can be fairly well explained by individual-level explanations [10], time-varying environmental factors may matter too. In the case of election attitudes captured through polling, it is well known that these are unpredictable over time, despite the remarkable predictability of actual election results based on information available well before election day [9]. Scandals, threats foreign and domestic, and policy successes may influence individual attitudes toward particular candidates. To the extent that these causes of individual opinion change external to campaign activity operate as an aggregate shock to the public, these too can be controlled for through a temporal fixed effect. Temporal fixed effects control for many potential omitted variables that are baked into the political environment yet changing over time.

With two-way fixed effect designs, the major threat to inference is time-varying interactions with latent traits. For example, it is likely the case that there are latent features of individuals that do influence their receptivity to Presidential advertising, most notably perhaps how much television they watch. To the extent that these time-varying latent traits are correlated with campaign behavior, these would confound any estimates that we produce. It is not possible to conclusively rule these out, however particular threats to inference to the extent they can be made specific and well-proxied can also be tested.

We mention briefly one important limitation of our design, and one advantage not typically available with most designs or data sources. In an ideal world, the analysis of Presidential spending on individual attitude change would be sensitive to where and when individuals are actually paying attention to these advertisements. In the best of circumstances, we would have data indicating whether an individual actually watched a particular ad, or perhaps only slightly worse had the television on and tuned to the channel when a particular ad was run. This is imaginable data, but far beyond what exists today. A potential concern in our design is the lack of fit between what individuals are watching and what is being spent. If the Trump campaign only spends money on a particular network (e.g. Fox News) or a particular program (e.g. *The Apprentice*), then it is hard to see how that treatment could possibly have an effect on those who only watch a different network or do not watch that program. Of course, it is easily verified that campaigns are not particularly selective in their approach to selecting outlets, particularly toward the end of the campaign they tend to saturate media markets in electorally competitive states across all times and places. As a result, increasing intensity of campaign advertisement overall is typically related to increasing intensity of ads across the board. Moreover, the fact that Clinton out raised and out spent Trump in many markets, but to varying degrees, creates a unique opportunity to uncover meaningful variation in total spending on individual attitudes. That said, it is the case that there is likely some slippage with campaigns intending to "concentrate their fire" in particular networks and times of day to court particular audiences (and also to take advantage of differing costs of ads across time), and the probable effect of this slippage baked in the design is to attenuate the estimates of campaign advertising's true effect. While it would be desirable to have an analysis that could examine campaign advertisement levels by channel and modulate these intensities by the media consumption behavior of the particular respondents [8], this is not possible with the UAS data.

Prior observational research has had the same limitations but additionally has often collapsed all variation in over time spending [13, 18, 17].

Turning from the downsides to the positive, one major recently developing concern in the analysis of polling data is that response bias may have gotten larger, particularly in polls that do not use careful probability sampling methods like the Daybreak study [6]. Moreover, the way in which they are not representative is potentially related to the outcomes of the survey, particular the sort of Presidential election surveys which are our focus. For example, if there is a scandal involving a candidate, that candidate's supporters might become relatively more reluctant (or depending on the circumstances, perhaps more enthusistic) about expressing their opinions in a survey. Polls vary wildly over the course of an election cycle, likely at least in part because of these dynamics in survey response bias. A benefit of the recontact design for individuals is that it attenuates these temporal dynamics. Individuals could still decide not to participate in a particular week, but when a person decides to not respond it is at least visible to the researcher. Although we assume for the purposes of our preanalysis plan that failures to respond are instances of data that is missing at random, we do intend to examine the validity of that assumption and, if evidence of a problem is detected, to suggest appropriate supplementary analyses.

1.3 Pilot Data

In order to develop a pre-registered analysis plan for the Daybreak poll, we conduct a pilot study. In a typical experimental approach, features such as sample size and treatment assignment mechanism are designed by the researcher prior to any examination of outcome variables, often through fielding a smaller pilot study. In an observational study, sample size and treatment pattern are not subject to the control of researchers in the same way or to the same degree. That said, analysts often make choices about how to measure treatment, procedures for the inclusions or exclusions of units, and so forth, that are analogous to decisions about sample size and assignment mechanism that an experimenter would make. Simply put, without a preliminary consideration of the observational data, it is difficult to develop a comprehensive analysis plan.

We began our pilot study by extracting counts of individuals within each Designated Market Area (DMA). Media markets define the geographical areas where ads are broadcast. Knowing the distribution of survey respondents across these areas allows us to determine how many people were exposed to various levels of treatment, without giving any actual indication of how that treatment affected them. We then classified these DMAs based on their exposure to campaign advertisements by week. The terminology of de Chaisemartin and coauthors is useful [2]. In their framework, there are four distinct treatment groups in a staggered adoption design: positive switchers, negative switchers, stayers, and never treated. Positive switchers are those that receive an increased level of campaign advertisements in one time period to the next. Negative switchers are exposed to a lower level of campaign advertisements in the subsequent period than in the previous one. Stayers had a (roughly) consistent level of positive exposure across time period, while never treated units are those that have not yet received a treatment. Note that the staggered adoption design, while similar to the context we examine, is importantly different because in a staggered adoption design, changes in treatment are irreversible once they have occurred. Campaign ads roughly follow a staggered adoption pattern, but it is not unusual for campaigns to drawdown their fire on particular areas or even abandon them entirely, making their approach a weak fit for our case. While we ultimately find conceptualizing our observational study through a staggered adoption design to be too difficult given the important differences, the terminology and developments in the staggered adoption design inform our thinking about how to design this study.

1.4 Analysis Plan

Using two-way fixed effects, our study seeks to estimate the impact of television ad spending by 2016 Presidential campaigns on voting intentions using an unbalanced, seventeen week panel survey where individuals are recontacted on a weekly basis. The ultimate goal of our design is, therefore, to estimate the following statistical models:

$$\Pr(\text{VoteClinton})_{it} = \beta \cdot \log_{10}\{\text{ClintonSpendingIntensity}_{it} + 1\} + \psi_i + \theta_t + \epsilon_{it} \tag{1}$$

$$\Pr(\text{VoteTrump})_{it} = \beta \cdot \log_{10}\{\text{TrumpSpendingIntensity}_{it} + 1\} + \psi_i + \theta_t + \epsilon_{it} \tag{2}$$

$$\Delta \Pr(\text{VoteClinton})_{it} = \beta \cdot \log_{10}\left\{\frac{\text{ClintonSpendingIntensity}_{it} + 1}{\text{TrumpSpendingIntensity}_{it} + 1}\right\} + \psi_i + \theta_t + \epsilon_{it} \tag{3}$$

where ψ_i is the individual fixed effect, θ_t is the time fixed effect, and ϵ_{it} is the disturbance term. Because the range of the outcome variables are constrained in each equation, generalized linear models are theoretically appealing, however there are technical limitations that prevent their use. In particular, as Chamberlin (1980) notes, [3] fixed effects with generalized linear models such as logit or probit have an incidental parameters problem and are biased if the number of periods is relatively small, and our panel has at most seventeen periods of observation on each indvididual. Even with remarkable progress in the analysis of difference-in-difference or two-way fixed effect specifications, there is still no solution we are aware of to this limitation [23]. In supplementary analyses, we will fit a conditional logit specification with individual fixed effects a la Chamberlain (1980) on a dichotomized outcome variable, however for the main analysis we consider the robustness we obtain via two-way fixed effects to be more desirable than a form for the likelihood function with the theoretically correct range. Further, we note that the key explanatory variable is highly skewed and likely subject to diminishing marginal returns because of the saturation effect, therefore logging this variable is appropriate in our main specifications. In many locations, the amount of spending observed is 0 in which case the log would be undefined, for this reason we add \$1 of spending to every place. We intend to examine other functional forms in supplementary analyses, however the log specification also ties most cleanly to our preferred interpretation of the coefficients. In particular, the coefficient may be interpreted as the percentage increase in spending across the board that would be necessary to generate a 1% shift in the outcome variable.

In implementing this analysis, an important observation is that what a candidate spent at time t is not necessarily the entirety of the relevant treatment for individual attitudes at the same time. Indeed, we are mindful that most observational studies in the space have treated all spending observed for the *entirety* of a Presidential campaign as the treatment, implicitly treating spending in the far past as equally relevant to election outcomes as today's spending,[5] We are aware of only two prior studies that have examined individual respondents over time. While Franz and Ridout [8] uses a measure well-tailored to the individual's viewing habits, it is unclear what time horizon of ads they considered relevant for that individual. Gerber et al. [11] provides two approaches to this issue, one is to use a model of geometric decay via lagged dependent variables and the other is lagged regressors. While time-series models may be appropriate, they raise the technical complexity of estimation considerably (especially with fixed effects) in a way that is particularly troubling since these authors also "find relatively little evidence of time-series dynamics. Simple models, in other words, lead to roughly the same substantive conclusions as more elaborate models"[11].

As we view our investigation as more about *whether* there is persuasion from Presidential campaign ads and not *how durable* the persuasion is, we prefer not to estimate the rate of decay directly, but rather to examine robustness of estimates of persuasive advertising to various models of geometric decay. In particular, for each analysis we conduct, we posit a fixed δ discount factor that is applied to actual past spending to produce a derived variable, which is similar but importantly different, called "candidate spending intensity." In particular

$$\text{CandidateSpendingIntensity}_{it} = \text{CandidateSpending}_{it} + \sum_{j=1}^{\infty} \delta^j \cdot \text{CandidateSpending}_{i(t-j)} \tag{4}$$

For example, $\delta = 0$ means campaign spending intensity is the same as actual campaign spending only on that day t, an unlikely assumption if media effects are indeed persuasive. By contrast $\delta = 1$ implies that all spending observed weeks past contributes equally to today's spending intensity, also an unlikely assumption

[5] It is interesting to note that is another potential source of attenuation bias in prior observational work

if memories are fallible and individual political opinion has a mean-reverting tendency. Implicitly, $\delta = 1$ has been the typical approach of prior observational studies [e.g. 18]. More plausible given the findings of Gerber and coauthors [11] is a moderate δ value somewhere in between these extremes. For example, $\delta = 0.5$ implies that candidate spending today is worth twice as much as spending yesterday, four times what spending is worth the day before that, eight times what spending is worth the day before that, and spending more than a week old is worth a tiny fraction of today's spending. We have no prior point of view on which δ is most appropriate, beyond a (rebuttable) skepticism about extremes. Indeed, it seems likely to us that to the extent advertising is effective, individuals probably differ in how durable the effect of this advertising is on them, so the δ might vary across study samples. The geometric model is obviously an approximation in the best of cases. That said, the geometric model of decay is flexible enough to capture a wide variety of discounting patterns to a first approximation provided a broad enough range of values are chosen. Indeed, we propose to conduct our regressions for all δ between 0 and 1 in increments of 0.01 (or finer, computation times permitting). In examining robustness across a range of values, we are able to avoid some of the complexities of time-series analysis while still addressing the important issue in a way consistent with prior practice. If no effect is found for any δ, our results and discussion will focus on the case $\delta = 0.5$. In particular, we will report equivalence testing of our estimates against β at our three meaningful levels[6] using the δ-level of 0.5. We will include in the Supplementary Information equivalence testing for these β at a range of other δ levels as well. If a significant β is found for some δ, we will focus on presenting the results with δ that leads to the smallest significant β and discuss the robustness (or lack thereof) to different choices about δ.

1.5 Power Analysis

In order to further bolster our analysis plan, we also investigate the power of our dataset against these potential hypotheses. To that end, let us restate equation (1).

$$\Pr(\text{VoteClinton})_{it} = \beta \cdot \log_{10}\{\text{ClintonSpendingIntensity}_{it} + 1\} + \psi_i + \theta_t + \epsilon_{it}$$

Recall that ψ_i is the individual fixed effect, θ_t is the time fixed effect, and ϵ_{it} is the disturbance term. A power analysis should take the β of interest as given, predict the outcome variable with an appropriate model of randomization, and reestimate $\hat{\beta}$ in a way that ideally will be close to its theoretical value, but depending on the treatment and randomization may not be. Replicating that analysis many times can give a sense for

[6] To reiterate, these levels imply a 10% increase in spending for 1% increase in support, doubling spending for a 1% increase in support, and multiplying spending by ten times for a one percent increasing support

the probability of detecting an effect. In considering how the power analysis should look like in our analysis, we note that *treatment* in this case is a completely determined function of a particular individual's location. Once one knows which individual is in the study, their location and treatment is fixed and therefore the typical experimental conceptualization of randomization of treatment assignment is inappropriate. That said, the survey designers could have picked a different person from each place, and all that matters about the individual in our model is their fixed effect. Provided one knows the values of the fixed effects that are available, as well as the distribution of the disturbance terms, one can calculate the probability of voting for Clinton (Trump) for various hypothetical values of β and then calculate the probability of detecting an effect significantly different from 0.

A key challenge, however, is that neither the fixed effects nor the disturbance terms are known beforehand. That said, it is possible to estimate their values on *untreated* units. Although determining the effect of campaign spending requires low-level and confidential geographic identifiers, the public release of the Daybreak data files indicates respondents state and it is possible to therefore infer which units could not have observed Presidential campaign spending as of Week 1, Week 2, and so forth of the 2016 campaign. An important assumption driving this analysis is that the untreated units have the same distribution of fixed effects and distrubance terms as the treated units, which is not verifiable before the study is conducted.

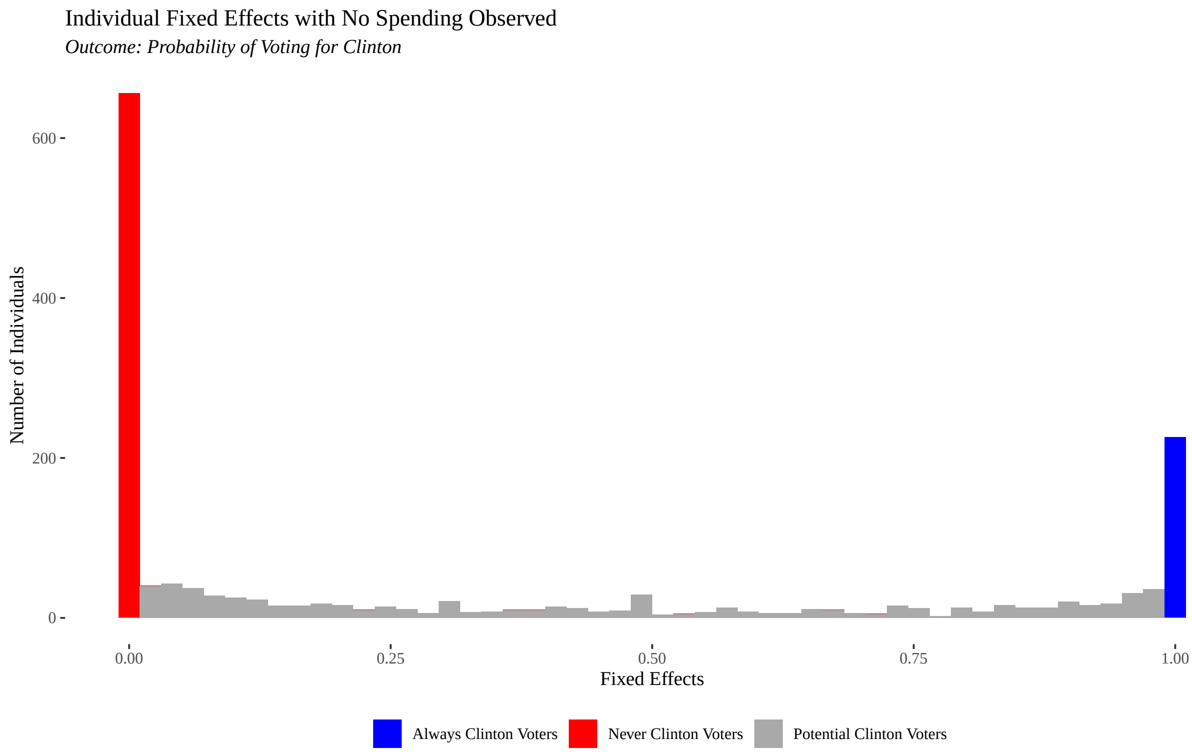
With untreated units, we estimate a regression like the following:

$$\Pr(\text{VoteClinton})_{it} = \psi_i + \theta_t + \epsilon_{it}$$

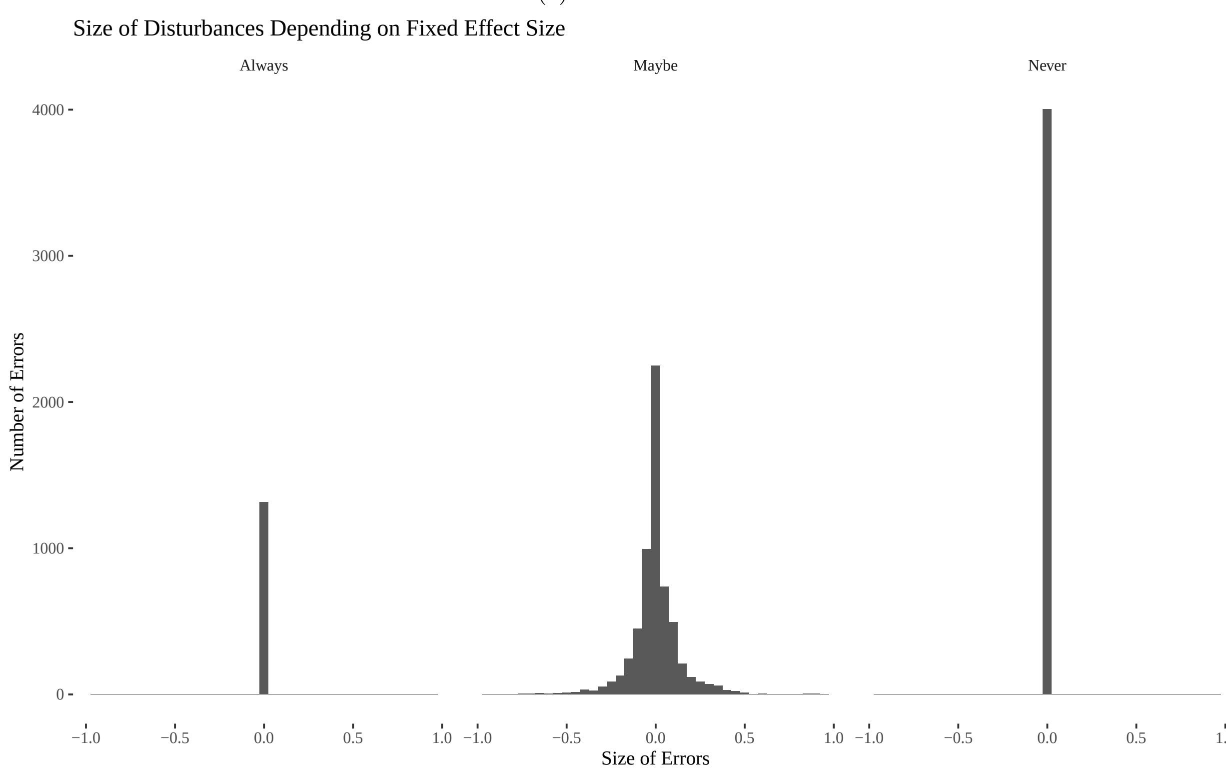
and "plug-in" the estimated values of the parameters to simulated data that we can use to estimate statistical power. In practice, we do this simulation as follows. For each unit that we know will be exposed to a given, geographically-defined treatment pattern[7], we draw a fixed effect at random and with replacement from the pool of estimated fixed effects from the untreated units. We then add the appropriate disturbance term by sampling from a sub-pool of residual terms depending on the size of the fixed effects. In particular, this sub-sampling approach is justified by the fact that a significant proportion of the electorate expresses 100% or 0% support for Clinton (Trump) across all time periods. These "extreme" individuals have fixed effects that are very small or very large and residuals that are close to 0 (see Figures 1a and 1b). While the number of committed respondents in a survey is rightly regarded as subject to some uncertainty, pooling residuals between committed and uncommitted respondents would likely add more noise to our simulations of power than is ideal. Our approach to power analysis assumes that there is no randomization over the treatment

[7]Recall from the pilot study that we know how many individuals there are in each media market, but not which media market any individuals are in, therefore we know how many individuals exposed to a particular treatment pattern there will be but not which individuals they actually are

plan (i.e. number of units receiving a possible pattern of exposure to ads), but there is randomness over which individuals gets assigned a particular treatment plan. The treatment pattern is fixed in so far as the number of units per county is fixed via the survey designers and the treatment pattern a unit receives is completely determined by the geography. Yet who the survey designers actually pick to fill the geographic quota is random. Our power analysis emulates this by simulating draws over all potential assignments of individuals (and their fixed effects).



(a) Fixed effects



(b) Residuals on fixed effect estimates

Figure 1: Inverse Logit Regression of Unexposed Units

In order to estimate the fixed effects ψ_i that we use, we must decide how broadly or narrowly to construe the pool of untreated units. We focus on respondents who could not have yet received treatment but are located in a place that may later receive some Presidential campaign ads. Ad spending is heavily back-loaded, meaning that a large proportion of it occurs towards the end of the general election campaign. Around half of all spending for televised ads by the two presidential campaigns had not yet occurred just four weeks prior to election day. As a result, out of the 210 media markets in the data, 147 had not seen any spending by that same point in time. A respondent residing in the Honolulu media market for example can therefore be considered unexposed up until 3 weeks before the election date. By contrast, a resident of the Cleveland media market cannot be a part of our ex ante power analysis because the campaigns had already spent a quarter million USD there in the first week of the general election campaign. But not only did the gross amount of spending occur towards the end of the campaign, potential exposure of respondents is similarly back-loaded. Table 1 shows the number of respondents who were exposed to television ads by week, illustrating the treatment pattern found in the data. Strikingly, in the penultimate week of the campaign, more than half of all respondents had not received any exposure.

Figure 2 shows the composition of the pool of untreated respondents and the number of observations we have on each particular individual. For roughly one in three individuals in the sample of 1,605 unexposed individuals we have more than ten responses to the survey, while we have a similar but slightly larger number of individuals with fewer than five responses.

With simulated panels that "plug-in" individual fixed effects and disturbances from the unexposed survey respondents, we proceed to refit regressions like in Equations (1)-(3). Note that the size of δ has an influence on ability to determine if a particular β is significant, because implicitly less discounting means that the treatment intensity is higher. Figure 3 shows the likelihood of rejecting the null of $\beta = 0$ if $\delta = 0.5$ for various β ranging from 2.3, which implies a shift of 1% in probability of voting for a candidate would require an increase in spending intensity of 1% across the board, all the way to to 1.4e-3, which implies a shift of 1% in the probability of voting for a candidate would require spending 100 million times as much as the campaign spends presently (i.e. one billion percent increase in the explanatory variable). The figure demonstrates a more than 80% chance of rejecting the null if $\beta > 2.5e - 3$ for this δ.

As discussed in the analysis section, the findings about a particular β are dependent on how one aggregates spending over time in the key explanatory variable. Therefore, δ has an influence on power, with smaller δ and heavier discounting making it harder to detect an effect because the implied treatment dosage is smaller. Figure **??** shows more systematically for a range of δ values how power changes. Importantly, the probability of rejecting the null if the effect is truly greater than $\beta > 2.5e - 3$ is still 0.8 if δ is as small as 0.1 (implying a 90% reduction in the value of spending *per day*). It is worth repeating that we consider the

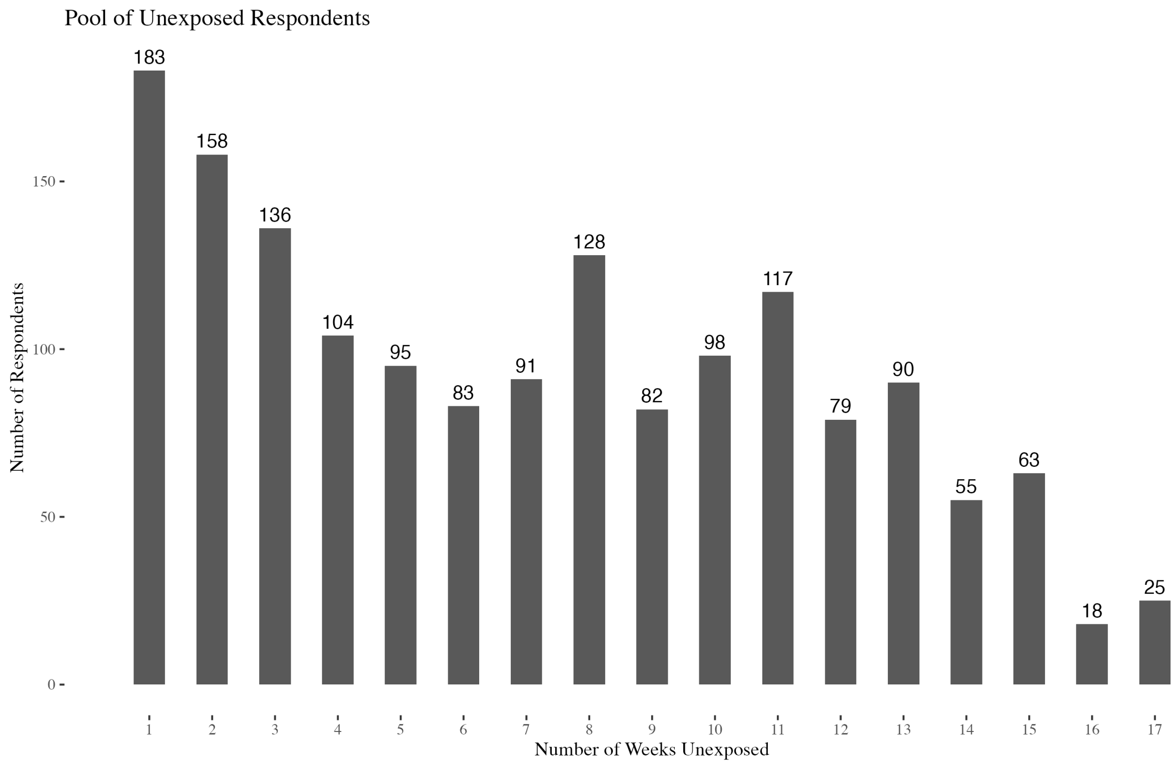


Figure 2: Number of respondents by number of weeks they have been unexposed

smallest meaningful effect to be one that implies it would take more than 10 times as much spending across the board to produce a one percent shift in the electorate ($\beta_{1000}^{OLS} = 0.0096$), therefore we appear to have more than adequate power assuming no misspecification.

Figure 4 illustrates the statistical power of detecting an effect of televised campaign advertising on voter intentions across a range of hypothetical beta values (β), given different discount factors (δ). As the figure demonstrates, for all tested discount factors ranging from 0.1 to 0.99, the statistical power is consistently high, indicating a high probability of detecting even small effects of campaign spending on voter intentions. Only with hypothetical elasticities of attitudes with respect to spending that asymptotically approach 0 does statistical power drop. This is true even under conservative assumptions about discount values, i.e. the decay of treatment effects.

This power analysis is reassuring and indeed expected given the size of the Daybreak panel (i.e. $n = 50,163$) and the extent of the variation in the explanatory variable. That said, the power analysis does assume that there is no misspecification. To further bolster our power analysis, we consider several forms of model misspecification. In particular, we consider two alternative models. In the first, we assume that the data generating process involves censoring dependent variable values larger than 1 or smaller than 0. In the power analysis previously described, if our true equation implies a probability of voting for Clinton of 132%

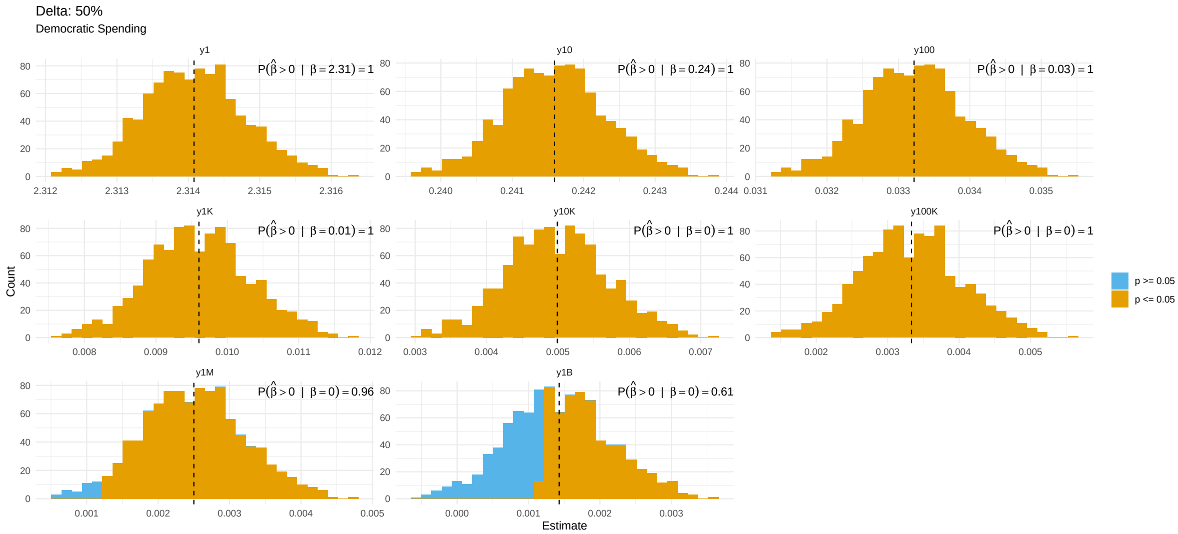


Figure 3: Probability of rejecting the null of no-effect if $\delta = 0.5$ for various true β ranging from 2.3 to $1.4e{-}3$, implying anywhere from a 1% increase in spending to a 100 million fold increase in spending to produce a 1% shift in voting preferences.

or -79%, we allow that and fit the equation anyway. The Daybreak Survey does not include response values that are not valid probabilities. Censoring 132% to be 100% or -79% to equal 0% violates the functional form assumptions of the true model, but makes the simulated data more similar to what we will eventually analyze. In the second alternative model, we assume the data generating process involves an inverse logit, which also enforces the constraint on the response variable but does so with a different functional form. Figures similar to the above demonstrate that either form of misspecification leads to bias in the estimates, but in a conservative direction (i.e. it is always harder to find an effect for a particular combination of β and δ if the data generating process is censored or the likelihood has an alternative functional form than if the model for analysis fits the data generating process). We also anticipate that misspecification of this type makes detecting an effect harder, as censoring limits the range of variation at which treatment may observably influence attitudes. Once spending by a candidate gets an individual to voting with them for certainty, any marginal spending in these models has no effect. Put differently, mis-specification of this type creates attenuation bias. Even if misspecification introduces a conservative bias to the approach, the figures in the Supplementary Appendix show more than adequate power to detect spending-attitude elasticises at the level we consider substantively meaningful.

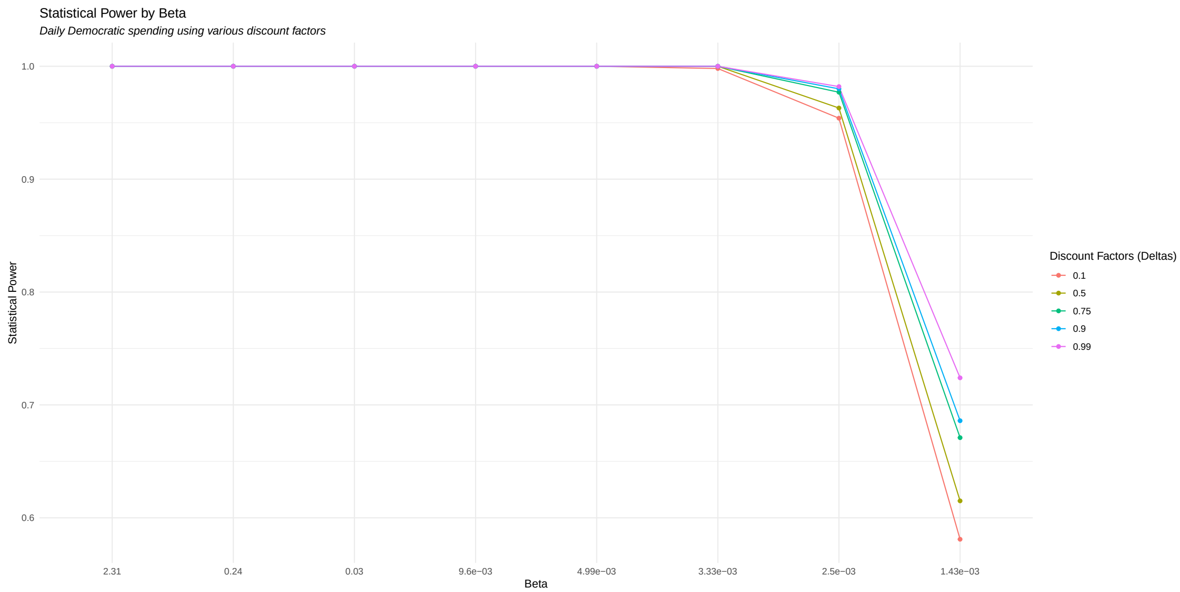


Table 1: Patterns of Campaign Ad-Spending Exposure during the 2016 Campaign

	Week 1	Week 5	Week 9	Week 13	Week 17	Week 18
Had prior ad exposure	232	570	981	878	1791	3753
Newly exposed (in last week)	0	0	202	0	226	37
Exposure increased (in last week)	179	363	639	792	1380	1814
Exposure decreased (in last week)	1253	280	342	86	863	2011
Not exposed (in last week)	1273	2180	2237	2541	2008	77
Not had prior ad exposure	1273	2180	2237	2541	2008	77
Eventually exposed	1268	2175	2232	2536	2003	72
Total respondents	1505	2750	3218	3419	3799	3830

2 Data Availability

We commit to sharing all relevant data and materials upon acceptance of our Stage 2 manuscript. We confirm that the Understanding America Study has collected geolocation information in compliance with privacy and consent requirements. They have further agreed to make the datasets used to conduct the study available, if requested, to other accredited researchers under the same conditions as we have been allowed access.

3 Code Availability

We commit to sharing all code used in the preparation of this report, including code used to simulate data, conduct the power analysis, and analyze pilot data. It will be publicly available on the Understanding America Study website.

Table 2: Design Table

Question	**Hypothesis**	**Sampling plan (e.g. power analysis)**	**Analysis Plan**	**Interpretation given to different outcomes**
What is the effect of increased intensity of spending on television ads by a U.S. Presidential candidate on the likelihood voters will intend to vote for that candidate?	Increased intensity of spending on television ads by a U.S. Presidential candidate is associated with increased likelihood voters will intend to vote for that candidate.	Sample of more than 4,500 voters surveyed weekly over seventeen weeks. Power analysis conducted to ensure sufficient power to detect small effect sizes.	Use two-way fixed effects regression models to estimate the impact of television ad spending on voting intentions, accounting for individual and time fixed effects.	Rejecting the null hypothesis would indicate a persuasive effect of campaign spending. Failure to reject would suggest null or negligible effects. Large effects would indicate substantial impact of advertising.
Does an increase in observable spending by Clinton have a relationship with vote intentions toward Clinton?	An increase in observable spending by Clinton has a positive relationship with vote intentions toward Clinton.	Same as above.	Specific regression model focusing on Clinton's ad spending and its impact on vote intentions for Clinton.	Positive significant results would indicate Clinton's ads are effective. Null results would suggest no effect.
Does an increase in observable spending by Trump have a relationship with vote intentions toward Trump?	An increase in observable spending by Trump has a positive relationship with vote intentions toward Trump.	Same as above.	Specific regression model focusing on Trump's ad spending and its impact on vote intentions for Trump.	Positive significant results would indicate Trump's ads are effective. Null results would suggest no effect.
Does an increase in net observable spending by Clinton over Trump have a relationship with net intentions to vote for Clinton over Trump?	An increase in net observable spending by Clinton over Trump has a positive relationship with net intentions to vote for Clinton over Trump.	Same as above.	Regression model focusing on the net difference in ad spending between Clinton and Trump and its impact on net vote intentions.	Positive significant results would indicate a competitive advantage for Clinton's ad spending over Trump's. Null results would suggest no differential effect.

References

[1] David E. Broockman and Joshua L. Kalla. “When and Why Are Campaigns’ Persuasive Effects Small? Evidence from the 2020 U.S. Presidential Election”. en. In: *American Journal of Political Science* (Aug. 2022), ajps.12724.

[2] Clément de Chaisemartin et al. *Difference-in-Differences Estimators for Treatments Continuously Distributed at Every Period*. arXiv:2201.06898 [econ]. Dec. 2023.

[3] Gary Chamberlain. “Analysis of Covariance with Qualitative Data”. en. In: *Review of Economic Studies* 225-238.1 (1980).

[4] Nate Cohn. “How One 19-Year-Old Illinois Man Is Distorting National Polling Averages”. In: *The New York Times, Upshot Column* (Oct. 2016). The 2016 Race.

[5] Alexander Coppock, Seth J. Hill, and Lynn Vavreck. “The small effects of political advertising are small regardless of context, message, sender, or receiver: Evidence from 59 real-time randomized experiments”. en. In: *Science Advances* 6.36 (Sept. 2020), eabc4046.

[6] Carina Cornesse et al. “A Review of Conceptual Approaches and Empirical Evidence on Probability and Nonprobability Sample Survey Research”. en. In: *Journal of Survey Statistics and Methodology* 8.1 (Feb. 2020), pp. 4–36.

[7] Erika Franklin Fowler, Michael M. Franz, and Travis N. Ridout. *Political Advertising in the United States*. eng. OCLC: 1041051339. Boulder: Routledge, 2018.

[8] Michael M. Franz and Travis N. Ridout. “Does Political Advertising Persuade?” en. In: *Political Behavior* 29.4 (Oct. 2007), pp. 465–491.

[9] Andrew Gelman and Gary King. “Why Are American Presidential Election Campaign Polls so Variable When Votes Are so Predictable”. In: *British Journal of Political Science* 23.4 (1993), pp. 409–451.

[10] Andrew Gelman et al. “The Mythical Swing Voter”. In: *Quarterly Journal of Political Science* 11.1 (2016), pp. 103–130.

[11] Alan S. Gerber et al. “How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment”. en. In: *American Political Science Review* 105.1 (Feb. 2011), pp. 135–150.

[12] Kenneth Goldstein and Travis N. Ridout. “MEASURING THE EFFECTS OF TELEVISED POLITICAL ADVERTISING IN THE UNITED STATES”. en. In: *Annual Review of Political Science* 7.1 (May 2004), pp. 205–226.

[13] Gregory A. Huber and Kevin Arceneaux. "Identifying the Persuasive Effects of Presidential Advertising". en. In: *American Journal of Political Science* 51.4 (Oct. 2007), pp. 957–977.

[14] Joshua L. Kalla and David E. Broockman. ""Outside Lobbying" over the Airwaves: A Randomized Field Experiment on Televised Issue Ads". en. In: *American Political Science Review* 116.3 (Aug. 2022), pp. 1126–1132.

[15] Leonard M. Lodish et al. "How T.V. Advertising Works: A Meta-Analysis of 389 Real World Split Cable T.V. Advertising Experiments". In: *Journal of Marketing Research* 32.2 (1995), pp. 125–139.

[16] Domenico Montanaro. "Do early campaign ads really make a difference in the presidential election?" In: *NPR* (2024). Heard on Weekend Edition Sunday.

[17] John Sides, Lynn Vavreck, and Christopher Warshaw. "The Effect of Television Advertising in United States Elections". en. In: *American Political Science Review* 116.2 (May 2022), pp. 702–718.

[18] Jörg L Spenkuch and David Toniatti. "Political Advertising and Election Results". en. In: *The Quarterly Journal of Economics* 133.4 (Nov. 2018), pp. 1981–2036.

[19] Jim Tankersley. "The advertising decisions that helped doom Hillary Clinton". In: *Washington Post* (Nov. 2016). Accessed: 2016-11-12.

[20] Understanding America Study. *Methodology.* Accessed: 2024-08-02. 2024.

[21] Understanding America Study. *Recruitment.* Accessed: 2024-08-02. 2024.

[22] *Understanding America Study 2016 Presidential Election Data Page.* `https://uasdata.usc.edu/page/UAS+2016+Presidential+Election`. Accessed: 2024-06-13.

[23] Jeffrey M Wooldridge. "Simple approaches to nonlinear difference-in-differences with panel data". en. In: *The Econometrics Journal* 26.3 (Sept. 2023), pp. C31–C66.