Optimal Targeting of Television Advertisements∗

Mitchell Lovett† Michael Peress‡

November 17, 2011

Abstract

In this paper, we study the targeting of advertising for the largest media expenditure for most firms—television. While television advertising was once used to “broadcast” to the widest possible audience, today with the explosion of channels and show content, it has the potential to be targeted to a narrower audience. In this paper, we investigate this potential in the context of congressional elections and present a method that estimates the effect of advertising. We find that advertising significantly influences candidate choice, but has a relatively small effect on turnout. We also develop and apply an approach for identifying the optimal allocation of media buys across television programs. Our results suggest that advertising can be effectively targeted, that candidates generally target in a way consistent with the most desirable generic strategy, and that the total effect of advertising is to increase relative vote shares roughly 2% over not advertising. By applying the technique we have developed campaigns could on average double the effectiveness of their television advertising campaigns. Consistent with prior expectations, close races and incumbents come closest to the optimal benefit of advertising.

∗We’d like to thank Eric Bradlow, Jean Pierre Dubé, Paul Ellickson, Ron Goettler, Brett Gordon, Wes Hartmann, Günter Hitsch, Sanjog Misra, Ryan Moore, and Ron Shachar; participants of seminars at Carnegie Mellon University, the Booth and Harris Schools at University of Chicago, the Simon School at University of Rochester, and UC San Diego; and audience members at the 8th Invitational Choice Symposium (2010), Marketing Science (2010), the Econometrics Society Summer Conference (2011), SICS (2011), Polmeth (2011), and QME (2011).
†Simon School of Business, University of Rochester. mitch.lovett@simon.rochester.edu
‡Department of Political Science, University of Rochester. mperess@mail.rochester.edu
1 Introduction

In the last two decades, the number and variety of television programs has grown dramatically, greatly increasing both the potential to target advertisements and the challenge of the targeting problem (Soberman 2005). These changes have occurred alongside shifting demographics that are disrupting existing wisdom regarding television advertising, drawing into question the value of traditional ways of targeting and measuring audiences (Neff 2011). For this new environment, we provide an approach to optimally target advertisements and use this framework to evaluate current practice.

Optimally targeting television ads involves allocating a budget to overlapping “segments” (shows). This problem is similar to targeting in traditional media (e.g., print, outdoor, radio) and new media (Yao and Mela 2011, Danaher, Lee and Kerbache 2010). For instance, targeting sponsored search ads involves selecting from overlapping keywords, targeting display ads involves selecting sites with overlapping readerships, and selecting mobile apps for in-app ads involves selecting from overlapping user bases. This similarity and the fact that television is the largest advertising category (IDC 2011) makes television a valuable setting to study optimal targeting decisions.

We study targeted advertising in the setting of elections. Campaign spending in 2008 reached over $5 billion, with advertising media the largest expenditure category. Not surprisingly, the literature on political advertising has received recent attention (Gordon et al. forthcoming), including consideration of the geographic allocation problem (Shachar 2009, Gordon and Hartmann 2010, Fletcher and Slutsky 2011) and the content selection problem (Hoegg and Lewis forthcoming, Lovett and Shachar 2011). We focus on optimally allocating ads across television programs. Our approach involves simulating advertising exposures, estimating the advertising exposure effect, and framing optimal targeting decisions as a non-linear program. We use this approach to demonstrate both how current practice diverges from optimal decisions and how campaigns could use our approach to improve decisions.

In the process of studying targeting, we add new evidence to the literature on political advertising (Gordon et al. forthcoming). This literature distinguishes two effects—the effect on whether an individual turns out to vote (turnout) and, conditional on voting, the effect on which party the voter chooses (candidate choice). Several studies conclude that advertising affects candidate
choice (Huber and Arceneaux 2007, Gerber et al. 2010). In contrast, the evidence on the turnout effect is mixed—for example, Huber and Arceneaux (2007) find no significant turnout effect, while Shachar (2009) finds a small but significant average effect by modeling heterogeneity in advertising response. We add to this literature by developing and applying a new approach for estimating advertising effects. We control for the endogeneity of advertising by including fixed effects at the media market level. Our identification is analogous to a difference-in-differences strategy.

We develop a multiple imputation approach to link the various datasets we employ (Kamakura and Wedel 1997). The ideal, single-sourced data that contains both advertising exposure and voting behavior is not available. In our approach, we build on a combination of individual level political data and aggregate viewing data and link the data by imputing the exposure. The result is individual-level data with the full set of variables. The imputed exposure is critical to estimating the correct advertising effect both because not accounting for exposure biases political advertising effects downward (Goldstein and Freedman 2002, Freedman, Franz and Goldstein 2004a) and because the individual level variation in exposure is essential to our identification strategy.

We apply our methodology to congressional races in the 2004 general election. We find that without controlling for endogeneity, both the turnout and the candidate choice effects of advertising are significant and positive, though the turnout effect is much smaller. Once introducing the controls, the turnout effect diminishes and becomes insignificant, while the candidate choice effect increases in strength. Interestingly, this is exactly the direction of endogeneity bias we would expect if there were in fact a candidate choice effect and candidates targeted their ads based on this effect. We estimate an average elasticity of approximately 0.075, and with a one standard deviation change above and below the average level of individual exposures we find average turnout changes by 0.2% and average candidate choice shifts by 3.5%. Thus, we find that in the context of congressional elections the candidate choice effect is significant and meaningful, while the turnout effect is smaller and less important to election outcomes.

Using these estimates, we evaluate how well candidates target ads and how they can improve. With a dominant candidate choice effect, we would expect candidates to target relatively cheap programs with audiences that have high turnout rates and contain largely swing voters. Candidate strategies appear to match this generic strategy, but concentrate too little on the cheapest, most effective programs. Optimal strategies can lead to more than doubling the effectiveness of adver-
tising in congressional races. In addition, we identify a simple set of heuristics such as targeting the top 5 highest turnout shows, which can be used to obtain nearly two-thirds of the value from optimal targeting. Finally, we identify circumstances in which candidates perform best. We find that incumbents and candidates in very close races perform closest to optimal.

This paper is organized as follows. In the next section we discuss related literature on optimal targeting of television ads. We then introduce our model in Section 3. We describe the datasets we use in Section 4 and discuss the estimation procedure and identification strategy in Section 5. We then present the estimation results in Section 6 and the targeting model and evaluation in Section 7. We conclude with a summary of our results and some limitations.

2 Related Literature

Designing optimal media plans for television has long been recognized as important. Considerable attention has focused on the scheduling (timing) of advertising (Doganoglu and Klapper 2006, Dubé, Hitsch and Manchanda 2005, Mesak and Zhang 2001, Naik, Mantrala and Sawyer 1998). A number of recent theoretical papers discuss the incentives to target advertisements and the resulting market outcomes (Iyer, Soberman and Villas-Boas 2005, Soberman and Sadoulet 2007) and two recent papers examine whether advertising generates value by being a matchmaker for consumers and products (Anand and Shachar 2009, 2011). Importantly, this latter literature has focused on providing explanations for why targeted advertising is practiced and testing those explanations. In contrast, we focus on how well campaigns are targeting advertisements and develop a methodology to improve targeting strategies.

An earlier literature offered a number of specific approaches to improving targeting decisions and examined how well ads were targeted.\(^1\) For instance, Assael and Poltrack (1996), Cannon (1988), and Danaher and Rust (1992) study the merits of using different data sources and integration strategies to make rank-order program selections. Shachar and Anand (1998) examine network show tune-in ads aired in November, 1995, and find that networks are nearly optimal in the number of ads they choose to air. In contrast, Currim and Shoemaker (1990) find that in 1983, across 22 product categories, advertising placements were not targeted to the heavy users of the brand or

\(^1\)For reviews see Cannon and Seamons (1995) and Cannon, Smith and Williams (2007). Abe (1997) is one of the most recent models, but uses single-source data from the 1980’s that is not available for current political settings.
product category. We add to this literature both by providing new optimal and heuristic approaches to targeting and by evaluating current strategies against these optimal ones.

Historically, the main approach to targeting, called “double-sourcing,” was to first identify demographic segments that have high indexes for purchase, and then pick shows that index high on those targeted demographics (Cannon, Smith and Williams 2007). With this method, the decisions are based on cross-tab summaries from two separate data sources. Efforts to collect all the important variables in one “single-sourced” dataset have faced extremely high costs, sample size and representativeness problems, and resistance from the media planning process (Assael and Poltrack 1996, Phelps 1987). Our approach requires more detailed (but readily available) information than the typical double-sourced data, yet provides similar benefits to costly single-sourced data. With this imputed single-source data, we optimize the allocation of advertisements across programs (i.e., targeting decisions), going well beyond rank order program selections. Further, we examine how much of the benefit from optimally allocating can be obtained using simple heuristics.

3 Model

We consider the voting and viewing decisions of potential voters in congressional elections in order to model how advertising by the candidates influences voting behavior (turnout and candidate choice). Conceptually, individuals respond to advertising exposure, not advertising levels, but we do not observe both exposure and voting choices for the same individuals. Hence, we develop a model of viewing behavior that allows us to impute which advertisements an individual sees. The purpose of this viewing model is to support estimating the effect of advertising exposure and to allow us to simulate the effect of counterfactual advertising allocations on voting behavior.

There are \( J \) elections, indexed by \( j \), each featuring a Democratic and Republican candidate, whose party is denoted by \( k \in \{ D, R \} \), and no major third party challenger. There are \( M \) media markets, indexed by \( m \). Congressional districts may overlap multiple media markets and media markets may overlap multiple congressional districts. Within each media market, Democratic and Republican candidates are able to run ads on \( P \) different television programs, indexed by \( p \). Let \( c \) denote the index for opportunities to advertise on a program and let \( a_{k,m,j,p,c} \) denote whether an ad

\[\text{In 2008 a joint Nielsen- Arbitron effort to develop single-sourced data was cancelled after cost and sample issues became insurmountable (Assael and Poltrack 1996).}\]
was aired by the candidate of party $k$ in media market $m$ and congressional district $j$ on program $p$ at occasion $c$. We let $a_{k,m,j,p,c}$ denote whether an ad was run by non-congressional candidates of party $k$ in media market $m$ on program $p$ at occasion $c$.

3.1 Individual Viewing Decisions and Exposure to Advertisements

For each program $p$ and each viewing occasion $c$, individual $n$ chooses between watching, $w_{n,p,c} = 1$, and not watching, $w_{n,p,c} = 0$. The probability that the individual watches is

$$\Pr(w_{n,p,c} = 1| x_n) = \Lambda(\gamma_p' x_n)$$

where $\gamma_p$ is a vector of parameters specific to program $p$, $x_n$ are the observable characteristics of individual $n$, and $\Lambda(z) = \frac{e^z}{1+e^z}$ is the logistic cdf.

Let $a_{k,m,p,c}$ denote the set of all ads by party $k$ broadcast in media market $m$ on program $p$ at occasion $c$,

$$a_{k,m,p,c} = \sum_{j=0}^{J} a_{k,m,j,p,c}$$

Let $m_n$ denote the media market in which individual $n$ lives. The exposure of individual $n$ to ads by party $k$ is given by,

$$e_{n,k} = \sum_{p=1}^{P} \sum_{c=1}^{C_p} w_{n,p,c} a_{k,m_n,p,c}$$

where $C_p$ is the total possible advertising opportunities for program $p$. Note that an individual’s exposure is an interaction of whether an ad was aired and whether the individual watched, summed over all opportunities to air ads. These opportunities include viewing ads aired by other congressional candidates in the same media market and ads that are run by candidates for other offices in the same media market. We define exposure in this way to allow ads by these other candidates to affect congressional voting behavior. Considering all exposures is particularly important for turnout decisions, which we expect are influenced by advertisements from all races rather than advertisements for just one (perhaps less important) race. In our empirical analysis we investigate the robustness of our results to separating exposure to congressional ads and exposure to other ads.
3.2 Individual Voting Decisions

Individuals choose between $y_n = 0$ (not voting), $y_n = 1$ (voting for the Democratic candidate), and $y_n = 2$ (voting for the Republican candidate). We define,

$$t^*_n = \xi_{t,j,m} + \beta_t' x_n + \alpha_t e_{n,T} + \varepsilon_t^n$$

where $x_n$ denotes the observable characteristics of individual $n$, $\beta_t$ and $\beta_v$ specify the effect of observable characteristics on turnout and candidate choice, $e_{n,D}$, $e_{n,R}$, and $e_{n,O}$ denote the exposure of individual $n$ to ads by the Democratic, Republican, and other candidates, $e_{n,T} = e_{n,D} + e_{n,R} + e_{n,O}$ denotes the total exposure of individual $n$ to ads, and $\alpha_t$ and $\alpha_v$ specify the effects of advertising exposure on turnout and candidate choice. We assume that $\varepsilon_t^n$ and $\varepsilon_v^n$ are normally distributed with variance 1 and correlation $\rho$. We also include fixed effects, $\xi_{t,j,m}$ and $\xi_{v,j,m}$, that are common to all individuals living in media market $m$ and congressional district $j$. We assume that an individual turns out if $t^*_n \geq 0$ and that conditional on turning out, the individual votes for the Republican candidate if $v^*_n \geq 0$. For notational simplicity, we set $\bar{t}^*_n = \xi_{t,j,m} + \beta_t' x_n + \alpha_t e_{n,T}$ and $\bar{v}^*_n = \xi_{v,j,m} + \beta_v' x_n + \alpha_v (e_{n,R} - e_{n,D})$. We present the probabilities of the voting decisions as

$$\Pr(y_n = 0 | \bar{t}^*_n) = \Phi(-\bar{t}^*_n)$$

$$\Pr(y_n = 1 | \bar{t}^*_n, \bar{v}^*_n) = \Phi(\bar{t}^*_n) - \Phi_2(\bar{t}^*_n, \bar{v}^*_n; \rho)$$

$$\Pr(y_n = 2 | \bar{t}^*_n, \bar{v}^*_n) = \Phi_2(\bar{t}^*_n, \bar{v}^*_n; \rho)$$

where $\Phi_2$ denotes the bivariate standard normal cdf with correlation $\rho$.

We allow total advertising exposure (including exposure to ads by groups that support third party candidates) to influence turnout decisions. This specification is most consistent with the

\footnote{Note that in practice we place some constraints on the coefficients that reduce the set of variables in $x_n$ and limit the functional form to linear effects.}
notion that television advertisements remind potential voters that an election is approaching and activate their interest in the election.\(^4\) We allow the difference in advertising exposure to influence candidate choice.

The model we employ is sometimes referred to as the “nested” or “sequential” probit, because the choice of candidate is nested under the choice to turnout. Unlike in a multinomial choice model, the utility for abstention is not compared to the maximum of the utilities for choosing the Democratic candidate and the Republican candidate. In a multinomial choice model with the typical brand-focused utility set-up, a single coefficient (say, \(\alpha\)) would represent the effect of advertising exposure on turnout and candidate choice and a positive coefficient would imply both mobilization and persuasion effects, confounding the two effects. In our framework, the parameter \(\alpha_t\) uniquely captures the effect of advertising exposure on turnout and \(\alpha_t = 0\) implies that there is no effect on turnout. Similarly, the parameter \(\alpha_v\) uniquely captures the effect on candidate choice and \(\alpha_v = 0\) implies that there is no effect on candidate choice.

Our decision to depart from the standard multinomial choice framework is consistent with work in both political science (Freedman, Franz and Goldstein 2004b, Huber and Arceneaux 2007, Gerber et al. 2010) and marketing (Che, Iyer and Shanmugam 2007, Gordon and Hartmann 2011). In all cases, the authors realized the restrictiveness of allowing a single coefficient to determine the effect of advertising on turnout and candidate choice; however, as outlined above, our approach does differ from the approaches adopted by Che, Iyer and Shanmugam and Gordon and Hartmann.

4 Data

We use a number of distinct datasets for our empirical analysis. In this section we describe each of these including the data on program viewing, potential voters, advertising, advertising costs, and aggregate turnout and vote shares. We conclude the section with a discussion of the final sample.

\(^4\)We considered some alternative specifications motivated by alternative notions of the function of TV advertising. We also estimated models in which the turnout effect was allowed to differ across the two parties. These differences were not statistically significant. Further, as the literature also suggests there are differences in the effect of positive and negative advertising on turnout and candidate choice (Che, Iyer and Shanmugam 2007, Lovett and Shachar 2011), we tested such specifications by splitting the advertising by these designations. Again, we did not find statistically significant differences in the effects.
4.1 Program Viewing Data

The Simmons National Consumer Survey is a national survey that collects information on consumer behavior, attitudes, demographics, and, essential to our study, political variables. We use a subsample of the survey corresponding to voting age American adults (N = 24,868). We use only the relevant set of variables that are also available in the National Annenberg Election Survey. We require the variables to be present in both datasets so that we can link the two datasets in our estimation procedure. Specifically, we include demographic variables indicating gender, race, age, education level, marital status, employment status, income level, previous service in the armed forces, and region as well as political variables indicating self-reported party identification, political ideology, and voter registration status. The individual-level Simmons data is not available. We instead access the data through a computer program, which provides cross-tabulations. From this dataset, we collected cross-tabulations of program viewing percentages and frequencies by each of the demographic and political variable levels. This data includes 602 programs across cable, network weekly, network weekday, and sports programs.

To give an initial view on this data, we consider the audience variation across programs. In order to target advertising, this variation needs to be sufficiently large along two key dimensions—the probability of turning out and the probability of voting for the Republican candidate. Figure 1 contains scatter plots of the television programs in our dataset, according to the portion of viewers registered to vote (on the Y-axis) and the percentage of conservative minus the percentage of liberal identifiers among the program viewers (on the X-axis). As apparent from the figure, the programs exhibit considerable variation on both dimensions. The percent of registered voters ranges from 46% for Run of the House (WB) to 89% for Meet the Press (NBC). Net conservative identifiers ranges from -25% for Now with Bill Moyers (PBS) to 40% for Sue Thomas FB:Eye (PAX).

4.2 Potential Voter Data

From the 2004 National Annenberg Election Survey (NAES), we obtain a sample of potential voters. This study incorporates a number of different data components, of which we employ two. The first is the election panel (N = 8,665), which provides individual-level data on the voting behavior (i.e., turnout and candidate choice) and the individual-level characteristics found in the Simmons data (i.e., the demographic and political variables listed above). The second dataset is used as a means
of linking aggregate and individual level data in both estimating the viewing and voting models. This data set is the rolling cross section ($N = 81,423$) component of the NAES, which contains a very large sample of the demographic and political variables also available in the Simmons data (and listed above), but does not contain the actual voting behavior of respondents.

### 4.3 Advertising Data

Our advertising data come from the 2004 Wisconsin Advertising Project (WAP) (Goldstein and Rivlin 2005). The data cover the 100 largest media markets in the United States, including about 80% of the U.S. population and both network and cable ads. The data provides us with a detailed coding of the ads including when they aired, where they aired, who aired them, and on what program they aired, as well as an estimate of the cost of the ad. In our analysis, we focus on the general election campaign and consider all ads run between Labor Day and Election Day.

One complication of the separate data sources for advertising and program viewing is that the program titles for the WAP data do not correspond exactly with the titles for the Simmons data. As a result, we created a link-list to match the 5,100 program titles on which ads aired in the WAP data with the program codes in the Simmons data. While many of the additional titles were easy to link such as various misspelled program names and listing each movie instead of a “movie night” listing, some program titles could not be linked. In the end, we were able to identify matching programs that accounted for over 95% of the aired congressional ads.

### 4.4 Advertising Cost Data

We also collected data on the forecasted cost of ads. Television ad prices are negotiated between ad buyers and television stations or cable providers on a case by case basis. However, a starting point of these negotiations are cost forecasts published by SQAD Inc. These estimates report the average cost per rating point of running ads for each media market and each day part.\(^5\) These estimates may not perfectly reflect the costs the campaigns actually pay for ads, but after talking with ad buyers, we believe these estimates accurately reflect the costs the campaigns believe they will pay at the time they make their advertising decisions.

Table 1 presents the average cost per thousand viewers and the ad spending (in thousands of

\(^5\)A ratings point is defined as a percentage point of the population who watch a program.
Figure 1: Characteristics of the Average Viewer – In the left panel, plus signs indicate the characteristics of the average viewer of the show. The size of the plus sign is proportional to the rating of the show. In the right panel, we included a selected sample for illustration purposes. The size of the text is proportional to the rating of the show.

GRPs) for Democratic and Republican congressional candidates by day part. We find substantial differences in the cost of reaching viewers by day part—prime time ads cost more than three times as much per viewer as early morning ads. This difference likely arises from the value commercial advertisers place on reaching different demographic groups. We note that actual ad placements are generally consistent with seeking out these cheaper dayparts—the candidates purchase many GRPs during the early morning and daytime and purchase relatively few during prime time.

<table>
<thead>
<tr>
<th>Day Part</th>
<th>Cost per Thousand Viewers</th>
<th>Congressional GRPs (1000s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Morning (12:30pm-9am)</td>
<td>$6.02</td>
<td>116.8</td>
</tr>
<tr>
<td>Day Time (9am-4pm)</td>
<td>$6.26</td>
<td>72.7</td>
</tr>
<tr>
<td>Early Fringe (4pm-6pm)</td>
<td>$7.95</td>
<td>26.0</td>
</tr>
<tr>
<td>Early News (6pm-7pm)</td>
<td>$9.19</td>
<td>58.1</td>
</tr>
<tr>
<td>Prime Access (7pm-8pm)</td>
<td>$11.32</td>
<td>50.5</td>
</tr>
<tr>
<td>Prime Time (8pm-11pm)</td>
<td>$18.73</td>
<td>31.9</td>
</tr>
<tr>
<td>Late News (11pm-1:30pm)</td>
<td>$13.75</td>
<td>23.0</td>
</tr>
<tr>
<td>Late Fringe (11:30pm-12:30pm)</td>
<td>$9.74</td>
<td>15.6</td>
</tr>
</tbody>
</table>

Table 1: Summary Statistics for Day Parts.

---

*A GRP, or gross ratings point, is a measure of the size of the audience reach by a media campaign. A GRP is the product of the frequency of airing and the size of the audience. For example, 10 GRPs would correspond to 10% of viewers seeing an ad once, or 5% of viewers seeing an ad twice.*
4.5 Aggregate Market and District Level Data

Finally, we sought to obtain data on the aggregate voting behavior for the units constructed from intersecting congressional districts and media markets (henceforth, market-district). We purchased proprietary data from uselectionatlas.org on the number of votes for the Democratic and Republican congressional candidates at the county-congressional district level. To calculate voter turnout, we used census data on the voting age population by county and congressional district. We aggregated both the population and the voting data from the county-level to market-districts. Finally, we defined congressional turnout as the total number of congressional votes in the district divided by the voting age population.

4.6 Final Sample

Because of missing data, we are not able to use all congressional districts in our final sample. First, the 2004 WAP data only covers the 100 largest media markets in the United States. As a result, we exclude any congressional districts that intersect with unobserved media markets. Second, the 2004 NAES did not sample in Alaska and Hawaii. Hence, we exclude all districts in those states. Finally, we exclude races where the losing major party candidate received less than 20% of the two party vote share. While this shrinks our sample size, we drop these congressional districts to guard against any biases arising from the potentially different motivations candidates and voters have in races that are uncompetitive. Our final sample consists of 219 congressional districts, 36 of which are served by multiple media markets.

5 Estimation and Identification

5.1 Estimating the Show Viewership Model

To estimate the parameters that characterize show viewership, $\gamma_p$, we employ a minimum distance estimator (Newey and McFadden 1994). Define $x_{n,i}$ as one arbitrary characteristic $i$ from those contained in $x_n$. Based on the Simmons data, we can compute quantities of the form $\hat{\Pr}(w_{n,p,c} = 1|x_{n,i} = q)$, where $i$ denotes the variable being conditioned on and $q$ denotes a value that the variable takes.\footnote{In Technical Appendix A.1, we discuss how $\hat{\Pr}(w_{n,p,c} = 1|x_{n,i} = q)$ is computed.} For example, this could be the percentage of very liberal viewers in the Simmons
sample that watched *60 Minutes* on a given viewing occasion. We would like to compare these to their theoretical counterparts, which are given by,

\[
\Pr(w_{n,p,c} = 1|x_{n,i} = q) = \frac{\sum_{x:x_i=q} \Pr(x) \Lambda(\gamma_p' x)}{\sum_{x} \Pr(x)} \quad (9)
\]

Here, \( \Pr(x) \) denotes the probability mass function of the demographic characteristics. This distribution is not directly observed in our television viewership data, but we observe a large sample of \( x_n \) in the NAES rolling cross section. We use the empirical distribution to obtain an estimate of the (population) distribution of \( x_n \).

Noting that \( N_{rcs} \) is the sample size from the NAES rolling cross section, we estimate the (model-based) probability that an individual \( n \) with characteristic \( i \) at a value \( q \) watches program \( p \) on occasion \( c \),

\[
\Pr(w_{n,p,c} = 1|x_{n,i} = q) \approx \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} 1\{x_{n,i} = q\} \Lambda(\gamma_p' x_n) \quad (10)
\]

Based on this, we can form the following moment conditions,

\[
\hat{h}_{p,i,q}(\gamma_p) = \hat{\Pr}(w_{n,p,c} = 1|x_{n,i} = q) - \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} 1\{x_{n,i} = q\} \Lambda(\gamma_p' x_n) \quad (11)
\]

For each \( i \), we select moments for all of the subgroups, defined by the set of values \( q \) that each \( x_{n,i} \) can take.\(^8\) This corresponds to using the proportion of each subgroup that watches each show as moments in our estimation. Let \( \hat{h}_p(\gamma_p) \) denote the vector of moments for program \( p \). By choosing these moments, we have that \( \hat{h}_p(\gamma_p) \xrightarrow{prob} 0 \) if and only if \( \gamma_p = \gamma_{p,0} \) (where \( \gamma_{p,0} \) denotes the true parameter vector characterizing the data generating process for one program), so the moments will define a minimum distance estimator. We therefore have an exactly identified minimum distance estimator and we estimate \( \gamma_{p,0} \) by solving the nonlinear system \( \hat{h}_p(\gamma_p) = 0 \).\(^9\)

Note that the estimator we employ here is not a standard method of moments estimator for at least two reasons. First, the moment conditions are formed based on two different samples—the Simmons data is used to compute \( \hat{\Pr}(w_{n,p,c} = 1|x_{n,i} = q) \) while the NAES rolling cross section is used to compute \( \hat{\Pr}(x) \). Second, the moment conditions contain sums of the data in both the

---

\(^8\)For example, if \( x_{n,i} \) denotes gender, then \( x_{n,i} \) can take on the values 1 (for females) and 0 (for males).

\(^9\)In principle we could increase the efficiency by adding additional moment conditions formed from higher-order cross-tabulations of show viewership and the demographic variables. We also could use these higher-order moments to identify interactions in the viewing functions.
numerator and the denominator. Because our estimator is not a standard method of moments estimator, we rely on specialized formulas for deriving the asymptotic distribution. We derive the large sample distribution of the minimum distance estimator in Technical Appendix A.2 and report the estimation results in Technical Appendix A.3.

5.2 Estimating the Voter Decision Model

We are interested in estimating the vector of parameters, $\theta = (\alpha_v, \alpha_t, \beta_v, \beta_t, \rho)$, along with the vectors of fixed effects, $\xi^t$ and $\xi^v$. We write the log-likelihood as,

\[ l(\theta, \xi^t, \xi^v) = \sum_{n=1}^{N} \left\{ y_n = 0 \right\} \log Pr(y_n = 0|\theta, \xi^t, \xi^v) + 1 \left\{ y_n = 1 \right\} \log Pr(y_n = 1|\theta, \xi^t, \xi^v) + 1 \left\{ y_n = 2 \right\} \log Pr(y_n = 2|\theta, \xi^t, \xi^v) \] (12)

If we allow for a correlation between the error terms of the turnout and candidate choice equations, then the expressions $Pr(y_n = 1|\theta, \xi^t, \xi^v)$ and $Pr(y_n = 2|\theta, \xi^t, \xi^v)$ involve computing rectangles of the bivariate normal distribution. We compute these integrals using the GHK simulator (Geweke, Keane and Runkle 1994).

While our sample size is large, it is not large enough to allow us to accurately estimate the fixed effects from the individual level data. Instead, we apply a share inversion approach (Berry 1994, Berry, Levinsohn and Pakes 1995, Chintagunta and Dubé 2005), estimating the fixed effects by matching the predictions of the model to the observed turnout and Republican vote shares in each district. For district $j$ and media market $m$, we denote the turnout share by $s^t_{j,m}(\theta, \xi^t_{j,m})$ and Republican vote share by $s^v_{j,m}(\theta, \xi^t_{j,m}, \xi^v_{j,m})$. According to the model, we have,

\[ s^t_{j,m}(\theta, \xi^t_{j,m}) = \frac{1}{|N_{j,m}|} \sum_{n \in N_{j,m}} \Phi(t^*_n) \] (13)

\[ s^v_{j,m}(\theta, \xi^t_{j,m}, \xi^v_{j,m}) = \frac{\sum_{n \in N_{j,m}} \Phi_2(t^*_n, v^*_n)}{\sum_{n \in N_{j,m}} \Phi(t^*_n)} \] (14)

where $N_{j,m}$ denotes the set of NAES rolling cross section respondents living in congressional district.
and media market \(m\). In our estimation procedure, we match the predicted shares given in equations (13) and (14) to the observed turnout share, \(\hat{s}^t_{j,m}\), and the observed Republican vote share, \(\hat{s}^v_{j,m}\). Specifically, we have the constraints,

\[
s^t_{j,m}(\theta, \xi^t_{j,m}) = \hat{s}^t_{j,m} \tag{15}
\]

\[
s^v_{j,m}(\theta, \xi^t_{j,m}, \xi^v_{j,m}) = \hat{s}^v_{j,m} \tag{16}
\]

We thus estimate the model parameters \((\theta, \xi^t, \xi^v)\) via constrained MLE (Chintagunta and Dubé 2005), maximizing (12) over \((\theta, \xi^t, \xi^v)\), subject to constraints (15) and (16) for all \(j\) and \(m\).

Following Berry (1994) and Berry, Levinsohn and Pakes (1995), we apply a nested fixed point approach to solve the constrained optimization problem. To solve the share inversion sub-problems, we depart from Berry (1994) and Berry, Levinsohn and Pakes (1995). Instead, for a given value of \(\theta\), the only unknown value in (15) is \(\xi^t_{j,m}\). We use a one-dimensional zero-finder to solve for each \(\xi^t_{j,m}\) imposing the constraint in (15). With the value of \(\xi^t_{j,m}\) known, we again use a one-dimensional zero-finder to calculate \(\xi^v_{j,m}\) as the solution to (16). In Technical Appendix Section A.4, we prove that for each \(\theta\) there exists a unique solution to the share equations.

### 5.3 Multiple Imputation

An additional complication with the procedure above is that we do not actually observe \(e_{n,k}\), which are needed in the likelihood. Instead, we have the probability model \(\Pr(w_{n,p,c} = 1 | x_n) = \Lambda(\gamma_p x_n)\), where \(\gamma_p\) is a parameter characterizing the viewing decisions for program \(p\), which was estimated with the method described in Section 5.1. This model allows us to simulate \(w_{n,p,c}\) using independent draws from the Bernoulli(\(\Lambda(\gamma_p x_n)\)) distribution. We then calculate exposure using (3).

We follow the multiple imputation literature (Rubin 1987, Schafer 1997) and estimate the model based on 5 draws for \(e_{n,k}\). Repeating this process 5 times allows us to properly account for the uncertainty in the imputation model and also produces estimates that are more efficient than one would obtain with a single draw (Schafer 1997). We then perform the entire constrained maximum likelihood estimation on each of the 5 replicated data sets separately. We follow Rubin’s (1987)
approach for calculating the estimates and standard errors from the results of the 5 imputed datasets (see Technical Appendix A.5 for further details).

Because we use imputation to link multiple data sets, we review the role of each dataset. In the first stage, we estimate the parameters $\gamma_p$ using the Simmons and NAES rolling cross section data (RCS). We use these estimates and the advertising data to impute advertising exposure for the NAES election panel data (using the individual characteristics $x_n$ from that dataset). As a result, we have the “fused” individual-level data for our estimation. In addition, to estimate the second stage, we need to match the observed shares to the model predicted shares. For this, we use the NAES RCS data since the sample size is roughly 10 times larger. To use the RCS data we also need to impute advertising exposure, which we do in the same manner only using the RCS values of $x_n$. Finally, for all the counterfactual simulations we report in the paper, we use the imputed advertising exposure and calculate the latent variables (and calculate the individual probabilities) for the observations in the NAES RCS sample.

5.4 Identification

Our primary focus is on the identification of the advertising effects. As is typical in estimating advertising effects from non-experimental data (Dubé, Hitsch and Manchanda 2005, Gordon and Hartmann 2011), aggregate unobserved characteristics that affect voting decisions could also influence campaign advertising decisions and, hence, correlate with the observed advertising levels. In our case, these unobserved characteristics could be any unobserved tendency of the area to more likely turnout or vote for the Republican (instead of the Democrat). In particular, these characteristics might reflect some (known to campaigns) tendency such as past higher voting percentages. In our setting, these endogeneity biases are likely to differ depending on the “true” advertising effect. To understand these biases we consider the extreme cases when advertising has only a candidate choice or turnout effect.

Consider first the case that advertising has a true effect on candidate choice (and not on turnout). In this case, advertising in media markets with higher turnout would generate more votes. As a result, candidates (candidates for President, Governor, and Senate as well as House candidates with multiple media markets) would advertise more in high turnout media markets. This would induce a positive correlation between advertising and turnout that would bias our estimates.
of the turnout effect upward. The trade-offs between different campaign expenditures can also lead to bias. Since GOTV efforts mobilize voters to turnout (Gerber and Green 2000), the incentive to spend on GOTV efforts increases as the proportion of core supporters increases. Hence, in districts with higher proportions of core supporters, the relative incentive to spend on television advertising is lower, which could bias our estimates of the candidate choice effect of advertising downward.

Alternatively, if the true effect of advertising is on turnout (and not on candidate choice), candidates would target media markets with large proportions of core supporters to air more ads. This would induce an inverted U relationship between Republican vote shares and advertising that could bias our estimates in favor of finding a candidate choice effect. In addition, markets and districts with lower advertising costs (and/or higher costs for GOTV efforts) are likely to have more advertisements, but less GOTV efforts, and vice versa for markets with higher advertising costs (and/or lower costs of GOTV efforts). This induces a negative correlation between advertising and turnout.

In both cases the endogeneity bias due to the unobserved characteristics will lead us to precisely the wrong conclusion. If advertising in fact has a candidate choice effect, we would find a stronger turnout effect than exists and would underestimate the strength of the candidate choice effect. If advertising in fact has a turnout effect, we would find a stronger candidate choice effect than exists and would underestimate the strength of the turnout effect.

Our approach to solving this endogeneity problem builds on a difference-in-differences logic. Equations (4) and (5) include (i) main effects for the individual-level characteristics (demographics and political variables), (ii) fixed effects at the market-district level, and (iii) advertising exposure. The individual-level characteristics control for well-known effects at the individual level (e.g. Wolfinger and Rosenstone (1980)) and the fixed effects control for market-district level aggregate unobserved characteristics (or candidate actions). The fixed effects also absorb the effect of advertising levels (or overall spending). However, an ad can only affect the voting behavior of individuals who see the ad. We compute advertising exposure as the interaction of advertising decisions and individual-level characteristics, as seen in equation (3). Thus, conceptually, we estimate the advertising exposure effect as the effect of the interaction between individual-level characteris-

---

11 As a result, these fixed effects have the added advantage that we do not have to worry about specific variables to include as congressional district and media market level controls. Such controls include the incumbent party, the quality of the challenger, money spent on GOTV efforts, the number of presidential visits, etc. Such variables will already be captured by the fixed effects.
tics and advertising levels. Identification comes from the differences across market-districts in the correlation between demographics and voting behavior. Thus, our key identifying assumption is that differences in the direct effect of individual-level characteristics across market-districts are uncorrelated with advertising.

6 Estimation Results

6.1 Estimates of Voting Model

In Table 2, we report estimates of the effect of television advertising on voter turnout and candidate choice. We first report results without fixed effects. In column (1), we find that ad exposure increases turnout and that advantages in advertising exposure lead to greater relative vote share. The demographic and political variables generally have sensible coefficients. Consistent with expectations and the literature (Wolfinger and Rosenstone 1980), registered, more educated, and older voters are more likely to vote and blacks are less likely to vote. Similarly, blacks and younger voters tend to vote more Democratic, while liberal voters are more likely to vote Democratic and conservative voters are more likely to vote Republican.

The results reported in column (1) do not account for endogeneity in ad spending. As discussed in Section 5.4, we expect endogeneity biases could arise. We report the results controlling for fixed effects in column (2). We find that the effect of ad exposure on turnout is smaller and no longer statistically significant. Further, we find that ad exposure affects candidate choice and that the estimated effect is approximately two times larger than the estimated effect without fixed effects. These results are consistent with the biases we expect when not controlling for unobserved characteristics under a true candidate choice effect. This also suggests that the estimates in column (1) suffer from endogeneity bias. Without controlling for fixed effects, one would falsely conclude that advertising has a larger, significant effect on turnout and a smaller effect on candidate choice. We also note that the standard errors increase in column (2) largely because the fixed effects explain the majority of the variation in advertising exposure.

In columns (3) and (4), we present the full model that includes correlation in the unobserved

\footnote{In fact, we further restrict that the individual level characteristics component of this interaction must be “mediated” by the television viewing decisions of the individuals. In other words, we attribute to the advertising exposure effect only the part of the correlation between advertising and individual-level characteristics that also correlates with viewing decisions.}
idiosyncratic voting decision errors. We find that this correlation is small and insignificant, that none of the significant qualitative results change as a result of this inclusion, and that the significance levels for the variables of interest are essentially the same. \(^{13}\) Since performing counterfactual simulations with the correlation is more computationally demanding, for all future calculations and predictions, we use the estimates from the model in column (2).

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Turnout:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Ad Exposures</td>
<td>0.115*</td>
<td>0.054</td>
<td>0.150**</td>
<td>-0.047</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.163)</td>
<td>(0.059)</td>
<td>(0.145)</td>
</tr>
<tr>
<td>Registered</td>
<td>1.473***</td>
<td>1.516***</td>
<td>1.511***</td>
<td>1.582***</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.072)</td>
<td>(0.068)</td>
<td>(0.073)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.278**</td>
<td>-0.265*</td>
<td>-0.314***</td>
<td>-0.236*</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.108)</td>
<td>(0.098)</td>
<td>(0.111)</td>
</tr>
<tr>
<td>Education</td>
<td>0.202***</td>
<td>0.193***</td>
<td>0.210***</td>
<td>0.190***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Age</td>
<td>0.141***</td>
<td>0.146***</td>
<td>0.147***</td>
<td>0.153***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.018)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.020</td>
<td>0.007</td>
<td>-0.034</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.055)</td>
<td>(0.053)</td>
<td>(0.056)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Voting:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advantage in Ad Exposures</td>
<td>0.741**</td>
<td>1.315*</td>
<td>0.730**</td>
<td>1.146*</td>
</tr>
<tr>
<td></td>
<td>(0.240)</td>
<td>(0.663)</td>
<td>(0.235)</td>
<td>(0.519)</td>
</tr>
<tr>
<td>Ideology</td>
<td>0.822***</td>
<td>0.798***</td>
<td>0.804***</td>
<td>0.772***</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.033)</td>
<td>(0.804)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Black</td>
<td>-0.939***</td>
<td>-0.929***</td>
<td>-0.954***</td>
<td>-0.938***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
<td>(0.141)</td>
<td>(0.142)</td>
<td>(0.137)</td>
</tr>
<tr>
<td>Education</td>
<td>0.005</td>
<td>0.006</td>
<td>0.010</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.021)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.088***</td>
<td>-0.090***</td>
<td>-0.083***</td>
<td>-0.093***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.090+</td>
<td>-0.060</td>
<td>-0.090+</td>
<td>-0.048</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.052)</td>
<td>(0.050)</td>
<td>(0.049)</td>
</tr>
<tr>
<td>( \rho )</td>
<td>-</td>
<td>-</td>
<td>0.027</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.101)</td>
<td>(0.096)</td>
</tr>
<tr>
<td><strong>Fixed Effects?</strong></td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>3,436</td>
<td>3,436</td>
<td>3,436</td>
<td>3,436</td>
</tr>
</tbody>
</table>

Table 2: Main Estimation Results – One star indicates statistical significance at the 5% level. Two stars indicates statistical significance at the 1% level. Three stars indicates statistical significance at the 0.1% level. A plus sign indicates statistical significance at the 10% level.

\(^{13}\) We also ran a number of robustness tests to ensure against alternative ideas about the effect of advertising, including whether the turnout effect differs for the parties, whether positive and negative ads have different effects, whether prime time ads or news ads have different effects, whether congressional ads have different effects from ads for other races, and whether the effect of ad exposure has curvature beyond that implicit in the probit function. In all cases, our estimates did not provide statistical support for the added complications. To be clear, we do not take these robustness checks as implying these additional effects are irrelevant, but rather that our estimation approach cannot tease apart these finer distinctions. These results are available upon request.
6.2 Advertising Exposure Effect Sizes and Elasticities

We now provide a sense of the size of the advertising effects we presented above by calculating marginal effects of advertising and the average elasticity. First, we calculate the average effect of a uniform increase in advertising exposure. We hold all variables at observed values and use the estimated parameter values to predict a baseline for turnout and Republican vote share.\textsuperscript{14} We then increase and decrease Democratic exposure for every observation by one standard deviation (45 ad exposures) and observe the effect on turnout and Republican vote share. We perform the same calculation for Republican candidates. The results are reported in Table 3. We find that for both parties, a change in exposure from one standard deviation below the baseline to one standard deviation above leads to an increase in that party’s vote share of 3.5%. While not a huge effect, it is substantial enough to swing a close election.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Turnout</th>
<th>Rep. Vote Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>54.5%</td>
<td>49.2%</td>
</tr>
<tr>
<td>Dem. Exposure Increased (One Sd.)</td>
<td>54.6%</td>
<td>47.4%</td>
</tr>
<tr>
<td>Rep. Exposure Increased (One Sd.)</td>
<td>54.6%</td>
<td>50.9%</td>
</tr>
<tr>
<td>Dem. Exp. Increased - Dem. Exp. Decreased</td>
<td>0.2%</td>
<td>-3.5%</td>
</tr>
<tr>
<td>Rep. Exp. Increased - Rep. Exp. Decreased</td>
<td>0.2%</td>
<td>3.5%</td>
</tr>
</tbody>
</table>

Table 3: Marginal Effects of Advertising Exposure.

Second, we calculate advertising exposure elasticities in order to provide unit-free estimates that can be compared to advertising effects in other contexts. We calculate the elasticity of the turnout rate with respect to total ad exposures and the elasticities of the Democratic and Republican vote shares (as a share of the two-party vote) with respect to Democratic and Republican ad exposure, respectively.\textsuperscript{15} We calculate the quantities for each individual in the NAES rolling cross section survey and report the average elasticities across respondents who received some advertising. The results for the turnout elasticity confirm that the effect is small in magnitude—doubling the total ad exposure of an individual will on average increase his probability of turning out by only 0.5%. The elasticities to Democratic and Republican exposures are roughly equal in magnitude and approximately 0.07. In addition, to compare to commercial contexts, we calculate the elasticity

\textsuperscript{14}By construction this baseline matches aggregate data since we constrained it to during estimation. We note that this baseline represents an average rate for voters residing in the 219 congressional districts in our sample. The turnout rate in our sample of congressional districts will be higher than the turnout rate in the excluded districts because we have excluded uncompetitive congressional races from the analysis.

\textsuperscript{15}The details of these calculations are given in Technical Appendix A.6.
for the total effect (i.e., not normalizing for relative share) of advertising. These estimates are (as expected) slightly higher and approximately 0.075.

These estimates compare favorably with those in the existing literature. While in commercial settings the average effects are typically higher—Assmus, Farley and Lehmann (1984) identified a grand mean of 0.2 and Sethuraman and Tellis (1991) and Tellis (1988) found an average of 0.11—considerable variation exists across products with new products typically near 0.3 and existing products closer to 0.01 (Assmus, Farley and Lehmann 1984). Since political parties are “existing” products with new faces, our lower elasticity is not too surprising. We also note that our estimate is somewhat larger than another recent study of political advertising (Gordon and Hartmann 2011).

<table>
<thead>
<tr>
<th>Elasticity</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elasticity of Turnout to Total Exposure</td>
<td>0.005</td>
</tr>
<tr>
<td>Elasticity of Democratic Relative Vote Share to Democratic Exposure</td>
<td>0.074</td>
</tr>
<tr>
<td>Elasticity of Republican Relative Vote Share to Republican Exposure</td>
<td>0.071</td>
</tr>
<tr>
<td>Elasticity of Democratic Vote Share to Democratic Exposure</td>
<td>0.077</td>
</tr>
<tr>
<td>Elasticity of Republican Vote Share to Republican Exposure</td>
<td>0.075</td>
</tr>
</tbody>
</table>

Table 4: Elasticities to Advertising Exposure.

7 Evaluating and Improving Campaign Decisions

In this section, we apply our framework to the study of actual campaign decisions and provide guidance as to how campaigns might improve their targeting decisions. First, we establish that the programs contain sufficient diversity along the necessary dimensions to support targeting. Second, we use a map that integrates the show dimensions and the actual candidate advertising to qualitatively describe whether campaigns act in accordance with the best generic strategy. Third, we evaluate candidate advertising strategies quantitatively by computing optimal targeting decisions. We evaluate how much improvement the candidates could expect to obtain from reallocating their existing funds to different television programs and we compare these optimal strategies to a set of heuristic strategies.

7.1 Targeting Opportunities

Our method allows us to simulate for each show the two critical dimensions for voting behavior—the predicted voter turnout rate and the predicted Republican support rate. To simulate from
these distributions we use the model estimates and the NAES data. Each individual in the NAES rolling cross-section sample is characterized by a vector of demographic and political variables $x_n$. We set ad exposure to zero and compute for each individual the probability of turning out using $T_n^* = \Phi(\xi_{j,m}^t + \hat{\beta}_t^j x_n)$ and the probability of voting Republican using $V_n^* = \Phi(\xi_{j,m}^v + \hat{\beta}_v^j x_n)$. We then allocate individuals to programs by drawing for each individual whether this person watches a given airing of program $p$ using $w_{n,p,c} \sim \text{Bernouli}(\Lambda(\gamma_{p}^t x_n))$. By considering all individuals that view the program, we estimate the joint distribution of voting tendencies for the program audience, $(T_n^*, V_n^*|w_{n,p,c} = 1)$ and calculate the average voting tendency of the viewers of each show as,

\[
T_{n,p}^* = \frac{\sum_{n=1}^{N_{rcs}} w_{n,p,c} T_n^*}{\sum_{n=1}^{N_{rcs}} w_{n,p,c}} \tag{17}
\]

\[
V_{n,p}^* = \frac{\sum_{n=1}^{N_{rcs}} w_{n,p,c} V_n^*}{\sum_{n=1}^{N_{rcs}} w_{n,p,c}} \tag{18}
\]
Figure 3: Advertising in Congressional Elections for Democratic and Republican Candidates by Day Part – Each letter represents a program with the size of the letter proportional to the number of GRPs run on the program. D’s represent ads by Democratic congressional candidates and R’s represent ads by Republican congressional candidates.
In Figure 2, we plot the locations of TV programs in the predicted turnout/candidate choice space (in Technical Appendix A.7, we present the full distribution for several shows). A large number of shows have middle of the road audiences and these shows vary greatly in their propensity to turnout. Many shows have Democratic leaning audiences, but far fewer lean Republican. Both of these tendencies in Figure 2 were apparent in Figure 1 (which is based directly on the Simmons data). This provides some face validity to our model predictions. However, important differences arise because Figure 2 considers the predictive ability of all viewer characteristics. Figure 2 depicts a cluster of shows with low turnout and heavy Democratic leaning that is absent in Figure 1. Even the middle of the road shows differ in important ways. For example, *American Idol* (Fox) and *Friends* (NBC) swap their relative vertical positions between the two figures due to *American Idol’s* larger proportion of black viewers and smaller proportion of college-educated viewers, both of which reduce turnout probabilities.

We use this map to characterize the options that candidates have for implementing generic strategies based on the turnout effect or the candidate choice effect of advertising. Candidates pursuing a candidate choice strategy have many options to target shows with middle of the road audiences that have a high likelihood of voting.

### 7.2 Observed Targeting Strategies

We now overlay candidate decisions on the show map to qualitatively evaluate how effectively campaigns target. In Figure 3 we report the advertising levels for the Democratic and Republican candidates across the 8 dayparts. The figure demonstrates that both parties heavily target ads at programs with high turnout rates and generally middle of the road audiences. This behavior is consistent with a belief that advertisements primarily affect candidate choice, but not with a belief that ads affect turnout. The figure also shows that the candidates target most of their advertising towards the cheapest day parts. Broadly speaking, candidates appear to follow the generic strategy suggested by the estimated advertising effects—to place ads on programs that are relatively cost effective and that have audiences with high turnout rates and large proportions of swing voters.
7.3 Best Responses Strategies

Consider any congressional district $j$ that is entirely contained in media market $m$. Let $B_{j,m}^k$ denote the budget of the candidate from party $k$, let $a_{k,j,m} = (a_{k,j,m,1}, \ldots, a_{k,j,m,P})$ denote the vector of ad placements for the candidate of party $k$ in congressional district $j$ and media market $m$, and let $c_m = (c_{m,1}, \ldots, c_{m,P})$ denote the corresponding costs per ad. We compute the budget for each candidate in each district by multiplying the number of ads the candidates ran on each program by the estimated price of the ads on that program. Given ad spending by the other candidates, the optimal targeting strategy for the Democratic and Republican candidates, respectively, is given by,

$$\min \left\{ a_{D,j,m} : a_{D,j,m} \geq 0, c_m^e a_{D,j,m} \leq B_{j,m}^D \right\} s_{j,m}^D(a_{D,j,m}, a_{R,j,m}) \quad (19)$$

$$\max \left\{ a_{R,j,m} : a_{R,j,m} \geq 0, c_m^e a_{R,j,m} \leq B_{j,m}^R \right\} s_{j,m}^R(a_{D,j,m}, a_{R,j,m}) \quad (20)$$

The nonlinear programming problems given in (19) and (20) can be solved numerically. While the number of variables and constraints is fairly large—candidates optimize their advertising over 602 television programs subject to 602 positivity constraints and a budget constraint—the constraints are linear and the constraint matrix sparse. We use a c++ constrained optimizer that takes advantage of the sparsity structure of the problem. With this solver, each optimization problem can be completed in minutes, allowing us to generate solutions to the hundreds of such problems necessary to produce the results in this subsection. The results we present are averages over the districts in our sample, although we note that only about a third of the districts see positive ad spending. We simulate the individual level responses in the same manner as in the construction of the program maps (see Section 7.1 for the details).

Congressional districts which overlap with multiple media markets involve a more complicated decision for the candidates, but we abstract away from this and focus on reallocating funds across television programs, holding constant the geographic allocation decisions of the candidates. We also abstract away from the fact that action by one candidate may motivate the other candidate to act as

---

16Here, $s_{j,m}^e(a_{D,j,m}, a_{R,j,m})$ incorporates ad spending by other candidates in the same media market, though for compactness we do not explicitly incorporate this in the notation.
well. We compute best-responses given the opponent’s current strategy and do not compute Nash Equilibrium strategies for two reasons. First, to solve for the Nash equilibrium strategies, we could optimize one candidate’s objective function subject to that candidate’s constraints and the set of constraints generated by the first-order conditions from the other candidate’s optimization problem. This new problem, however, would be a non-linear programming problem with complementarity constraints, a problem which is much more difficult to solve numerically. Second, while it is not entirely accurate to assume that one candidate switching to a best-response would be met with no reaction from his opponent, we argue that it would be more inaccurate to assume that a switch in strategy by one player would immediately lead his opponent to switch to the best-response. In practice, we expect to see some reaction—perhaps in the form of mimicking behavior by the opponent—but to the extent that a candidate switches to a new strategy right before the election, we expect to see a relatively small reaction by his opponent.

In Figure 4, we report the number of GRPs suggested by the best response strategies, for all shows in our sample, by day part, along with the probability of turning out and the probability of voting Republican for the average viewer of each show. A few patterns are immediately evident—a small number of early morning and daytime shows see the most ads. The optimal spending patterns for the parties are very similar. The optimal strategies involve spending on high turnout shows. The optimal candidate strategies in Figure 4 can be directly compared to the observed candidate strategies in Figure 3. The optimal spending patterns differ from the observed spending patterns in two important respects. First, while observed spending is concentrated on the two cheapest dayparts, the optimal strategy suggests a much higher degree of concentration. Second, while observed spending is concentrated on high turnout programs, the optimal strategy suggests the spending should be much more concentrated on a small number of high turnout shows.

The gains candidates could achieve from reallocating their existing funds optimally across television programs are reported in Table 5. The average Democratic vote share when the Republicans advertised at current levels, but the Democrat aired no ads is 48.7%. The Democratic Party achieved a vote share of 50.8% with its actual ad strategy. Thus, we estimate the benefit of those ads were a vote share increase of 2.1%. Similarly, the average Republican vote share under actual and no advertising is 52.9% and 50.5% respectively, leading to an increase of 2.4%. We note that actual vote shares for Democratic and Republican candidates are inconsistent in the table because
for each party, we averaged the results over districts that saw positive ad spending by that party and the parties did not advertise in exactly the same districts.

Had the Democratic candidates played a best response strategy, the average vote share would have been 53.2%, which represents a gain more than twice as large as they achieved. The Democratic Party could have swung 9 districts simply by reallocating their existing advertising budget to different television programs. Put a different way, by re-optimizing, the average Democratic congressional candidate would have increased the effectiveness of his plan by 117.8%. For Republican candidates, we obtained comparable results. Simply by reallocating their existing budget to a different set of television programs, Republican candidates could have more than doubled the effectiveness of their advertising campaigns and swung 8 districts their way.

Though we can solve the optimization problems given by (19) and (20), to do so requires both relatively demanding and advanced techniques (minimum distance estimation, multiple imputation, share inversion, nonlinear optimization) and costly data that is even more difficult to obtain before the election takes place. As a result, it is possible such a task may be beyond the resources and capabilities of some campaigns. To provide some sense of what simpler strategies might achieve, we also evaluate several heuristics. Evaluating these heuristics provides insight into the magnitude and nature of improvement we achieve with our optimization procedure. In addition, they provide campaigns with a means to avoid the more demanding analysis, while gaining meaningful improvement in practice.

Candidates in practice appear to already use a combination of heuristics. For instance, we find that advertising was targeted to local news programs, nightly news broadcasts, a selected set of talk and game shows, and programs with relatively high ratings. These allocations could arise from some combination of heuristics related to program type and rating, and such heuristics could roughly approximate the desired generic strategy. We build on this apparent use of heuristics by offering a new set of heuristics, motivated by observations about the computed best response functions. The best response functions suggest that spending should be targeted towards the two cheapest dayparts and that shows with very small audiences should be avoided. In addition, shows with high turnout rates and shows with many swing voters should be targeted, with the turnout rate being the more important factor.
Figure 4: Aggregated Best Response Strategy for Democratic and Republican Candidates—Each letter represents a program with the size of the letter proportional to the number of GRPs suggested by the best response strategies. D’s represent ads by Democratic congressional candidates and R’s represent ads by Republican congressional candidates.
The first heuristic we consider is to spend equally on the five or ten programs with the highest turnout rates. There are a large number of shows in our sample with few viewers—for example Kudlow and Kramer on CNBC has the highest turnout rate in our sample, but is viewed by less than 1% of the population. In fact, the top 10 high turnout programs are all viewed by less than 2% of the population. Moreover, all of these programs are cable news programs, which likely have overlapping viewers. Buying many ads on these programs does not achieve a high reach and cannot be an effective strategy since only a very small group of viewers will be seeing the ads. For this reason, we consider heuristics which only consider shows that are viewed by at least 2% of the population. By dividing their spending equally between the five such programs by turnout rate, the Democratic candidates could have improved their advertising effectiveness by 17.7%. Similarly, if the Democratic candidates would have spread their spending equally among the 10 shows with the highest turnout rates, the Democratic candidates could have increased their effectiveness by 14.1%.

<table>
<thead>
<tr>
<th></th>
<th>Democratic Candidates</th>
<th>Republican Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No Ads</strong></td>
<td>48.7%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Observed Strategy</strong></td>
<td>50.8%</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>Heuristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Top 5 by Turnout</td>
<td>51.3%</td>
<td>3</td>
</tr>
<tr>
<td>Top 10 by Turnout</td>
<td>51.2%</td>
<td>3</td>
</tr>
<tr>
<td>Top 5 in EM and DAY by Turnout</td>
<td>52.2%</td>
<td>5</td>
</tr>
<tr>
<td>Top 10 in EM and DAY by Turnout</td>
<td>52.1%</td>
<td>4</td>
</tr>
<tr>
<td>Top 5 in EM, DAY, and EN, by Turnout</td>
<td>51.9%</td>
<td>4</td>
</tr>
<tr>
<td>Top 10 in EM, DAY, and EN, by Turnout</td>
<td>51.9%</td>
<td>3</td>
</tr>
<tr>
<td>Top 5 Moderate in EM and DAY by Turnout</td>
<td>52.2%</td>
<td>6</td>
</tr>
<tr>
<td>Top 10 Moderate in EM and DAY by Turnout</td>
<td>52.1%</td>
<td>4</td>
</tr>
<tr>
<td>Top 5 in EM and DAY by Reg. Voters</td>
<td>52.1%</td>
<td>4</td>
</tr>
<tr>
<td>Top 10 in EM and DAY by Reg. Voters</td>
<td>51.8%</td>
<td>3</td>
</tr>
<tr>
<td>Top 5 Moderate in EM and DAY by Reg. Voters</td>
<td>52.1%</td>
<td>4</td>
</tr>
<tr>
<td>Top 10 Moderate in EM and DAY by Reg. Voters</td>
<td>51.8%</td>
<td>3</td>
</tr>
<tr>
<td><strong>Optimal Strategy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.2%</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 5: Evaluation of Optimal and Heuristic Strategies – Reports the average vote share under each strategy, the number of districts that flipped which party won (Party Flips), and the average percent improvement over the increase in vote share due to actual advertising levels (Avg. % Imprv.) for each of the parties.

The generic strategy also suggests focusing on the cheaper day parts, including the early morning
and daytime day parts. The candidates could instead spend equally on the top five (or top ten) high turnout programs among the two cheapest day parts. We find that by employing this heuristic, the candidates would improve effectiveness by 66.1% and 59.1% respectively. The turnout rate for each program, however, is not directly observed in the Simmons data, but was calculated based on our model. Suppose that instead candidates spend their money on the top 5 (or top 10) programs with the highest percentage of registered voters in the early morning or daytime day parts. In these cases, the improvements would be 62.1% and 42.0% respectively. This is not quite as good as the high turnout/low cost heuristic, but by applying this very simple heuristic, the candidates could achieve nearly two-thirds of the maximum possible improvement. We report additional results in Table 5. We note that the results for Republicans are quite similar. Overall, the results suggest that candidates could obtain significant improvements by employing very simple heuristics. Nonetheless, optimization could lead to further improvement and could allow flipping an extra two or three districts even compared to the best heuristics we examined.

We also examine the candidate and race characteristics that are associated with better or worse targeting. We evaluate the relative improvement that the optimal allocation generates over the observed allocation. Because we are interested in how well the ads are targeted, we consider the average, not the cumulative effect of targeting. Hence, we use relative improvement, which we calculate by dividing the improvement by the budget in terms of 1000 early morning GRPs, denoted $B_{jk,\text{GRPs}}$. We regress relative improvement on candidate and race characteristics, including indicators for each of the closeness levels in the pre-Labor Day Cook’s Political Report, an indicator for the frontrunner also taken from the Cook’s Political Report, an indicator for incumbency, and the number of terms served (coded as 0 for non-incumbents).

In Table 6, we report the results pooled across parties. The $R^2$ of the regression is 11%, so we do not explain much of the variance in relative improvement. We find a marginally significant negative effect for the closest races. Moreover, while not statistically distinguishable, it is encouraging that the closeness variables increase in magnitude in the expected direction. For example, the

\[ \text{Relative improvement for the Republican candidate is given by,} \]

\[ \frac{s_{v,j,m}(a_{D,j,m},a_{R,j,m}) - \hat{s}_{v,j,m}}{B_{jk,\text{GRPs}}} , \]

\[ \text{where} \ a_{R,j,m} \ \text{is the best response strategy of the Republican candidate,} \ a_{D,j,m} \ \text{is observed strategy of the Democratic candidate, and} \ \hat{s}_{v,j,m} \ \text{is the observed vote share for the Republican candidate. For the Democratic candidate, the relative improvement is given by} \]

\[ \frac{1-s_{v,j,m}(a_{D,j,m},a_{R,j,m}) - \hat{s}_{v,j,m}}{B_{jk,\text{GRPs}}} , \]

\[ \text{where} \ a_{D,j,m} \ \text{is the best response strategy of the Democratic candidate and} \ a_{R,j,m} \ \text{is the observed strategy of the Republican candidate.} \]

\[ \text{We did not find significant differences between the parties.} \]
average improvement per 1000 early morning GRPs in relative vote share over the actual allocation is 0.31 for the not close districts, 0.30 for the Likely districts, 0.29 for the Leaning districts, and 0.25 for the Toss Up districts. This result suggests that as races get closer, optimizing does not improve the average gains per GRP as much as it does in less close races. One interpretation of this result is that candidates in the closest races work harder at targeting and as a result achieve results closer to the optimal level. We also find a highly significant, negative effect for Incumbent. The average improvement per 1000 early morning GRPs in terms of relative vote share over the actual allocation is 0.32 for challengers and 0.28 for incumbents. This indicates that incumbents do better than challengers at targeting their advertising. This could be the result of knowledge or organizational relationships that incumbents have developed during previous races in that district. We do not find significant effects for frontrunner or tenure. Overall, the results are consistent with the theoretical incentives and learning one might expect in political races.

<table>
<thead>
<tr>
<th>Coefficient (Std. Err.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Republican</td>
</tr>
<tr>
<td>Likely</td>
</tr>
<tr>
<td>Leaning</td>
</tr>
<tr>
<td>Toss Up</td>
</tr>
<tr>
<td>Frontrunner</td>
</tr>
<tr>
<td>Incumbent</td>
</tr>
<tr>
<td>Tenure</td>
</tr>
<tr>
<td>R-squared</td>
</tr>
</tbody>
</table>

Table 6: Candidate Relative Improvement Regression – Note that all estimates are scaled up by 100 for ease of display. One star indicates statistical significance at the 5% level. Two stars indicates statistical significance at the 1% level. Three stars indicates statistical significance at the 0.1% level. A plus sign indicates statistical significance at the 10% level.

8 Conclusion

In this paper we study optimal targeting strategies for television advertising. We develop an approach for simulating exposure to television ads where overlapping segments (programs) are targeted. Lacking single-sourced data, this approach allows us to link separate surveys on television viewing and voting behavior using multiple imputation, leading to a “fused” data set with individual level measures of advertising exposure and voting behavior. This fused data enables a new approach for estimating advertising effects from observational data, leveraging the individual level variation in advertising exposures. We use these estimates to calculate the optimal targeting strategies and
examine how well campaigns target their ads.

Our results suggest that political advertising significantly influences candidate choice, but has a relatively small effect on turnout. Generically, targeting strategies should involve advertising on relatively cheap shows with high turnout rates and many swing voters. Actual candidates strategies are broadly consistent with this prescription. Nonetheless, we find that candidates could on average double the effectiveness of their advertising through better targeting. In particular, candidates could improve by more heavily targeting the most effective, cheapest programs and by targeting a smaller number of shows with very high turnout rates. Achieving the full gains of optimal targeting would require a high degree of sophistication from the campaigns, but we identified a number of simple heuristics that approximate optimal targeting. The simplest such heuristic—spending equally on the five programs in the two cheapest dayparts with the highest percentage of registered voters—would achieve about 60% of the benefit of optimal targeting, in comparison to the candidates current strategies.

Much of our approach could be extended to other products and media. First, our approach for simulating exposure to advertisements could be applied to other media and industries. Targeting on television is similar to targeting on radio, in magazines, on billboards, and on the web. In fact, the Simmons data we employ also has information on radio listening and magazine reading. The Simmons or similar data may also prove useful for applying our approach to other industries. Doing so would require identifying variables in the Simmons or similar data that are good proxies for primary demand (as percent registered voters are good proxies for turnout) and secondary demand (as ideology and party identification are good proxies for candidate choice). Fortunately, the Simmons data contains many purchase measures and a host of demographic and psychographic variables, although we note that some research has pointed to the limited use of demographics and psychographics at predicting secondary demand (Fennell et al. 2003).

Second, we develop a new method of estimating advertising effects from observational data that controls for endogeneity. Here, our new technique would primarily be of interest to academics, who often are in the position of studying advertising without the ability to manipulate advertising. For managers hoping to apply the techniques we observe in this paper, experimentation likely offers a superior approach for assessing advertising effects. While we believe our approach represents a reasonable way of uncovering advertising effects from observational data, we would of course not go
so far as to suggest that managers, who have the ability to test their campaigns in an experimental setting, forgo the opportunity to do so.

A number of special features of political advertising were beneficial for our study, but differ from many forms of commercial advertising. For example, because elections happen at a particular point in time, we are able to abstract away from scheduling issues, such as those considered in Dubé, Hitsch and Manchanda (2005). The fact that most congressional districts do not overlap multiple media markets means that we can abstract away from geographic allocation issues, such as those considered in Gordon and Hartmann (2010). The fact that individuals can vote at most once means that we can abstract away from differences in product use which may be relevant for many product categories. The special features of our particular application allowed us to address a topic—the targeting of advertisements across television programs—which has seen relatively limited recent attention. Beyond the insights our study generates for congressional campaigns, the techniques we introduce can be applied more broadly to address targeted advertising and television advertising in particular.

References


35
A Online Technical Appendix

A.1 Calculating Exposure Per Viewing Occasion

Consider an individual \( n \) with covariates \( x_n \). For each category in the Simmons data, we would like to estimate a logistic model of the form, \( \Pr(w_{n,p,c} = 1|x_n) = \Lambda(\gamma_p'x_n) \), where \( \Pr(w_{n,p,c} = 1|x_n) \) denotes the probability of viewing the program on a given viewing occasion. In our data, respondents are characterized by their viewing frequency. For example, for cable shows, the Simmons survey recorded whether the respondent viewed the show in the last 7 days and whether the respondent viewed the show during the last month. In this case, we have to convert the tabulations in the Simmons data into appropriate units. For example, we observe the proportion of black viewers among those that reported watching Larry King Live in the last week and the proportion of black viewers among those that reported watching Larry King Live in the last month. Larry King Live at the time ran 5 times a week. We had to construct an estimate of the proportion of black viewers during a single viewing occasion. There are five types of categories that we observed in the Simmons data-weekday network shows, weekly network shows, cable shows, sports, and local news. We developed a separate transformation for each of these categories to convert the demographic shares to shares for a single viewing occasion.

**Weekday Network Shows:** For weekday network shows, we observe an indicator of whether the show was viewed yesterday. In this case, no transformation was necessary.

**Weekly Network Shows:** For weekly network shows, we observe an indicator of whether the show was viewed in the last 7 days. In this case, no transformation was necessary.

**Cable Shows:** For cable shows, we must use a complicated transformation. For cable shows, we observe whether the individual watched the show during the last week and whether the individual watched the show during the last month. Denote these categories as \( l \) and \( l' \). Let the random variable \( w_{n,p,t} \) be an indicator of whether the individual watched the program in a given day. Assume for simplicity that there are 28 days a month and that the cable program airs \( k \) days a week (and consequently \( 4k \) days a month). We assume a split population model of viewing. With probability \( 1 - \vartheta \), an individual never watches the program. Conditional of being a “watcher” of the program, the probabilities of watching on a given day are independent and are given by \( \varphi \). Based on this model, we have,

\[
\Lambda(\gamma_l'x_n) = \Pr \left( \sum_{t=1}^{k} w_{n,p,t} \geq 1 \right) = \varphi(1 - (1 - \vartheta)^k)\Lambda(\gamma_l'x_n) = \Pr \left( \sum_{t=1}^{4k} w_{n,p,t} \geq 1 \right)
\]

\[
\varphi(1 - (1 - \vartheta)^{4k})
\]

This system defines two equation with two unknown, so we can back out the probabilities \((\vartheta, \varphi)\) from the observed data. Define \( r = (1 - \vartheta)^{k} \). We can divide the second equation by the first to obtain the cubic polynomial,

\[
r^3 + r^2 + r + 1 - \frac{\Pr(\sum_{t=1}^{4k} w_{n,p,t} \geq 1)}{\Pr(\sum_{t=1}^{k} w_{n,p,t} \geq 1)} = 0
\]
Using the cubic formula, we arrive at,

\[ r = \sqrt[3]{\zeta + \sqrt{\zeta^2 + (\frac{2}{3})^3}} + \sqrt[3]{\zeta - \sqrt{\zeta^2 + (\frac{2}{3})^3}} - \frac{1}{3} \tag{23} \]

where,

\[ \zeta = \frac{1}{27} \left[ 7 - 27 \left( 1 - \frac{\Pr(\sum_{t=1}^{4k} w_{n,p,t} \geq 1)}{\Pr(\sum_{t=1}^{k} w_{n,p,t} \geq 1)} \right) \right] \tag{24} \]

We can further determine that,

\[ \vartheta = 1 - r^{1-k} \varphi = \frac{\Pr(\sum_{t=1}^{k} w_{n,p,t} \geq 1)}{1 - r} \tag{25} \]

which leads us to the estimate \( \vartheta_{n,p} \varphi_{n,p} \) for the proportion of the population that views the cable show on a given airing.

**Sports Shows:** For sports shows, we observe the number of times the event was viewed in the last 12 months—once, twice, three times, or four or more. We assume that there are \( k \) possible instances in which the show can be viewed over a period of 12 months. We observe whether the sporting event was viewed exactly zero, one, or two times over this time period. We can calculate these probabilities as,

\[ \Pr(\sum_{t=1}^{k} w_{n,p,t} = 1) = \vartheta \left( \begin{array}{c} k \\ 1 \end{array} \right) \varphi^1 (1 - \varphi)^{k-1} \tag{26} \]

\[ \Pr(\sum_{t=1}^{k} w_{n,p,t} = 2) = \vartheta \left( \begin{array}{c} k \\ 2 \end{array} \right) \varphi^2 (1 - \varphi)^{k-2} \tag{27} \]

Notice that,

\[ \varphi = \frac{\Pr(\sum_{t=1}^{k} w_{n,p,t} = 2)}{\frac{1}{2} \Pr(\sum_{t=1}^{k} w_{n,p,t} = 1) + \Pr(\sum_{t=1}^{k} w_{n,p,t} = 2)} \tag{28} \]

We use,

\[ \Pr(\sum_{t=1}^{k} w_{n,p,t} = 0) = 1 - \vartheta + \vartheta (1 - \varphi)^k \tag{29} \]

to obtain,

\[ \vartheta = \frac{\Pr(\sum_{t=1}^{k} w_{n,p,t} = 0) - 1}{(1 - \varphi)^k - 1} \tag{30} \]

which leads us to the estimate \( \vartheta_{n,p} \varphi_{n,p} \) for the proportion of the population that views the cable show on a given airing.

**Local News:** For local evening news, we observe, for each network, an indicator of whether
the individual watched early evening news and whether the individual watched late evening news. Since we only observe one event, we cannot easily apply a split population model as before. Instead, we assume that the probability of watching local news on a given day is given by \( \varphi \). Based on this, we can estimate that,

\[
\Pr\left( \sum_{t=1}^{5} w_{n,p,t} \geq 1 \right) = 1 - \Pr\left( \sum_{t=1}^{5} w_{n,p,t} = 0 \right) = (1 - \varphi)^5
\]

Hence, we can estimate,

\[
\varphi = 1 - \left( \Pr\left( \sum_{t=1}^{5} w_{n,p,t} \geq 1 \right) \right)^{1/5}
\]

For early local morning news, we do not observe such a variable, so we proxy for it using the major morning news show on the network.

A.2 Deriving the Asymptotic Variance of the Minimum Distance Estimator

Recall that for the minimum distance estimator, the moments are given by,\(^{19}\)

\[
\hat{h}_{p,i,q}(\gamma_p) = \hat{\Pr}(w_{n,p,c} = 1|x_i = q) - \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \{ x_{n,i}^{rcs} = q \} \Lambda(\gamma'_p x_n)
\]

\[
= \hat{\Pr}(w_{n,p,c} = 1|x_i = q) - \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \sum_{x:x_i=q} \{ x_{n}^{rcs} = x \} \Lambda(\gamma'_p x)
\]

where \( p \) denotes the television program, \( i \) denotes the covariate, and \( q \) denotes the specific value that the covariate takes. To derive the asymptotic distribution, let us further assume that \( N_{sim} = \kappa N_{rcs} \) for \( \kappa > 0 \) (i.e. assume that the Simmons sample size grows proportional to the NAES rolling cross section sample size). The estimated probabilities \( \hat{\Pr}(w_{n,p,c} = 1|x_i = q) \) implicitly depend on the Simmons sample size and are computed using the formula,

\[
\hat{\Pr}(w_{n,p,c} = 1|x_i = q) = \frac{1}{N_{sim}} \sum_{n=1}^{N_{sim}} \sum_{x:x_i=q} \{ x_{n}^{sim} = x \} \Lambda(\gamma'_p x)
\]

Of course, we do not have access to the individual level Simmons data, \( x_n^{sim} \) and \( w_{n,p,c}^{sim} \).

We let \( \gamma_{p,0} \) denote the parameters of the data generating process that characterize viewership for show \( p \) and we let \( \hat{\gamma} \) and \( \gamma_0 \) denote the stacked vectors of \( \gamma_p \) and \( \gamma_{p,0} \) respectively. We let \( \hat{h}(\gamma) \) denote the stacked vector of moment conditions. The asymptotic distribution of the minimum distance estimator is given by,

\[
\sqrt{N_{rcs}} (\hat{\gamma} - \gamma_0) \overset{\text{dist.}}{\longrightarrow} N(0, H^{-1} V H^{-1})
\]

where,

\(^{19}\)Through this section, we use \( x_{n}^{rcs} \) to denote the vector of covariates, as observed in the NAES rolling cross section, and we use \( x_{n}^{sim} \) to denote the vector of covariates, as observed in the Simmons data.
\[ H = \frac{\partial}{\partial \gamma} \hat{h}(\gamma_0) \] (37)

\[ V = \text{Var}(\sqrt{N_{rcs}} \hat{h}(\gamma_0)) \] (38)

Estimating \( H \) is in principle straightforward—we can estimate it using \( \hat{H} = \frac{\partial}{\partial \gamma} \hat{h}(\hat{\gamma}) \), i.e. we can estimate it by taking numerical derivatives of the moment conditions at the estimated parameter value. Fortunately, the outer matrices \( H \) are block diagonal by program, so we avoid the need to invert very large matrices when calculating \( H^{-1} V H^{-1} \).

To characterize \( V \) we rewrite,

\[ \hat{h}_{p,i,q}(\gamma_{p,0}) = \hat{h}_{p,i,q}^{\text{sim}}(\gamma_{p,0}) - \hat{h}_{p,i,q}^{\text{rcs}}(\gamma_{p,0}) \] (39)

where,

\[ \hat{h}_{p,i,q}^{\text{sim}}(\gamma_{p,0}) = \text{Pr}(w_{n,p,c} = 1 | x_i = q) \] (40)

\[ \hat{h}_{p,i,q}^{\text{rcs}}(\gamma_{p,0}) = \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \sum_{x:x_i = q} 1 \{ x_{rcs}^n = x \} \Lambda(\gamma_{p,0}^t x) \] (41)

Since the Simmons and NAES RCS samples are independent, we have,

\[ \text{Var}(\sqrt{N_{rcs}} \hat{h}(\gamma_0)) = \text{Var}(\sqrt{N_{rcs}} \hat{h}_{p,i,q}^{\text{sim}}(\gamma_0)) + \text{Var}(\sqrt{N_{rcs}} \hat{h}_{p,i,q}^{\text{rcs}}(\gamma_0)) = V^{\text{sim}} + V^{\text{rcs}} \] (42)

Next, notice that,

\[ V_{p,q,i,p',q',i'}^{\text{sim}} = \text{Cov} \left( \sqrt{N_{rcs}} \hat{\text{Pr}}(w_{n,p,c} = 1 | x_i = q), \sqrt{N_{rcs}} \hat{\text{Pr}}(w_{n,p',c} = 1 | x_{i'} = q') \right) \] (43)

We can write,

\[ \sqrt{N_{rcs}} \hat{\text{Pr}}(w_{n,p,c} = 1 | x_i = q) \xrightarrow{\text{prob.}} \frac{1}{N_{sim}} \sum_{n=1}^{N_{sim}} \sum_{x:x_i = q} 1 \{ x_{sim}^n = x, w_{n,p,c} = 1 \} \] (44)

Applying a Law of Large Numbers to the denominator, we have,

\[ \sqrt{N_{rcs}} \hat{\text{Pr}}(w_{n,p,c} = 1 | x_i = q) \xrightarrow{\text{prob.}} \frac{1}{\sqrt{N_{sim}}} \sum_{n=1}^{N_{sim}} \sum_{x:x_i = q} 1 \{ x_{sim}^n = x, w_{n,p,c} = 1 \} \] (45)

We can similarly apply a multivariate law of large numbers to the stacked vector of the quantities \( \sqrt{N_{rcs}} \hat{\text{Pr}}(w_{n,p,c} = 1 | x_i = q) \). Now, consider the random variables,

\[ Z_{p,i,q} = \frac{1}{\sqrt{N_{sim}}} \sum_{n=1}^{N_{sim}} \sum_{x:x_i = q} 1 \{ x_{sim}^n = x, w_{n,p,c} = 1 \} \] (46)

We can compute the variance of \( Z_{p,i,q} \) to be,
\[ \text{Var}(Z_{p,i,q}) = \text{Var} \left( \frac{1}{\sqrt{N_{\text{sim}}}} \sum_{n=1}^{N_{\text{sim}}} \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \]  
\[ = \frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \text{Var} \left( \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \]  
\[ = \text{Var} \left( \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \]  
\[ = E \left( \left( \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right)^2 \right) - \left( E \left( \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \right)^2 \]  
\[ = E \left( \left( \sum_{x:x_i=q} E \left( 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \right)^2 \right) - \left( E \left( \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \right)^2 \]  
\[ = \left( \sum_{x:x_i=q} \text{Pr}(x, w_{n,p,c} = 1) \right) - \left( \sum_{x:x_i=q} \text{Pr}(x, w_{n,p,c} = 1) \right)^2 \]  
\[ = \left( \sum_{x:x_i=q} \text{Pr}(x) \Lambda(\gamma_{p,0} x) \right) - \left( \sum_{x:x_i=q} \text{Pr}(x) \Lambda(\gamma_{p,0} x) \right)^2 \]  

We can compute the covariance of \(Z_{p,i,q}\) and \(Z_{p',i',q'}\) to be,

\[ \text{Cov}(Z_{p,i,q}, Z_{p',i',q'}) \]

\[ = \text{Cov} \left( \frac{1}{\sqrt{N_{\text{sim}}}} \sum_{n=1}^{N_{\text{sim}}} \sum_{x:x_i=q} 1\{x_n^{\text{sim}} = x, w_{n,p,c} = 1\}, \frac{1}{\sqrt{N_{\text{sim}}}} \sum_{m=1}^{N_{\text{sim}}} \sum_{y:y_i'=q'} 1\{x_m^{\text{sim}} = y, w_{m,p',c} = 1\} \right) \]
\[
\frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \sum_{m=1}^{N_{\text{sim}}} \text{Cov} \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\}, \sum_{y:y_i'=q'} 1\{x_{m}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
\frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \text{Cov} \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\}, \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
= \text{Cov} \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\}, \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
E \left( \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \left( \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right) \right)
\]

\[
- E \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) E \left( \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
= E \left( \left( \sum_{x:x_i=q} \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right) \right)
\]

\[
- E \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) E \left( \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
= E \left( \left( \sum_{x:x_i=q} \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) \right)
\]

\[
- E \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) E \left( \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
= E \left( \left( \sum_{x:x_i=q, x_i'=q'} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1, w_{n,p',c} = 1\} \right) \right)
\]

\[
- E \left( \sum_{x:x_i=q} 1\{x_{n}^{\text{sim}} = x, w_{n,p,c} = 1\} \right) E \left( \sum_{y:y_i'=q'} 1\{x_{n}^{\text{sim}} = y, w_{n,p',c} = 1\} \right)
\]

\[
= \sum_{x:x_i=q, x_i'=q'} \text{Pr}(x) \Lambda(\gamma_{p,0}^0 x) \Lambda(\gamma_{p',0}^0 x) - \left( \sum_{x:x_i=q} \text{Pr}(x) \Lambda(\gamma_{p,0}^0 x) \right) \left( \sum_{x:x_i'=q'} \text{Pr}(x) \Lambda(\gamma_{p',0}^0 x) \right)
\]

Notice that when \(i = i'\) and \(q \neq q'\), we have,
\[
\sum_{x:x_1=q,x_i=q'} \Pr(x) \Lambda(\gamma_{p,0}^i x) \Lambda(\gamma_{p',0}^i x) = 0
\]  
(65)

We can apply Sluksky’s theorem to determine that,

\[
P_{\text{sim}}^{\text{var}} \xrightarrow{\text{prob.}}
\]

\[
\left\{
\begin{array}{ll}
\frac{N_{\text{rcs}}}{N_{\text{sim}}} \left( \sum_{x:x_1=q,x_i=q'} \Pr(x) \Lambda(\gamma_{p,0}^i x) \right) \left(1 - \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma_{p,0}^i x) \right) & , p = p', i = i', q = q' \\
- \frac{N_{\text{rcs}}}{N_{\text{sim}}} \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma_{p,0}^i x) \right) \left( \sum_{x:x_i=q'} \Pr(x) \Lambda(\gamma_{p',0}^i x) \right) & , i = i', q \neq q' \\
\frac{N_{\text{rcs}}}{N_{\text{sim}}} \left[ \left( \sum_{x:x_i=q,x_i=q'} \Pr(x) \Lambda(\gamma_{p,0}^i x) \Lambda(\gamma_{p',0}^i x) \right) - \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma_{p,0}^i x) \right) \left( \sum_{x:x_i=q'} \Pr(x) \Lambda(\gamma_{p',0}^i x) \right) \right] & , \text{otherwise}
\end{array}
\right.
\]

A similar approach can be used to derive an expression for \( V_{\text{sim}}^{\text{rcs}}_{p,i,q,p',i',q'} \). Define,

\[
W_{p,i,q} = \frac{1}{\sqrt{N_{\text{rcs}}}} \sum_{n=1}^{N_{\text{rcs}}} \sum_{x:x_1=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x)
\]  
(67)

We can characterize the variance of \( W_{p,i,q} \) as,

\[
\text{Var}(W_{p,i,q}) = \text{Var} \left( \frac{1}{\sqrt{N_{\text{rcs}}}} \sum_{n=1}^{N_{\text{rcs}}} \sum_{x:x_1=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right)
\]  
(68)

\[
= \frac{1}{N_{\text{rcs}}} \sum_{n=1}^{N_{\text{rcs}}} \text{Var} \left( \sum_{x:x_1=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right)
\]  
(69)

\[
= \text{Var} \left( \sum_{x:x_1=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right)
\]  
(70)

\[
= \mathbb{E} \left[ \left( \sum_{x:x_1=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right)^2 \right] - \mathbb{E} \left[ \sum_{x:x_i=q} 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right]^2
\]  
(71)

\[
= \sum_{x:x_1=q} \mathbb{E} \left[ 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x)^2 \right] - \left( \sum_{x:x_i=q} \mathbb{E} \left[ 1 \{ x_{n}^{\text{rcs}} = x \} \Lambda(\gamma_{p,0}^i x) \right] \right)^2
\]  
(72)

\[
= \sum_{x:x_1=q} \sum_{y} 1 \{ y = x \} \Lambda(\gamma_{p,0}^i x)^2 \Pr(y) - \left( \sum_{x:x_i=q} \sum_{y} 1 \{ y = x \} \Lambda(\gamma_{p,0}^i x) \Pr(y) \right)^2
\]  
(73)

\[
= \sum_{x:x_1=q} \Pr(x) \Lambda(\gamma_{p,0}^i x)^2 - \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma_{p,0}^i x) \right)^2
\]  
(74)

We can characterize the covariance of \( W_{p,i,q} \) and \( W_{p',i',q'} \) as,
\[ \text{Cov}(W_{p,i,q}, W_{p',i',q'}) \]

\[ = \text{Cov} \left( \frac{1}{\sqrt{N_{rcs}}} \sum_{n=1}^{N_{rcs}} \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x), \frac{1}{\sqrt{N_{rcs}}} \sum_{m=1}^{N_{rcs}} \sum_{y:y_{i'}=q'} 1 \{ x_m^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right) \]  

\[ = \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \sum_{m=1}^{N_{rcs}} \text{Cov} \left( \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x), \sum_{y:y_{i'}=q'} 1 \{ x_m^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right) \]

\[ = \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x), \sum_{y:y_{i'}=q'} 1 \{ x_m^{rcs} = y \} \Lambda(\gamma_{p',0}y) \]

\[ = \text{Cov} \left( \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x), \sum_{y:y_{i'}=q'} 1 \{ x_n^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right) \]

\[ = \text{Cov} \left( \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x), \sum_{y:y_{i'}=q'} 1 \{ x_n^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right) \]

\[ = E \left[ \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x) \right] \left[ \sum_{y:y_{i'}=q'} 1 \{ x_n^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right] \]

\[ - E \left[ \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x) \right] \left[ \sum_{y:y_{i'}=q'} 1 \{ x_n^{rcs} = y \} \Lambda(\gamma_{p',0}y) \right] \]

\[ = E \left[ \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x) \Lambda(\gamma_{p',0}x) \right] \]

\[ = E \left[ \sum_{x:x_i=q} 1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x) \Lambda(\gamma_{p',0}x) \right] \]

\[ - \left( \sum_{x:x_i=q} E[1 \{ x_n^{rcs} = x \} \Lambda(\gamma_{p,0}x)] \right) \left( \sum_{y:y_{i'}=q'} E[1 \{ x_n^{rcs} = y \} \Lambda(\gamma_{p',0}y)] \right) \]

\[ = \sum_{x:x_i=q} \text{Pr}(x) \Lambda(\gamma_{p,0}x) \Lambda(\gamma_{p',0}x) - \left( \sum_{x:x_i=q} \text{Pr}(x) \Lambda(\gamma_{p,0}x) \right) \left( \sum_{y:y_{i'}=q'} \text{Pr}(x) \Lambda(\gamma_{p',0}x) \right) \]

Notice that when \( i = i' \) and \( q \neq q' \), we have,
\[
\text{Cov}(W_{p,i,q}, W'_{p',i,q'}) = - \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x) \right) \left( \sum_{x:x_i=q'} \Pr(x) \Lambda(\gamma'_{p',0} x) \right)
\]

We apply Slutsky’s theorem once again to obtain,

\[
V_{p,i,q,p',i',q'}^{\text{rcs prob.}} \xrightarrow{\text{prob.}} \]

\[
\left\{ \begin{array}{l}
\frac{\sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x)^2 - \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x) \right)^2}{\Pr(x_i=q)^2}, \\
- \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x) \right) \left( \sum_{x:x_i=q'} \Pr(x) \Lambda(\gamma'_{p',0} x) \right),
\end{array} \right.
\]

\[
\frac{\sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x) \Lambda(\gamma'_{p',0} x^2) - \left( \sum_{x:x_i=q} \Pr(x) \Lambda(\gamma'_{p,0} x) \right) \left( \sum_{x:x_i=q'} \Pr(x) \Lambda(\gamma'_{p',0} x) \right)}{\Pr(x_i=q)^2 \Pr(x_{i'}=q')}, \\
\text{otherwise}
\]

To obtain estimate of the quantities \( V_{p,i,q,p',i',q'}^{\text{sim}} \) and \( V_{p,i,q,p',i',q'}^{\text{sim}} \), we replace \( \Pr(x) \) and \( \Pr(x_i=q) \) with the estimates,

\[
\hat{\Pr}(x) = \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} 1\{x_{n,rcs} = x\}
\]

\[
\hat{\Pr}(x_i = q) = \frac{1}{N_{rcs}} \sum_{n=1}^{N_{rcs}} \sum_{x:x_i=q} 1\{x_{n,rcs} = x\}
\]

Notice that \( \hat{\Pr}(x) \) never appears in the denominator, so we do not have to worry about empty cells leading to a division by zero since none of the tabulations \( \hat{\Pr}(x_i = q) \) are equal to zero.

Notice that our results imply that \( V \) is dense (there are few zeros in the matrix). The block diagonal pattern in \( H \) means that we can characterize,

\[
\sqrt{N_{rcs}} (\hat{\gamma}_p - \gamma_{p,0}) \xrightarrow{\text{dist.}} N(0, H_{p,p}^{-1} V_{p,p} H_{p,p}^{-1})
\]

Computing the full asymptotic variance matrix for \( \hat{\gamma} \) is much more difficult because the matrix contains about half a billion entries. Furthermore, sampling from the asymptotic distribution of \( \hat{\gamma} \) would require computing the Cholesky factorization of a matrix with about 25,000 rows and columns. Because sampling from \( \hat{\gamma} \) is so computationally costly and because we believe that uncertainty in \( \hat{\gamma} \) makes a relatively small contribution to the uncertainty in the second stage coefficients, we skip the step of sampling from \( \hat{\gamma} \) in our multiple imputation calculations and instead only use the asymptotic variance of \( \hat{\gamma} \) to provide standard errors for the first stage parameters.

### A.3 Estimates of Program Viewing Model

The minimum distance procedure produces 22,876 parameters (38 parameters * 602 shows), far too many to present in detail. To establish the face validity of these estimates, we instead regress the coefficients for each of the demographic levels (\( \gamma_{p,k} \)) on the cast demographic dummies and genre dummies.
We collected characteristics for all of the programs using numerous public sources, such as TV Guide, Wikipedia, and the Complete Directory of Prime Time Network and Cable TV Shows (Brooks and Marsh 2007). Based on these sources independent coders judged each program on several dimensions. First, the genre of each program was selected from a small set of genres with 13 programs not having an identifiable genre placed into an “other” category. Second, the cast was coded as either primarily “older” characters, “younger” characters, or a mix. The coders were instructed that older characters corresponded roughly to characters in their upper 40’s and up and younger characters to early 20’s and below. Third, the cast was coded for whether there were any leads—loosely defined as primary characters, hosts, or anchors—who were black, asian, Hispanic, or female on the show in that season.

The results for a subset of the show demographics are given in Table 7. Only the coefficients on the cast demographic dummies are reported in the table. In general, the results are quite intuitive—female viewers have a preference for shows with female leads, black viewers have a preference for shows with black leads, hispanic viewers have a preference for hispanic leads, younger viewers have a preference for younger characters, and older viewers have a preference for older characters. These effects are consistent with prior research on the match-up hypothesis for television viewing (Shachar and Emerson 2000). We use these estimates in order to impute the exposure to political advertisements (see Section 5.3).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Female</th>
<th>Black</th>
<th>Hispanic</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>18-24</td>
<td>25-34</td>
<td>45-54</td>
</tr>
<tr>
<td>Female Leads</td>
<td>0.242***</td>
<td>-0.014</td>
<td>0.052</td>
<td>0.064</td>
</tr>
<tr>
<td>Black Leads</td>
<td>-0.021</td>
<td>0.965***</td>
<td>0.067</td>
<td>0.030</td>
</tr>
<tr>
<td>Hisp. Leads</td>
<td>0.173</td>
<td>-0.128</td>
<td>0.170+</td>
<td>-0.087</td>
</tr>
<tr>
<td>Asian Leads</td>
<td>-0.329*</td>
<td>-0.475**</td>
<td>-0.112</td>
<td>-0.301*</td>
</tr>
<tr>
<td>Young Chars.</td>
<td>-0.070</td>
<td>-0.124</td>
<td>-0.019</td>
<td>0.282**</td>
</tr>
<tr>
<td>Old Chars.</td>
<td>-0.084</td>
<td>0.050</td>
<td>0.057</td>
<td>-0.164*</td>
</tr>
</tbody>
</table>

Table 7: Match Between Show Viewership Parameters and Cast Demographics – Results obtained by regressing the show viewership parameter \((\gamma_p)\) on cast demographics dummies and genre dummies. Only the coefficients on cast demographics are reported and standard errors are suppressed to save space. One star indicates statistical significance at the 5% level. Two stars indicates statistical significance at the 1% level. Three stars indicates statistical significance at the 0.1% level. A plus sign indicates statistical significance at the 10% level.

A.4 Proof of Existence and Uniqueness of Fixed Effects in Share Equations

The parameters \(\xi^t_{j,m}\) and \(\xi^v_{j,m}\) are estimated as the unique solutions to the share equations,

\[
\frac{1}{|N_{j,m}|} \sum_{n \in N_{j,m}} \Phi(\xi^t_{j,m} + \beta^t_n x_n + \alpha_t e_{n,T}) = \hat{s}^t_{j,m} \tag{88}
\]

\[
\frac{1}{|N_{j,m}|} \sum_{n \in N_{j,m}} \Phi(\xi^t_{j,m} + \beta^t_n x_n + \alpha_t e_{n,T}) = \hat{s}^v_{j,m} \tag{89}
\]

Here, we demonstrate that for each \(j\) and \(m\) with \(\hat{s}^t_{j,m} \in (0,1)\) and \(\hat{s}^v_{j,m} \in (0,1)\), there exist unique \(\xi^t_{j,m}\) and \(\xi^v_{j,m}\) that solve (88) and (89).

We begin by demonstrating that there is a unique \(\xi^t_{j,m}\) that solves (88). Notice that for each \(n\), the function \(\Phi(\xi^t_{j,m} + \beta^t_n x_n + \alpha_t e_{n,T})\) is continuous and strictly increasing in \(\xi^t_{j,m}\). Since
\[ \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) \] is the sum of continuous and strictly increasing functions in \( \xi_{j,m}^t \), it is also continuous and strictly increasing in \( \xi_{j,m}^t \). Notice further that \( \lim_{\xi_{j,m}^t \to -\infty} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) = 1 \) and \( \lim_{\xi_{j,m}^t \to \infty} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) = 0 \). Since \( \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) \) is an average of such functions, it must also be the case that \( \lim_{\xi_{j,m}^t \to -\infty} \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) = 0 \) and \( \lim_{\xi_{j,m}^t \to \infty} \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) = 1 \). Since \( \hat{s}_{j,m}^t \in (0, 1) \), the intermediate value theorem implies that there exists a \( \xi_{j,m}^t \) such that (88) holds. Since \( \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) \) is strictly increasing, there exists at most one \( \xi_{j,m}^t \) such that (88) holds.

Next, we show that for the \( \xi_{j,m}^t \) that uniquely solves (88), there is a unique \( \xi_{j,m}^v \) such that (89) holds. Recall that the bivariate normal c.d.f. satisfies the following properties,

1. For all \( x \), \( \Phi_2(x, y; \rho) = 0 \) is continuous and strictly increasing in \( y \)
2. \( \Phi_2(x, -\infty; \rho) = 0 \)
3. \( \Phi_2(x, \infty; \rho) = \Phi(x) \)

For each \( n \), the function,

\[ \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) \]

is continuous and strictly increasing in \( \xi_{j,m}^v \). Since,

\[ \sum_{n \in \mathcal{N}_{j,m}} \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) \]

is the sum of continuous and strictly increasing functions in \( \xi_{j,m}^v \), it is also continuous and strictly increasing in \( \xi_{j,m}^v \). Notice further that,

\[ \lim_{\xi_{j,m}^v \to \infty} \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) = \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) \]

\[ \lim_{\xi_{j,m}^v \to -\infty} \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) = 0 \]

Since (91) is an average of such functions, it must also be the case that,

\[ \lim_{\xi_{j,m}^v \to \infty} \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) \]

\[ = \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}) = \hat{s}_{j,m}^t \]

\[ \lim_{\xi_{j,m}^v \to \infty} \frac{1}{|\mathcal{N}_{j,m}|} \sum_{n \in \mathcal{N}_{j,m}} \Phi_2(\xi_{j,m}^t + \beta_t x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) = 0 \]
We therefore have that,

\[
\lim_{\xi_j,m \to \infty} \frac{1}{|N_{j,m}|} \sum_{n \in N_{j,m}} \Phi_2(\xi_{j,m}^t + \beta'_l x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta'_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) \leq 0
\]

\[
= \frac{s_{j,m}^t}{s_{j,m}^v} = 1
\]

\[
\lim_{\xi_j,m \to -\infty} \frac{1}{|N_{j,m}|} \sum_{n \in N_{j,m}} \Phi_2(\xi_{j,m}^t + \beta'_l x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta'_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho) \leq 0
\]

Since \( s_{j,m}^v \in (0, 1) \), the intermediate value theorem implies that when \( \xi_j,m \) satisfies (88), there exists a \( \xi_j,m^v \) such that (89) holds. Since,

\[
\frac{\sum_{n \in N_{j,m}} \Phi_2(\xi_{j,m}^t + \beta'_l x_n + \alpha_t e_{n,T}, \xi_{j,m}^v + \beta'_v x_n + \alpha_v (e_{n,R} - e_{n,D}); \rho)}{\sum_{n \in N_{j,m}} \Phi(\xi_{j,m}^t + \beta'_l x_n + \alpha_t e_{n,T})}
\]

is strictly increasing, there exists at most one \( \xi_j,m^v \) such that (88) and (89) hold, proving the result.

A.5 Additional Details on Multiple Imputation

Applying multiple imputation requires that we sample the exposure variables. We generate \( R \) different data sets with independent samples of the exposure variables and compute an estimate of the second stage parameters on each of these data sets. Assume that we have computed \( \{\hat{\theta}(r)\}_{r=1}^R \) with corresponding asymptotic variance matrices, \( \{\hat{V}(r)\}_{r=1}^R \). Our final estimator is given by,

\[
\hat{\theta} = \frac{1}{R} \sum_{r=1}^R \hat{\theta}(r)
\]

There are three sources of uncertainty about \( \theta_0 \). There is the uncertainty due to sampling variation in the second stage, there is uncertainty due to sampling variation in the first stage estimates \( \hat{\gamma} \), and there is uncertainty due to the fact that only a finite number of imputations are used.

There are two procedures that can be used to form the asymptotic sampling distribution for \( \hat{\theta} \). Rubin (1987) suggests that,

\[
\frac{\hat{\theta} - \theta_0}{\hat{s}} \sim t_{df}
\]

where,

\[
\hat{s} = \sqrt{\frac{1}{R} \sum_{r=1}^R \hat{V}(r) + (1 + \frac{1}{R}) \frac{1}{R-1} \sum_{r=1}^R (\hat{\theta}(r) - \hat{\theta})^2}
\]
\[ df = \left( R - 1 \right) \left( 1 + \left[ \frac{R+1}{R-1} \sum_{r=1}^{R} \left( \hat{\theta}(r) - \bar{\theta} \right)^2 \right]^{-1} \sum_{r=1}^{R} \tilde{V}(r) \right)^2 \] (102)

This method applies when the imputation method is “proper,” which means that each \( \hat{\theta}(r) \) is estimated based on an imputed data set in which \( \gamma \) is sampled from the posterior distribution of \( \gamma_0 \) and the missing variables are sampled from the model given the sampled value of \( \gamma \). In practice, this can be accomplished by sampling from the asymptotic distribution for \( \hat{\gamma} \), \( N(\hat{\gamma}, \sqrt{N}V_\gamma) \), which will approximate the posterior in large samples. Alternatively, Wang and Robins (1998) develop formulas for the asymptotic distribution for “improper” imputation methods, where the missing variables are sampled from the model given \( \hat{\gamma} \).

We apply Rubin’s formula for the standard errors, but ignore the first-stage sampling error by skipping over the step where \( \hat{\gamma} \) is sampled from the asymptotic distribution. We do this because, as we show in Appendix A.2, the asymptotic distribution is computationally costly to calculate and the Wang and Robins (1998) formula is also much more involved in our case. However, we believe this approach is reasonable. The first stage estimation is based on much larger sample sizes than the second stage estimation; the Simmons data and the NAES Rolling Cross Section, used in the first stage estimation, are about 10 and 20 times as large as the NAES Election Panel sample size, respectively.

A.6 Expressions for the Elasticities

Recall that for individual \( n \), the probability of turnout is given by,

\[ \Pr(y_n \neq 0|x_n) = \Phi(\bar{t}_n) \] (103)

where the elasticity with respect to total exposures can be calculated as,

\[ \frac{\partial \Phi(\bar{t}_n)}{\partial e_{n,T}} \left. \right|_{\Phi(\bar{t}_n)} = e_{n,T} \alpha_t \frac{\phi(\bar{t}_n)}{\Phi(\bar{t}_n)} \] (104)

We report the average elasticity, summing these calculations over respondents in the NAES Rolling Cross Section survey. Similarly, the probability of voting Republican is given by,

\[ \frac{\Pr(y_n = 2|x_n)}{\Pr(y_n \neq 0|x_n)} = \frac{\Phi_2(\bar{t}_n, \bar{v}_n)}{\Phi(\bar{t}_n)} \] (105)

so that the elasticity of voting Democratic with respect to Democratic exposures is given by,

\[ \frac{\partial}{\partial e_{n,D}} \left[ 1 - \frac{\Phi_2(\bar{t}_n, \bar{v}_n)}{\Phi(\bar{t}_n)} \right] = \frac{e_{n,D}}{1 - \frac{\Phi_2(\bar{t}_n, \bar{v}_n)}{\Phi(\bar{t}_n)}} \] (106)

and the elasticity of voting Republican with respect to Republican exposures is given by,

\[ \frac{\partial}{\partial e_{n,R}} \left[ \frac{\Phi_2(\bar{t}_n, \bar{v}_n)}{\Phi(\bar{t}_n)} \right] = \frac{e_{n,R}}{\Phi_2(\bar{t}_n, \bar{v}_n)} \] (107)
We have,

$$\frac{\partial}{\partial e_{n,D}} \left[ 1 - \Phi_2(t_n, \bar{v}_n) \right] = - \Phi(t_n) \frac{\partial}{\partial e_{n,D}} \Phi_2(t_n, \bar{v}_n) - \Phi_2(t_n, \bar{v}_n) \frac{\partial}{\partial e_{n,D}} \Phi(t_n)$$

where,

$$\frac{\partial}{\partial e_{n,D}} \Phi(t_n) = \alpha_t \phi(t_n)$$

$$\frac{\partial}{\partial e_{n,R}} \Phi_2(t_n, \bar{v}_n) = \alpha_t \phi(t_n) \phi(\bar{v}_n)$$

In the special case where $\rho = 0$, we can simplify the Democratic and Republican vote share elasticities to be,

$$\frac{\partial}{\partial e_{n,R}} \Phi_2(t_n, \bar{v}_n) = e_{n,R} \alpha_v \phi(\bar{v}_n)$$

which are similar to the turnout elasticity.

A.7 Illustration of Full Distribution of Programs

In the text, we only present figures with the average voting tendencies for each program, but our approach takes advantage of the distribution of viewers, not just these average tendencies. In Figure 5, we report the distribution for viewers of three programs with distinctive audiences—*Steve Harvey’s Big Time* (WB), *60 Minutes* (CBS), and *Cavuto on Business* (Fox News). We plot contour lines for the 20% and 50% quantiles for the audience members’ expected voting tendencies using a bivariate kernel density estimator. The graph depicts the variation within the audience of a show as well as the general voting tendencies of the audience. Although the show viewer profiles overlap, clear differences are apparent. Most viewers of *Steve Harvey’s Big Time* heavily prefer Democratic candidates and are relatively unlikely to vote, most viewers of *60 Minutes* prefer Democratic candidates and are likely to vote, and most viewers of *Cavuto on Business* prefer Republican candidates and are likely to vote.
Figure 5: Distribution of Voting Behavior for Three Shows.