The Informational Content of Campaign Advertising

Gregory J. Martin*

January 2020

Abstract

Understanding the mechanisms by which political advertising affects voters is crucial for evaluating the welfare effects of campaign finance and election regulation. This paper develops a method to distinguish between two alternative mechanisms for advertising influence: an informative channel in which voters learn about candidate attributes through advertising, and a prestige channel in which voters can be directly influenced by advertising even if it makes no contribution to the quality of information they possess about the sponsoring candidate. I show that both theories predict diminishing effects of advertising in voters’ experience with the candidate, but they have divergent implications for the strategic interaction between candidates. If advertising is informative, then opposing candidates’ advertising levels are strategic substitutes, whereas if advertising builds goodwill they are strategic complements. Using data from US Senate and gubernatorial elections from 2002-2014, I test for complementarities; results are consistent with the prestige hypothesis.

*Assistant Professor of Political Economy, Stanford Graduate School of Business, 655 Knight Way, Stanford CA 94305. Telephone: (650) 498-3321. Email: gjmartin@stanford.edu.
Elections perform two critical functions in a democracy: they allow citizens to express preferences over political outcomes, and they provide an accountability mechanism for public officials. Both of these functions work well under conditions of complete information, over candidates’ policy platforms or incumbents’ performance in office, respectively.

Real elections are, of course, far from the complete information benchmark. Voters must seek out what informative signals they can find as election day approaches. In many races, the most accessible such signals — in the sense of requiring the least exertion of search effort by voters — are paid advertisements run by candidates themselves or by third-party interest groups. The accessibility of advertising comes at the cost of the bias of its sources, whose primary objective is not to inform voters but to win election.

For voters, reliance on advertising thus involves a bias-variance tradeoff with unknown properties. This tradeoff can be worthwhile only if advertising decreases the variance of voters’ beliefs. That is, voters must learn something from advertising that they would not have learned in its absence.

It is possible that advertising conveys valuable information about the relative merits and policy positions of the candidates, to which at least some voters would not otherwise be exposed. Under this informative hypothesis, restricting campaign advertising, either directly by regulating advertisements or indirectly by limiting campaign contributions, would tend to degrade the quality of elected candidates and their policy congruence with constituents.

Alternatively, advertising may simply be vacuous priming that influences behavior in ways unrelated to the underlying characteristics of the candidate. Under this goodwill or prestige hypothesis, voters are manipulated into voting against their pre-advertising dispositions. Advertising restrictions would therefore tend to make election outcomes more representative of the voting population’s true preferences.

An informational mechanism for advertising influence is a necessary condition for advertising to improve voter welfare. In the absence of such a mechanism, voters get the bias cost without the variance reduction benefit, and increases in advertising do not improve the average quality of voter decisions. If there is a social cost associated with the fundraising activities that pay for campaign ads — for instance, that fundraising requires policy concessions to donors that the median voter dislikes — then advertising’s overall welfare effects will be even more negative.

The question of campaign advertising’s welfare implications is an important one for both scholars and policymakers, as all democracies regulate campaign communications to varying degrees. Extant regulatory approaches span the gamut from total prohibition on television advertising,\(^1\) to the post-\textit{Citizens United} American system where candidates’ advertising is limited

\(^{1}\text{A common system is a prohibition on privately financed electioneering communications, combined with mandated access to specific blocks of television time. These publicly provided television blocks are typically allocated on}\)
only by caps on contributions from individual donors, and outside groups are effectively unconstrained. The impact of these rules on the performance of the electoral mechanism hinges entirely on the importance of the informational channel of advertising influence.

The objective of this paper is, therefore, to distinguish informational from prestige effects of televised political advertising. This is challenging, because the theories make observationally equivalent predictions about voter responses to advertising. I demonstrate that an alternative approach, which relies only on the strategic interaction between candidates, can separate the two. The key insight is that if advertising is informative, then opposing candidates’ advertising levels are strategic substitutes. If, in contrast, advertising builds goodwill, then opposing candidates’ advertising levels are strategic complements. This method of empirical discrimination between the theories, unlike methods that rely on voter responses, does not depend on restrictive assumptions about the functional form of prestige effects. The approach assumes only that marginal returns to advertising are diminishing.

Across a variety of specifications that hold constant candidate- and market-level unobservables, the results are consistent with the prestige hypothesis. Candidates behave as if their own promotional advertising and their opponents’ attack advertising are strategic complements, not substitutes. This complementarity in best responses reveals that the prestige channel is the dominant mode of advertising influence.

I also conduct an analysis of the vote share impacts of advertising, using powerful instruments for advertising levels by candidates in US Senate and gubernatorial races. This analysis provides additional evidence that advertising has meaningful effects on election outcomes, consistent with existing work such as Gerber et al. (2011), Gordon and Hartmann (2013), and Spenkuch and Toniatti (2018). The combination of these two findings implies that advertising is largely wasteful from voters’ perspective: campaigns with balanced advertising levels provide no improvement in voters’ information or change in expected outcomes, but do cost money.

Large imbalances across parties and limited to a defined time window prior to the election.

In the 2016 federal election cycle, so-called Super PACs created in the wake of the ruling raised nearly $1.8B (Center for Responsive Politics, http://www.opensecrets.org) and aired more than 250,000 advertisements in the presidential race alone (Fowler et al., 2017b).

This result does not rule out the existence of the informational channel. There is high-quality experimental evidence (Kendall et al., 2015) that advertising can affect voters’ posterior beliefs about candidate types. Their setting (an Italian mayoral election) differs from mine on many dimensions, but one especially salient difference is that the overall level of advertising is much lower. One could imagine a more complex model where both channels are present, with relative weights that vary with the total quantity of advertising deployed. For instance, the first few doses of advertising might be informative, but further doses exhaust the informational benefit and any remaining effect is through the prestige channel. The result here indicates that at the levels of advertising actually observed in US state-wide elections, the prestige channel is the dominant one.

Pure-prestige advertising is wasteful under the usual consequentialist notion of voter utility: that utility is a function only of the elected candidate and the policies she enacts while in office. An expanded conception of voter utility that includes the “warm glow” of casting a vote for a heavily-advertised candidate might allow for a different
in advertising levels nonetheless have the potential to change election outcomes, a competitive
dynamic that induces candidates to exert effort on fundraising which could be spent elsewhere.
Limits on campaign advertising — at least, relative to the typical levels observed in American
elections — thus have the potential to improve voter welfare by reducing these costs without much
loss in voters’ knowledge of the candidates and issues at stake. This result is consistent with recent
findings on the effects of campaign expenditure limits, which find that higher caps systematically
advantage wealthier candidates, not more experienced or more competent candidates (Avis et al.,
2017; Fouirnaies, 2018).

In what follows, I introduce two models of advertising influence drawn from the literature on
product and political advertising. I show where their empirical predictions coincide and where
they diverge. I then test these predictions in a dataset of media-market-level advertising quantities
and vote shares in 265 American Senate and gubernatorial campaigns from 2002-2014. I conclude
with discussion of implications for campaign finance and regulation.

Informational and prestige effects of advertising

Models of advertising in political economics closely mirror models of advertising in product mar-
kets. The primary theoretical distinction between political and product advertising is that political
advertising is typically not paid for directly by candidates, but rather is financed by contributions
from a third party with distinct and perhaps competing interests. Aside from this complication,
the modeling technology and assumptions are quite similar to standard microeconomic treatments
of advertising, and can be similarly grouped according to their assumed information structure.

In one group, advertising reveals some information to voters. This information may be ver-
ifiable factual information contained in the content of the messages, as in Ashworth (2006) and
Polborn and Yi (2006), or indirect information revealed in the equilibrium of a signaling game, as
in Prat (2002). In both cases, advertising expenditures must be financed by outside groups, who
demand favors in exchange for their financial support. Voters trade off the positive signal adver-
tisements reveal against the knowledge that the messages must have been paid for by promises of
favors to special interests.6

In the second group, political advertising contains no information and is purely manipulative.
In Baron (1994), some fraction of voters are “uninformed” and respond to advertising in mechani-

---

5See Bagwell (2007) for a survey of the relevant industrial organization literature.
6These second-order inferences about the source of funding are specific to the American system of private cam-
paign finance and not a strictly necessary component of hard-information models. Kendall et al. (2015), in an appli-
cation outside of the US context, develop a model of verifiable information where voters learn only from the content
(or lack thereof) of the advertising message and not from any implication about how the ad was purchased.
cal fashion: higher levels of advertising by one candidate increase these voters’ propensity to vote for that candidate. The remaining “informed” voters are already partisans of one candidate or the other, and are unaffected by advertising. The effectiveness of political advertising in this setup thus depends primarily on the preponderance of the uninformed voter type in the electorate.

The economics of advertising literature refers to this latter kind of influence as “goodwill” or “prestige” effects of advertising (Bagwell, 2007). Even if consumers don’t learn anything about the characteristics of a product from an ad — and, therefore, exposure to the ad does not improve the expected utility of the consumer’s choice — the ad may build an aura of prestige around a product that increases the utility experienced by a consumer at purchase time. Scholars have argued that such goodwill advertising is cumulative. Ad campaigns can be thought of as adding to a stock of goodwill which persists over time (Clarke, 1976; Ashley et al., 1980; Seldon and Doroodian, 1989).

Political scientists and campaign practitioners often distinguish between “persuasive” versus “mobilizing” effects of ads (Kalla and Broockman, 2018; Lovett and Peress, 2015; Spenkuch and Toniatti, 2018). The former involves the vote-choice margin: for whom to vote, conditional on turning out. The latter involves the turnout margin: whether or not to show up to vote at all.

The framework here cuts across these categories. Both informative and prestige mechanisms can produce effects on either margin. Informative advertising might give a voter new information that convinces her to support a different candidate than she initially supported (persuasion), or she might learn that the candidates are similar enough that casting a vote is not worth the cost of getting to the polls (de-mobilization). Similarly, a prestige ad might inspire an irregular voter to turn out who might not have otherwise, and might also influence his vote choice once at the polls. Evidence that advertising primarily works on the turnout margin (Kalla and Broockman, 2018; Spenkuch and Toniatti, 2018) or that campaigns behave as if they are trying to persuade swing voters (Lovett and Peress, 2015) does not distinguish between these mechanisms. Accordingly, in my empirical specifications involving vote shares I use the candidate’s share of total possible votes, rather than votes cast, as the outcome. This definition captures both margins of influence.

One additional feature of political advertising that will prove to be important is the possibility of “going negative.” Unlike in product markets, a large fraction of candidates’ advertising budgets are spent attacking their opponents rather than promoting themselves.\(^7\) This difference stems directly from the difference in objectives. Product sellers care about the absolute quantity they sell, whereas politicians care about the number of ballots cast for them only relative to the number cast for their opponents. Converting a non-voter into a supporter is equally as valuable as converting an opponent’s supporter into a non-voter.

---

\(^7\) Fowler et al. (2017b) report that in presidential elections in the past two decades, between 30 and 60 percent of ads are pure attack ads; an additional 15-30 percent are “contrast” ads which contain some attacking elements.
It is therefore necessary to accommodate informational and prestige channels of both attack and promotional ads. I will make the assumption that an informative attack ad provides information about the sponsoring candidate’s opponent (Geer, 2006) and that a prestige attack ad reduces the stock of goodwill of the candidate’s opponent. Symmetrically, informative promotional ads provide information about the sponsoring candidate, and prestige promotional ads increase the sponsoring candidate’s stock of goodwill.\(^8\) With this extension, I sketch a model of each form of influence.

**A model of informative advertising**

I consider a race between two candidates for some office, the electorate for which is distributed across a set of media markets \(j\). Each market’s aggregate preferences are represented by a decisive voter\(^9\) with preferences over a candidate characteristic \(\xi_k\), \(k \in \{1, 2\}\). This characteristic is taken to be a scalar for simplicity,\(^{10}\) but results will extend straightforwardly to the multidimensional attribute setting.\(^{11}\) Voters will (imperfectly) observe the quality of promotional and attack advertising campaigns that the two candidates run, \((a^P_1, a^A_1, a^P_2, a^A_2)\). Initially, the voter’s prior belief about the joint distribution of \(\xi_k\) and \(a_k\) is that they follow the joint normal distribution:

\[
\begin{pmatrix}
\xi_k \\
 a^P_k \\
 a^A_k
\end{pmatrix}
\sim N\left( m_{0,j,k}, \Sigma_0 \right)
\]  

\(^{(1)}\)

Where \(m_{0,j,k} = (m^\xi_{0,j,k}, m^P_{0,j,k}, m^A_{0,j,k})'\). The prior mean on \(\xi\), \(m^\xi_0\), represents voters’ expectations about the average candidate’s attributes. The advertising priors \(m^P_0, m^A_0\) describe how well voters expect an average candidate to advertise on her own behalf, and how effective are the attacks they expect an average candidate’s opponent to mount.\(^{12}\) I allow for these priors to potentially vary across markets, to account for systematic differences in beliefs about candidate type

---

\(^8\)This is similar to the mechanism in Skaperdas and Grofman (1995) but restricted to two-candidate contests.

\(^9\)The assumption of a representative voter is for simplicity in explication of the model, but is not necessary for the results, which extend to the case of a mass of voters in each district with potentially heterogeneous preferences over attributes.

\(^{10}\)This is not a substantive restriction, as we can think of the scalar \(\xi_k\) as the projection of a (multidimensional) candidate attribute space onto the (unidimensional) voter utility space.

\(^{11}\)Characteristics could in general be policy positions, or “valence” attributes such as charisma or experience. Huber and Arceneaux (2007) find empirically that advertising has no effect on voter perceptions of candidates’ policy positions, but significant effects on perceptions of personal qualities. In principle, voters could learn both kinds of information from ads.

\(^{12}\)One way to think about this is to suppose that the voter has a model in her mind that describes how advertising quality is related to attributes, e.g.: \(a = f(\xi)\). The prior mean on ad quality \((m^P_0, m^A_0)\) then represents the average candidate’s advertising quality, e.g. \(f(m^P_0)\).
across markets: some markets may be initially favorable to some candidate and others hostile. \(\Sigma_0\) represents voters’ beliefs about the relationships (i.e., the correlation structure) between candidate attributes and advertising quality, and is common to all voters.

Next, voters observe a noisy signal about each candidate’s value of \(\xi\). This signal can be interpreted as the individual voter’s assessment of the candidate’s expected performance in office or policy preference alignment. Signals are unbiased and on average reflect the candidate’s true attributes, but are subject to an additive normal shock specific to each individual. The signal observed by the decisive voter in market \(j\) is thus:

\[
\hat{\xi}_{j,k} = \xi_k + \eta_{j,k}, \quad \eta_{j,k} \sim N(0, \sigma^2_{\xi,j,k})
\]

After voters observe the attribute signal, candidates choose advertising quantities in each of the media markets within the district, \((q_{j,k}^P, q_{j,k}^A)\). Increasing the quantity of advertising impressions purchased in some market gives the voters in that market more opportunities to observe the quality of the candidate’s campaign. I assume that each exposure is centered on the true quality level but has some iid additive normal error. Voters average over all the signals they observe, yielding the composite signals:

\[
\hat{a}_{j,k}^P = a_k^P + \bar{\eta}_{j,k}^P, \quad \bar{\eta}_{j,k}^P \sim N(0, \frac{1}{q_{j,k}^P} \sigma^2_{a, j,k})
\]

\[
\hat{a}_{j,-k}^A = a_{-k}^A + \bar{\eta}_{j,-k}^A, \quad \bar{\eta}_{j,-k}^A \sim N(0, \frac{1}{q_{j,-k}^A} \sigma^2_{a, j,-k})
\]

The substantive content of the above is that ad campaign quality is not a choice variable; candidates choose only quantities of advertising, which vary — through repetition — the precision of the signal that the average voter receives but not its mean. This is therefore a model of “hard” information, where candidates can choose to reveal (or not reveal) a signal about themselves but cannot freely choose the valence of that signal.\(^{14}\)

Following their observation of the advertising signals and the performance signal, voters update beliefs in Bayesian fashion. The larger is the number of ads that a voter sees, the more precise is the voter’s posterior estimate of the quality of the candidates’ campaigns, and correspondingly the more precise is her belief about the candidate attributes. Given the normality assumptions on priors and signals, the resulting posterior beliefs are also distributed normally.\(^{15}\) The posterior

\(^{13}\)Unbiasedness implies that candidates cannot manipulate their performance signal.

\(^{14}\)This does not literally require that candidates cannot control the content of their advertisements, or even that the message contained in an ad must be verifiable. All that is implied by the model is that candidates cannot control what voters infer about them, conditional on seeing an ad.

\(^{15}\)A derivation for the univariate case is given in Gelman et al. (2013), and extended to multivariate signals in
mean belief is given by:

\[
\hat{m}_{j,k} = m_{0,j,k} + \left( \Sigma_0^{-1} + \Phi_j^{-1} \right)^{-1} \Phi_j^{-1} \begin{pmatrix} \hat{\xi}_{j,k} - m_{0,j,k}^\xi \\ \hat{P}_{j,k} - m_{0,j,k}^P \\ \hat{A}_{j,-k} - m_{0,j,-k}^A \end{pmatrix}
\]

\[
\Sigma_{j,k} = \left( \Sigma_0^{-1} + \Phi_j^{-1} \right)^{-1}
\]

Where \( \Phi_j \) is the diagonal matrix with diagonal entries \( (\sigma_{\xi,j,k}^2, \sigma_{P,j,k}^2, \sigma_{A,j,-k}^2) \). The performance signals \( \hat{\xi}_{j,k} \) induce both a mean shift and a decrease in the variance in the decisive voter’s beliefs about each candidate’s type. The variance reduction is greater for candidate-market combinations which have smaller values of \( \sigma_{\xi,j,k}^2 \), e.g. smaller error variances in the quality signal. This reduction leaves proportionately less room for candidates’ advertising to influence beliefs.

Finally, each voter’s expected utility from voting for each candidate is a linear combination of the expected candidate attributes, plus an additive random shock:

\[
u_{j,k} = \beta_j \hat{m}_{j,k}^\xi + \gamma_j + \epsilon_{j,k}
\] (3)

And assuming that \( \epsilon_{j,k} \) is distributed iid Extreme Value Type I gives the standard homogeneous logit vote probabilities \( p_{j,k} \):

\[
p_{j,k} = \frac{e^{u_{j,k}}}{1 + \sum_{l\in\{1,2\}} e^{u_{j,l}}}
\] (4)

Here, the normalization is that the utility of the outside option (not voting) is zero. \( \gamma_j \) is the voting cost parameter in market \( j \): higher values mean greater likelihood that the voter casts a vote and vice versa. The probability of voter \( j \) not voting at all is \( v_{j,0} = \frac{1}{1 + \sum \epsilon_{j,l}} \). Differences in pre-advertising candidate preferences across markets are captured by variation in the prior means \( m_{0,j}^\xi \).

I will assume that \( \beta_j \) is positive for all \( j \), e.g. \( \xi_k \) is a valence characteristic which all voters like more of.\footnote{Ackerberg (2001).} The results which I present below continue to hold in the multidimensional attribute setting for any characteristic which voters weight positively.

The voter’s prior \( \Sigma_0 \) is a \( 3 \times 3 \) covariance matrix. I make three technical assumptions on the elements of this matrix, which fix the polarity of signals: \( \Sigma_0^{(1,2)} > 0 \), \( \Sigma_0^{(1,3)} < 0 \), and \( \Sigma_0^{(2,3)} < 0 \).

\footnote{Thinking about \( \xi_k \) as a projection from a multidimensional attribute space (containing policy preferences on many issues, measures of experience and competence, and so on) makes clear that this does not restrict the model’s applicability. There exists generically a direction of movement in the attribute space that all decisive voters like whenever the dimension of the attribute space exceeds the number of markets.}
These assumptions ensure that candidates want to advertise on their own behalf when voters’ priors underestimate the quality of their positive advertising (e.g. \( a_{j,k}^P - m_{0,j,k}^P > 0 \)) and want to run attack ads when voters’ priors underestimate the quality of their negative advertising (e.g. \( a_{j,-k}^A - m_{0,j,-k}^A > 0 \)). The logic behind the first two is straightforward, in ensuring that the direct effect of positive / negative advertising on voter beliefs about candidate quality go in the expected direction. The third assumption ensures that the indirect correlation between \( a_{j,k}^P \) and \( a_{j,-k}^A \) does not undo this direct effect.

Together, these assumptions rule out unnatural cases where, for instance, a candidate would like to run positive advertising on behalf of his opponent. Note that the third assumption reduces to

\[
\sum_{0}^{(1,2)} < 0
\]

as \( \sum_{0}^{(1,1)} \) becomes large, e.g. when voters are initially uncertain about candidate quality. This is exactly the scenario in which informative advertising makes sense from the candidates’ perspective.

### A model of prestige advertising

Again consider the same two-candidate contest with voters distributed across markets, and suppose there is a representative voter in each market \( j \). Unlike the previous section, where voters were rational Bayesians attempting to learn from signals, suppose here that the voter’s probability of voting for candidate \( k \) is an increasing function of \( k \)’s accumulated goodwill stock in market \( j \), \( g_{j,k} \equiv g(g_{0,j,k}^0 + q_{j,k}^P - q_{j,-k}^A) \), and a decreasing function of her opponent’s stock \( g_{j,-k} \equiv g(g_{0,j,-k}^0 + q_{j,-k}^P - q_{j,k}^A) \). Positive advertising is presumed to increase the candidate’s stock of goodwill, whereas negative advertising reduces the opponent’s.

The intercepts \( g_{0,j,k}^0 \geq 0 \) allow for some districts to favor one candidate over the other in the absence of any additional advertising. I presume the probability takes the form of a contest function (Baron, 1994; Skaperdas and Grofman, 1995; Jia et al., 2013):

\[
p_{j,k} = \frac{g(g_{0,j,k}^0 + q_{j,k}^P - g_{j,-k}^A)}{1 + g(g_{0,j,k}^0 + q_{j,k}^P - g_{j,-k}^A) + g(g_{0,j,-k}^0 + q_{j,-k}^P - g_{j,k}^A)}
\]

I presume that the goodwill function \( g \) is positive, twice differentiable, monotone increasing and globally concave. That is, \( g(\cdot) > 0 \), \( g'(\cdot) \geq 0 \), \( g''(\cdot) < 0 \). This implies \( \frac{\partial}{\partial q_{j,k}^P} g_{j,k} > 0 \), and \( \frac{\partial}{\partial q_{j,-k}^P} g_{j,k} < 0 \); in words, that positive advertising increases the candidate’s stock of goodwill, and her opponent’s negative advertising decreases it. The contest function structure implies that the relationship of vote probability to advertising levels also has these properties: \( \frac{\partial p_{j,k}}{\partial q_{j,k}^P} > 0 \), \( \frac{\partial p_{j,k}}{\partial q_{j,-k}^P} > 0 \).
Empirical predictions

I derive two empirical predictions from these models. On the first, the predictions are directionally equivalent; on the second they diverge. I discuss the intuition for each result below; proofs are relegated to the Appendix. I will later take both predictions to the data.

Variation of ad effectiveness with prior information

Consider a direct approach based on measuring voter responses to advertising, and testing how those responses vary with voters’ prior information about the candidate. One strategy would be to take advantage of an insight first employed formally by Ackerberg (2001, 2003) in studies of the market for single-serving yogurt, but which underlies the experimental analysis of Ansolabehere and Iyengar’s (1996) study of negative advertising: the effectiveness of a candidate’s informative advertising is decreasing in the precision of the voter’s prior information about the candidate. As Ansolabehere and Iyengar put it, “the voters who are most likely to learn from advertising are those who lack other sources of information.” In the terms of the model of informative advertising introduced above, \( \frac{\partial m_{j,k}}{\partial q_{j,k}} \) is increasing in \( \sigma^2_{\xi,j,k} \). We can state this hypothesis as:

**Proposition 1a.** In the informational model, the marginal effect of advertising is increasing in the variance of the performance signal \( \sigma^2_{\xi,j,k} \).

**Proof.** In Appendix A.1.

By observing variation in the effectiveness of advertising as a function of variation in levels of prior information, one might thus hope to measure the relative importance of the two effects. Electoral politics provides a strong shifter of voters’ prior information: incumbency status. Incumbency is associated with more precise information, as voters have had a chance to observe the incumbent’s — but not the challenger’s — past performance in office. Hence, if advertising is informative, we should expect lower effectiveness of incumbent advertising as compared to challenger advertising.

In the Results section I implement a test of this kind, using a within-candidate instrumental variables design that builds on existing identification strategies\(^\text{18}\) to address the endogeneity of

\[0, \frac{\partial^2 p_{j,k}}{\partial (q_{j,k})^2} < 0, \frac{\partial^2 p_{j,k}}{\partial (q_{j,k})^2} > 0.\]

\(^{17}\)See Appendix A.4 for derivation.

\(^{18}\)The method I employ is a mixture of the methods of Gordon and Hartmann (2013, 2016), who use price variation as instruments for advertising levels, and Ashworth and Clinton (2007); Huber and Arceneaux (2007); Spenkuch and Toniatti (2018) who use discontinuities at state boundaries.
advertising levels. The results show that, indeed, estimated effects of advertising on vote shares are substantially larger for challengers than for incumbents. This result is congruent with classic results in the campaign finance literature from candidate-level regressions (Jacobson and Carson, 2015) that challenger spending has larger effects on vote shares than does incumbent spending.

Unfortunately, this test is not dispositive. While it is consistent with informational advertising, it is also consistent with prestige advertising where increases to the accumulated stock of goodwill have diminishing marginal effects. In terms of the model of prestige advertising introduced above, suppose that incumbency increases the initial stock of goodwill $g^0_{jk}$ in all markets, due to voters’ exposure to ads in past campaigns. Given the assumption of concavity, the marginal effectiveness of advertising $\frac{\partial p_{jk}}{\partial q_{jk}}$ will then be lower for incumbents. This can be stated as:

**Proposition 1b.** In the prestige model, the marginal effect of advertising is decreasing in the initial goodwill stock $g^0$.

*Proof.* In Appendix A.3.

We generally expect that the initial goodwill stock will be higher for incumbent candidates than for challengers, because voters have been exposed to the incumbent’s advertising in previous campaigns. If an additional ad has a smaller marginal effect when previous exposure is higher, then we should expect to see smaller average partial effects for incumbents than for challengers, even if voters learn nothing from advertising. This result follows directly from the assumption of concavity in the voter reaction function $g$.

A directly related complication is that incumbent spending levels are typically higher than those of challengers. This implies that in observational data, the analyst observes marginal effects for challenger and incumbent expenditures locally to different points in the $g$ function. Unless one is willing to make strong functional form assumptions about the shape of $g$, this difference in levels confounds the incumbent-challenger comparison.

Hence, under plausible assumptions (voters know more about incumbents than challengers, and incumbents have greater accumulated goodwill stocks or on average spend more) the two models produce observationally equivalent predictions about the relationship of advertising effectiveness to incumbency status. I thus turn to a prediction where the models produce distinct predictions.

**Candidates’ strategic interactions**

I develop an alternative approach that uses information not on voter response to ads but on candidates’ strategic interactions: how candidates react to their opponents’ advertising choices. This
approach is analogous to what Athey and Stern (1998) call the correlation test of complementarity.\footnote{See also Arora and Gambardella (1990) and Arora (1996).}

**Informational model** If advertising is informational, then attack ads run by a candidate’s opponent are strategic substitutes with the candidate’s own promotional advertising. The intuition is the same as that underlying the incumbent/challenger differential described above: if the opponent’s attack ads provide information to voters about the candidate’s attributes, then the more attack ads the opponent runs, the more precise will voters’ beliefs be, and the more difficult will it be for the candidate’s own advertising to move voter beliefs. Consider a state with two markets, A and B. If candidate 1 runs attack ads in market A but not market B, then all else equal, candidate 2’s ads discussing her own merits are more cost effective in market B. The formal statement of the result is:

**Proposition 2a.** Suppose the informational model holds. Whenever a candidate captures less than 50% of the total possible votes,\footnote{Note that this refers to the share of all eligible voters, not the share of all votes actually cast.} and in any market where both candidates are willing to advertise, the marginal effect on vote shares of a candidate’s promotional advertising is decreasing in the level of her opponent’s attack advertising.

*Proof.* In Appendix A.2.

The two conditions in Proposition 2a — that both candidates are willing to advertise, and that one candidate is not too dominant in the election — do not generally bind in the empirical setting considered later. Most US Senate and gubernatorial campaigns advertise at positive levels, and use a mix of positive and negative ads. Turnout in most US Senate and gubernatorial elections is in the range of 40-60\%, implying that violating the second condition would require the combination of a relatively high-turnout election and a media-market-level share of votes cast in excess of 80\% for one candidate, which occurs in only one of 3476 observations in my data. Hence, the informational model predicts negative interactions between a candidate’s own promotional and her opponent’s attack advertising within a market, and thus a relationship of strategic substitutes between the two quantities.

The strict strategic substitutes relationship derived here depends crucially on the assumption of linkage between the two signals $a^p_k, a^A_k$, which is encoded in the model in the prior correlation matrix $\Sigma_0$. In the absence of such a connection — e.g., if $\Sigma_0^{(2,3)} = 0$ — observing the opponent’s attacks is uninformative about the candidate’s promotional advertising, and the predicted cross-partial derivative would be exactly zero. Another model of informational advertising that makes this kind of independence assumption is Polborn and Yi’s (2006) model of informative positive
and negative campaigning. Polborn and Yi assume that positive and negative messages inform voters on distinct, orthogonal dimensions. As both dimensions enter linearly in voters’ utility, there is no interaction between the two. Hence, their model predicts cross-partialis of zero. We will see in the empirical section that the data reject this prediction.

**Prestige model** In contrast, if advertising builds goodwill, then attack ads run by a candidate’s opponent are strategic complements with the candidate’s own promotional advertising. The more the opponent does to knock down the candidate’s stock of goodwill, the greater the marginal effectiveness of the candidate’s efforts to rebuild the stock. This result is a straightforward consequence of the concavity of the voter response function $g$.

**Proposition 2b.** Suppose the prestige model holds. The marginal effect on vote shares of a candidate’s promotional advertising is increasing in the level of her opponent’s attack advertising.

**Proof.** In Appendix A.4.

The import of Proposition 1b is that if the accumulation of goodwill stock has a concave effect on voter preferences, candidates with higher accumulated goodwill stocks (e.g., incumbents) will experience smaller marginal effects of advertising on vote shares. Because the function of negative advertising in the goodwill model is to reduce the opponent’s stock of goodwill, in fact, a decrease in the initial stock $g_{j,k}^0$ and an increase in the opponent’s attack advertising $q_{j,k}^A$ have effects on the marginal effect of $k$’s positive advertising that are equal in magnitude but opposite in sign, regardless of the shape of the response function $g$.\(^{21}\)

It is thus possible to distinguish the two models on the basis of a comparison of two comparative statics. In the informational model, the effect of incumbency (or an increase in any other non-advertising source of information to voters) and the effect of the opponent’s attack advertising both push in the same direction, of reducing the effectiveness of the candidate’s ads. In the prestige model, the two have opposite effects on the marginal impact of the candidate’s ads. In the next sections, I conduct a series of empirical analyses based on this idea.

**Data**

I take the models to the data in the context of US Senate and gubernatorial campaigns from 2002-2014. These races are a useful context to investigate due to their richness of candidate variation — especially in contrast to presidential campaigns, on which most studies of advertising have

\(^{21}\)Compare the expressions in Appendix sections A.3 and A.4.
focused — combined with sub-election level geographic variation to identify within-candidate advertising effects.

I use four primary types of data to estimate advertising effects and strategic interactions in these races. First, attributes of candidates’ advertising: quantities, tone, timing and location, which come from a combination of the Nielsen Company and the Wesleyan Media Project (WMP). Second, exogenous shifters of advertising exposures: a combination of state boundaries within media markets and variation in ad prices across market. The latter come from Spot Quotations and Data (SQAD) guides in the early years of the sample and Nielsen in the later years. Third, vote share data at the county level from the Congressional Quarterly elections database, which I aggregate to the level of the state-media market. Finally, a set of market, race and candidate level covariates — measures of pre-advertising political tastes, demographic characteristics, candidates’ fundraising ability, and so on — from a variety of sources. I describe the first two categories briefly here; remaining details of the data cleaning and joining steps as well as summary statistics are in Appendix B.

**Advertising data**

The most important component of the dataset is the level and tone of advertising chosen by each candidate in each of the media markets in the sample. As in Huber and Arceneaux (2007); Ashworth and Clinton (2007); Gordon and Hartmann (2013, 2016) and Spenkuch and Toniatti (2018) I rely on FCC regulations that require retransmission of broadcasts by all stations within geographically-defined Designated Market Areas (DMAs), meaning that voters in the same DMA see (in expectation) the same set of ads, and the candidate’s choice problem is how to allocate expenditures across DMAs within the state.

The two data sources I rely on — the Nielsen Ad Intelligence dataset for 2010-2014 and the Wesleyan Media Project (WMP) for 2002-2008 — measure advertising at the level of an ad buy, meaning a particular date, station, DMA, program, and time within the day. I measure advertising quantities in total impressions in the 18+ demographic — the estimated number of people over 18 who saw the ad — divided by the number of people over 18 in the media market. Nielsen reports 18+ impressions directly; for WMP data I infer impression quantity using WMP’s provided estimates of spot cost and average per-impression prices from SQAD.\(^{22}\) This measurement accounts for the heterogeneity in reach across ad spots: depending on the ratings of the underlying program, the same thirty-second spot might be viewed by an order of magnitude more viewers in one buy versus another. In addition to quantity, I also rely on measurements of the tone of ads; these come from the WMP for all years in the sample. The final dataset uses ad quantities aggregated to the

\(^{22}\) The method is the same as that used in Gordon and Hartmann (2013), and described in detail in Appendix B.
level of candidate-DMA-tone, where tone can be either positive or negative.

**Instruments**

Consistently estimating effects of advertising on votes requires some source of exogenous variation in levels of advertising. Instruments are necessary to avoid the selection problem that candidates may allocate advertising to markets on the basis of knowledge about the markets’ prior likelihood of supporting the candidate.

Fortunately, instruments which are likely to affect the candidates’ choices of where to advertise but are unrelated to unobserved political factors are available. I exploit variation in the (per-impression) prices of television advertising across markets within a state, which I collected from SQAD’s *Media Market Guides* and from the Nielsen Ad Intel data. SQAD publishes quarterly tables of prices of television advertising on a per-impression basis for all 210 DMAs in the United States.\(^{23}\) Because candidates have limited resources available to purchase advertising, they are likely to shift advertising into relatively low-cost markets at the expense of high-cost markets, relative to the allocation they would choose in a hypothetical world of costless advertising.

I use prices from the year *before* the relevant election as my instruments, in order to eliminate the potential problem that the same-year prices may be driven in part by the desirability of the market for political advertisers. Local station affiliates can derive significant revenues in the pre-election months from political advertisers, and hence it is at least conceivable that their price-setting process may reflect some market-level political unobservables. In the prior year, no election was happening and hence prices should reflect only the factors that make particular markets more or less attractive to standard, consumer-products advertisers.

A second source of exogenous variation arises from the design of DMAs. They are defined to encompass entire metropolitan areas, which often cross state boundaries. Hence, many states contain stub parts of DMAs whose main population centers are outside the state, but include some smaller urban or suburban communities in the stub state.

For example, the city of Vancouver, Washington, home to about 150,000 residents, lies directly across the Columbia river from the much larger city of Portland, Oregon. Vancouver, Portland, and Portland’s Oregon suburbs share a single television market, and Vancouver residents watch television broadcast by local Portland affiliates. Reaching these viewers requires buying advertising impressions on Portland television stations. From the perspective of a candidate running for a Washington state office, many of these impressions will be wasted on Oregon residents who are ineligible to vote in Washington elections.

\(^{23}\)Since the relevant demographic that candidates want to reach is voting-age adults, I use SQAD’s listed prices on a per-1000 impression basis within the “Adults 18+” demographic.
Discontinuities at state borders, as in the Vancouver example, increase the effective price of advertising for state-level candidates in markets that cross state boundaries. I therefore divide the raw prices by the fraction of the total DMA population that resides in the state of the election to construct an “effective price” paid by candidates in statewide elections interested in reaching the residents of only one state.

The instruments discussed thus far — effective prices of advertising in markets within a state — will be identical for both candidates within a given race. This feature may introduce an identification issue if the price instruments have the same, or nearly the same, effect on a candidate’s own advertising levels as they do on her opponent’s. To avoid this potential issue, I collected data on candidates’ fundraising activity,\(^{24}\) and computed each candidate’s total contributions. The idea is that candidates with bigger fundraising budgets are likely to be less sensitive to price differences across media markets. Thus, interacting the price instruments with candidates’ budgets generates instrumental variables which influence the allocation decisions of the two candidates in a race differently.\(^{25}\) I construct the candidate’s fundraising total (in dollars) divided by the average effective per-impression price.\(^{26}\) This composite variable, which I call the “effective ad budget,” has a simple interpretation as the number of in-state ad impressions that could be purchased if a candidate spent her entire budget in a given market.

Results

First stage

I first present the results of first stage regressions of advertising quantity on the instruments described previously. As expected, these variables are strong predictors of candidates’ total advertising expenditures in a market. The empirical specifications estimated take the form

\[
q_{jkt} = \alpha_{kt} + \beta_{kt}Z_{jkt} + \gamma X_{jkt} + \epsilon_{jkt}
\]

Where \(q_{jkt}\) is the total quantity of advertising, measured in impressions per capita\(^{27}\), purchased by candidate \(k\) in market \(j\) in election year \(t\); \(\alpha_{kt}\) is a candidate-year fixed effect absorbing the

\(^{24}\)Data for federal races comes from the Center for Responsive Politics (http://www.opensecrets.org). Data for state races is from Follow the Money (http://www.followthemoney.org).

\(^{25}\)It is of course possible that fundraising totals are correlated with the unobserved quality term \(\xi_k\). However, because I use exclusively within-candidate variation to identify the parameters, the fixed effects remove \(\xi_k\) from the error term.

\(^{26}\)The effective price is the price scaled by in-state share of DMA population.

\(^{27}\)Quantity measures are the total impressions in the 18+ demographic purchased in the DMA, divided by the voting-age population of the DMA in the election year.
candidate’s average advertising level across all markets; $Z_{jkt}$ is a measure of effective price for candidate $k$ in market $j$ in election year $t$; and $X_{jkt}$ are candidate-market-year specific control variables. Given the inclusion of candidate-year fixed effects, coefficients in the equation will be identified by within-candidate variation in $Z$ and $X$. A candidate running for state-level office will typically face different prices, and a different composition of the electorate, in several markets within the state. All results will cluster standard errors at the level of the race. Each cluster therefore contains two observations per market — one for each candidate — in all markets within a state-election. This clustering allows for within-race correlation induced by interactions across candidates and markets in the same race.

Table 1: First Stage Regressions of Advertising Quantity on (Effective) Prices

<table>
<thead>
<tr>
<th></th>
<th>Column (1)</th>
<th>Column (2)</th>
<th>Column (3)</th>
<th>Column (4)</th>
<th>Column (5)</th>
<th>Column (6)</th>
<th>Column (7)</th>
<th>Column (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log(Price)</td>
<td>$-0.009^{***}$</td>
<td>$-0.008^{***}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In-State Population Share</td>
<td></td>
<td></td>
<td>$0.009^{***}$</td>
<td>$0.006^{**}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log (Effective Price)</td>
<td></td>
<td></td>
<td></td>
<td>$-0.009^{***}$</td>
<td>$-0.007^{***}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effective Ad Budget</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$0.008^{***}$</td>
<td>$0.006^{***}$</td>
<td></td>
</tr>
</tbody>
</table>

Fixed Effects: 
- Cand-Year: Y
- Partisanship: Y
- Demographics: N

Number of Races: 265

R^2: 0.506, 0.576, 0.553, 0.587, 0.561, 0.593, 0.560, 0.596

*p < .1; **p < .05; ***p < .01
An observation is a candidate-DMA. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the number of advertising impressions per capita purchased by the candidate in the DMA. Instruments are (columns 1-2) the log of the DMA-average price per 1000 impressions in the previous year; (columns 3-4) the fraction of residents in the DMA who live in the state; (columns 5-6) the effective log price, computed as log previous year price per 1000 impressions divided by the fraction of in-state DMA residents; and (columns 7-8) the candidate’s total fundraising (in dollars) divided by the effective log price. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

Table 1 shows estimates of the parameter $\beta^q$ in equation 5 for various choices of the cost-shifter $Z_{jkt}$. Columns 1-2 of this table use logged average price in the DMA in the previous year; this varies by market and year but not across candidates advertising in the same market. Estimates are strongly negative, as expected; higher prices mean less advertising is possible for the same budget.

Columns 3-4 ignore the price variation and use only the share of the DMA population that is in the state in which the election is taking place; this varies by market and year, though almost all variation is across market. Again the effect is significant, but positive: having more of the market population reside in the state of the election increases the per-capita amount that state-level candidates advertise in the market.

Columns 5-6 combine the two sources into a single “effective price” by dividing the log price by the fraction of DMA population in state. This improves precision of the estimate substantially.

I use logged prices here because there are a few outlier markets with very high per-impression prices; taking logs thus improves the first stage precision substantially.
Finally, Columns 7-8 introduce the candidate’s fundraising budget, constructing an “effective ad budget” by dividing the candidates’ fundraising dollars by the effective log price. The interpretation of this variable is the quantity of in-state advertising that could be purchased if a candidate spent her entire budget in a given market. This varies across candidates as well as markets and years. The sign reverses here because of the division step (higher price means lower effective budget) but again the estimate is precisely estimated and nonzero.

Second stage

I next present second-stage models of vote share on advertising. Models estimated here take the form

$$\log s_{jkt} - \log s_{jk0} = \alpha_{kt}^s + \beta^s q_{jkt} + \gamma^s X_{jkt} + \epsilon_{jkt}^{s} \quad (6)$$

Where $s_{jkt}$ is candidate $k$’s share of total possible votes in market $j$ in year $t$, $s_{jk0}$ is the share of nonvoters (eligible voters who did not cast a vote) in market $j$ in year $t$, and $q_{jkt}$ is either the actual (in OLS specifications) or predicted (in IV specifications) value of the candidate’s per-impression advertising quantity in market $j$ in election $t$.\(^{29}\)

Again, the inclusion of candidate-year fixed effects mean that coefficients are estimated using within-race variation in advertising levels. E.g., a positive $\beta^s$ indicates that candidates on average perform better than their statewide share in markets where they purchased greater than their statewide average level of advertising. The fixed effects specification eliminates the potential for confounding from possible correlation between a candidate’s fundraising ability (and hence, her advertising budget) and her other electorally-valuable skills. It does not eliminate potential confounding from the fact that within a race, candidates choose advertising strategically and may preferentially target markets that are unobservably more or less favorable to the candidate even in the absence of advertising. The IV strategy aims to eliminate this source of bias by using variation in levels driven by factors — ad prices and the shape of media markets — plausibly exogenous to candidate support. Again, all reported standard errors are robust to clustering at the level of the race.

Table 2 presents results of estimating equation 6 on the full sample of candidates. Columns 1-2 are OLS, regressing log vote share differences on observed advertising quantities. Columns 3-4 are IV, using as instrument the “effective ad budget” whose first stage is shown in Table 1.

\(^{29}\)The log-difference specification is inspired by the homogeneous logit (Berry, 1994) form of equation 4, in which this transformation produces a linear function of the components of the voter’s utility function. All specifications in this section use the total quantity of advertising in impressions per capita, and do not distinguish between positive or negative advertising.
Table 2: Second Stage Regressions of Vote Share on (Predicted) Advertising Quantity

<table>
<thead>
<tr>
<th></th>
<th>Log Vote Share Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Impressions per Capita</td>
<td>0.719***</td>
</tr>
<tr>
<td></td>
<td>(0.144)</td>
</tr>
<tr>
<td>Predicted Impressions Per Capita</td>
<td>0.815**</td>
</tr>
<tr>
<td></td>
<td>(0.396)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
</tr>
<tr>
<td>Partisanship:</td>
<td></td>
</tr>
<tr>
<td>Demographics:</td>
<td></td>
</tr>
<tr>
<td>Number of Races</td>
<td>3.210</td>
</tr>
<tr>
<td>N</td>
<td>3,210</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.741</td>
</tr>
</tbody>
</table>

* p < .1; ** p < .05; *** p < .01
An observation is a candidate-DMA. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the log of the candidate’s vote share (as a fraction of population over 18) minus the log of the fraction of non-voters. Columns 1-2 are OLS. Columns 3-4 instrument for advertising quantity with the candidate’s total fundraising (in dollars) divided by the effective log price. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

Both OLS and IV yield positive and significant estimates of advertising effects. IV estimates are larger than the corresponding OLS, indicating that candidates’ strategic choices yield a negative bias. In other words, candidates appear to be targeting more ads to markets where their expected support is unobservably lower. This kind of targeting behavior is consistent with the Ansolabehere and Iyengar (1996) motive of de-mobilizing the opponent’s supporters with negative ads. It is also consistent with persuasion as the primary goal of advertising, as in Lovett and Peress (2015).

Table 3: Second Stage Regressions of Vote Share on (Predicted) Advertising Quantity: Incumbent Candidates

<table>
<thead>
<tr>
<th></th>
<th>Log Vote Share Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Impressions per Capita</td>
<td>0.516**</td>
</tr>
<tr>
<td></td>
<td>(0.220)</td>
</tr>
<tr>
<td>Predicted Impressions Per Capita</td>
<td>0.580</td>
</tr>
<tr>
<td></td>
<td>(0.730)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
</tr>
<tr>
<td>Partisanship:</td>
<td></td>
</tr>
<tr>
<td>Demographics:</td>
<td></td>
</tr>
<tr>
<td>Number of Races</td>
<td>173</td>
</tr>
<tr>
<td>N</td>
<td>1,144</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.734</td>
</tr>
</tbody>
</table>

* p < .1; ** p < .05; *** p < .01
An observation is a candidate-DMA. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the log of the candidate’s vote share (as a fraction of population over 18) minus the log of the fraction of non-voters. Columns 1-2 are OLS. Columns 3-4 instrument for advertising quantity with the candidate’s total fundraising (in dollars) divided by the effective log price. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

Tables 3 and 4 split the sample into, respectively, incumbent and challenger candidates. The difference between the two is clear, with effects in the incumbent sample roughly half the magni-
Table 4: Second Stage Regressions of Vote Share on (Predicted) Advertising Quantity: Challenger Candidates

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Vote Share Difference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impressions per Capita</td>
<td>0.844***</td>
<td>0.908***</td>
<td>(0.205)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Predicted Impressions Per Capita</td>
<td>0.942*</td>
<td>2.901***</td>
<td>(0.536)</td>
<td>(0.674)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partisanship:</td>
<td>N Y N Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics:</td>
<td>N Y N Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Races</td>
<td>230 230 230 230</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,066 2,066 2,066 2,066</td>
<td>0.727 0.869 0.727 0.852</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01

An observation is a candidate-DMA. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the log of the candidate’s vote share (as a fraction of population over 18) minus the log of the fraction of non-voters. Columns 1-2 are OLS; Columns 3-4 instrument for advertising quantity with the candidate’s total fundraising (in dollars) divided by the effective log price. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

In some specifications I cannot reject the null of zero effect in the incumbent sample, whereas this does not happen in the challenger sample. The data are thus consistent with the prediction, common to both the informational and the prestige model with concave voter reaction function, that the marginal effects of advertising are lower for incumbent than for challenger candidates.

Strategic interactions

I turn now to evaluating the hypotheses in Propositions 2a and 2b on the strategic interaction between the positive advertising of one candidate and the negative advertising of her opponent. The test I implement involves estimating the response function of one candidate’s promotional advertising to her opponent’s attack advertising. This approach is what Athey and Stern (1998) call the “correlation test” of complementarity. A positive slope of this function — more attack advertising from the opponent leads to more promotional advertising from the candidate — indicates complementarity, whereas a negative slope indicates substitution. I estimate specifications of the form:

$$q_{j,k,t}^P = \alpha_{k,t}^P + \beta Z_{j,k,t} + \gamma X_{j,k,t} + \delta q_{j,-k,t}^A + \epsilon_{j,k,t}$$

The empirical challenge of this approach is that the existence of common factors that push both candidates’ advertising levels in the same direction will tend to produce bias towards finding complementarity if these factors are omitted from the empirical specification (Arora, 1996). As an extreme example, suppose some state consists of two markets of similar size, in one of which turnout is typically very low (say, 10%) and in the other it is typically very high (say, 80%).
Candidates running in state-level elections will, naturally, concentrate campaign activity on the second market, which holds the vast majority of likely voters. A naive analysis might conclude that candidates are positively reacting to the opponent’s choice (running relatively more ads in markets where the opponent also ran relatively more ads) when in reality a common factor — expected turnout — is the driver of both choices.

I deal with this potential problem in four ways. First, I note that per the results in Table 1, the candidate fixed effects $\alpha$, market characteristics $X$ and cost shifters $Z$ are excellent predictors of advertising levels. This limits the scope of concern about the existence of unmeasured confounders, and also allows comparisons of estimates of $\delta$ in equation 7 with and without conditioning on $Z$ or $X$. This comparison allows for an estimate of the degree to which unobservable common factors could bias the results (Oster, 2019).

A second approach is to predict the opponent’s attack advertising level with an instrument that changes the expected level of attack advertising but is unrelated to market-specific unobservables. This IV approach measures responses to changes in opponents’ advertising levels that are orthogonal to the market-level unobservables that might also directly influence candidate spending. I use the opponent’s share of attack advertising in other markets (e.g., the fraction of the opponent’s impressions of attack advertising compared to total impressions, in all other markets but the one in question) as an instrument. The logic is that the same ads are often re-run many times, and thus the existence of pre-existing negative content makes running negative ads easier, regardless of the conditions on the ground in the market in question. By construction, this instrument is unrelated to market-specific unobservables.

Table 5: Estimated Best Response Functions

<table>
<thead>
<tr>
<th></th>
<th>Own Promote Impressions Per Capita</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Opponent’s Attack Impressions per Capita</td>
<td>0.421***</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
</tr>
<tr>
<td>(Predicted) Opponent’s Att. Imp. per Capita</td>
<td>0.002***</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>Effective Ad Budget</td>
<td></td>
</tr>
<tr>
<td>Partisanship:</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Demographics:</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2,944</td>
</tr>
<tr>
<td>Y</td>
<td>2,869</td>
</tr>
<tr>
<td>Number of Races</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td>223</td>
</tr>
<tr>
<td>R²</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>0.603</td>
</tr>
</tbody>
</table>

*p < .1; **p < .05; ***p < .01

An observation is a candidate-DMA. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the candidate’s promotional ad impressions per capita. Columns 1-2 are OLS; Columns 3-4 instrument for opponent’s attack ad quantity with the opponent’s fraction attack ads in other markets. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of Hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

Table 5 shows the results of these analyses. Columns 1 and 2 show estimates of $\delta$ in equation 7, respectively excluding and including cost-shifters $Z$ and market-level covariates $X$ from the

---

30The sample falls slightly relative to Tables 1 and 2, because I restrict here to races where both candidates did
specification. In both cases the estimated coefficient is large and positive. The magnitude of the effect falls, but only by about 15%, when these covariates are introduced into the specification. This small shift obtains even though both sets of covariates are strong predictors of ad levels. Accordingly, the proportional-selection test statistic for the hypothesis of zero true effect (Oster, 2019) comparing columns (1) and (2) is 1.10, above the usual robustness threshold of 1.\textsuperscript{31}

Results are very similar in magnitude and show a similar relative pattern in columns 3 and 4, which replace observed levels of attack advertising in market $j$ with levels instrumented by the opponent’s fraction of advertising that is negative in all markets except $j$. Standard errors widen significantly here, but the substantive conclusion is identical. Again, including market-level predictors lowers the estimate but by a small amount, roughly 13%.

It is of course still possible that the market-level predictors in Table 5 miss some important confounder that is known to candidates but unobserved in the data. A third approach deals with this possibility by taking advantage of the within-race time dimension of the data. Senate and gubernatorial campaigns last for months, and candidates do not wait until the final days of a campaign to deploy advertising. Competitive campaigns begin advertising far enough in advance of the election date that there is ample time to observe the opponent’s initial advertising choices and respond accordingly.

The existence of over-time, within-market variation in advertising levels allows me to estimate the candidate response function using market-candidate fixed effects; e.g., while holding constant any time-invariant unobservable characteristics specific to the match between candidate and DMA that may affect advertising levels. This can be thought of as a more complete implementation of the idea of conditioning on predictors of ad quantity, holding constant not just observable but also unobservable market characteristics. I can estimate a specification analogous to Equation 7 but using advertising disaggregated to the weekly level:

$$q_{j,k,t,w}^P = \alpha_{j,k,t}^P + \delta q_{j,-k,t,w}^A + \xi_w + \epsilon_{j,k,t,w}^P$$ (8)

Here, $w$ indicates the week of the campaign, defined relative to the election date. $\xi_w$ are week dummies which capture the temporal variation in the data. Advertising levels in all campaigns ramp up steadily as the election approaches. Controlling flexibly (e.g., with a complete set of dummy variables) for the number of weeks remaining is essential to eliminate the potential for bias in estimates of $\delta$ driven by both candidates’ advertising levels rising, in nonlinear fashion, as some advertising. It does not make sense to estimate within-race “reaction” functions of candidates to opponents who do no advertising at all. There are 223 such races in the data.

\textsuperscript{31}The statistic is calculated using $R_{MAX}^2 = 0.828$, which is the average $r$-squared value of race-by-race regressions of advertising levels on DMA fixed effects across all 223 races in the sample. The thought experiment is to estimate how much variation in advertising levels we could explain if all market-level unobservables common to both candidates in a race were perfectly observed.
the election approaches.

Table 6, column 1 estimates this model. The magnitude of the estimated $\delta$ comes down compared to the estimates in Table 5, but the estimate remains substantively positive and statistically different from zero.

Column 2 of the table adds lagged values of the opponent’s attack advertising choice from previous weeks, e.g. $q^A_{j,-k,t,w'}$ for $w' < w$. The idea is to allow for lags in candidate response time and measure current response to past actions by the opponent. The substantive conclusion is unchanged here: the slope of the reaction function is positive. The correlation declines slowly as we look farther back in time, but estimated lag coefficients out to three weeks prior are positive.

<table>
<thead>
<tr>
<th>Table 6: Estimated Best Response Functions, Weekly Ads Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Promote Impressions Per Capita</td>
</tr>
<tr>
<td>(1) (2) (3) (4) (5) (6)</td>
</tr>
<tr>
<td>Opponent’s Attack Impressions per Capita</td>
</tr>
<tr>
<td>0.178***</td>
</tr>
<tr>
<td>(0.055)</td>
</tr>
<tr>
<td>0.130***</td>
</tr>
<tr>
<td>(0.049)</td>
</tr>
<tr>
<td>0.268***</td>
</tr>
<tr>
<td>(0.057)</td>
</tr>
<tr>
<td>0.235***</td>
</tr>
<tr>
<td>(0.053)</td>
</tr>
<tr>
<td>0.151***</td>
</tr>
<tr>
<td>(0.048)</td>
</tr>
<tr>
<td>0.137***</td>
</tr>
<tr>
<td>(0.044)</td>
</tr>
<tr>
<td>1-week Lag Opponent’s Att. Imp. per Capita</td>
</tr>
<tr>
<td>0.064***</td>
</tr>
<tr>
<td>(0.021)</td>
</tr>
<tr>
<td>0.083***</td>
</tr>
<tr>
<td>(0.024)</td>
</tr>
<tr>
<td>0.078***</td>
</tr>
<tr>
<td>(0.022)</td>
</tr>
<tr>
<td>2-week Lag Opponent’s Att. Imp. per Capita</td>
</tr>
<tr>
<td>0.082**</td>
</tr>
<tr>
<td>(0.037)</td>
</tr>
<tr>
<td>0.118**</td>
</tr>
<tr>
<td>(0.052)</td>
</tr>
<tr>
<td>0.103**</td>
</tr>
<tr>
<td>(0.047)</td>
</tr>
<tr>
<td>3-week Lag Opponent’s Att. Imp. per Capita</td>
</tr>
<tr>
<td>0.056</td>
</tr>
<tr>
<td>(0.032)</td>
</tr>
<tr>
<td>0.093***</td>
</tr>
<tr>
<td>(0.029)</td>
</tr>
<tr>
<td>0.072***</td>
</tr>
<tr>
<td>(0.027)</td>
</tr>
<tr>
<td>Fixed Effects:</td>
</tr>
<tr>
<td>Cand-DMA-Year</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>Cand-DMA-Year</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Cand-Week</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>Cand-Week</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Cand-Week</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>Cand-Week</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>Number of Races</td>
</tr>
<tr>
<td>223</td>
</tr>
<tr>
<td>219</td>
</tr>
<tr>
<td>223</td>
</tr>
<tr>
<td>219</td>
</tr>
<tr>
<td>219</td>
</tr>
<tr>
<td>219</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>42.902</td>
</tr>
<tr>
<td>34.078</td>
</tr>
<tr>
<td>42.902</td>
</tr>
<tr>
<td>34.078</td>
</tr>
<tr>
<td>34.078</td>
</tr>
<tr>
<td>34.078</td>
</tr>
<tr>
<td>$R^2$</td>
</tr>
<tr>
<td>0.487</td>
</tr>
<tr>
<td>0.545</td>
</tr>
<tr>
<td>0.566</td>
</tr>
<tr>
<td>0.580</td>
</tr>
<tr>
<td>0.582</td>
</tr>
<tr>
<td>0.602</td>
</tr>
</tbody>
</table>

An observation is a candidate-DMA-week. Cluster-robust standard errors in parentheses (clustered by race). The dependent variable is the candidate’s promotional ad impressions per capita. The candidate-market fixed effect models include a dummy variable for each week prior to the election. Partisanship controls are the fraction of same-party and opposite-party partisans in the DMA. Demographic controls include the market’s population density, fraction of homeowners, age distribution, racial and ethnic composition, education levels, and income distribution, all interacted with a dummy for candidate partisanship (such that, for instance, the fraction of a market’s voters who are of hispanic origin can have different effects on Democratic versus Republican candidates’ vote shares.)

This panel approach deals with confounding from spatial variation — unobserved market characteristics that influence advertising levels — but opens the possibility of confounding due to unobserved temporal variation. While the week dummies in the specification eliminate the most obvious source of such temporal confounding, it remains possible that there are differential time trends that affect some races and not others. For example, suppose candidates in presidential years like to advertise heavily around the time of the national party conventions in August when public attention to politics is high. This would tend to produce upward-biased estimates of the slope of the candidate reaction function estimated within-market. To deal with this possibility, I can flip the direction of conditioning and include candidate-week fixed effects:

$$q^p_{j,k,t,w} = \alpha^p_{k,t,w} + \delta q^A_{j,-k,t,w} + \epsilon^p_{j,k,t,w}$$  \hspace{1cm} (9)$$

The specification in Equation 9 eliminates temporal confounding, now using within-candidate-week variation to identify the slope of the reaction function $\delta$. Columns 3 and 4 show specifications using contemporaneous values of opponent’s attack advertising, whereas columns 5 and 6
use the lagged quantities discussed previously. 3 and 5 exclude market level predictors while 4 and 6 include them. Again, the conclusion is unchanged: the slope of the reaction function is large, positive, and not much affected by the inclusion of market-level covariates that predict advertising levels.

**Discussion**

A growing body of evidence (Huber and Arceneaux, 2007; Gerber et al., 2011; Gordon and Hartmann, 2013; Spenkuch and Toniatti, 2018) demonstrates the existence of advertising effects on vote choices. The existing literature has tended to focus attention on single elections, and particularly presidential elections; my results confirm that this finding generalizes to a broad sample of elections with heterogeneous candidates and offices sought, across the US and over more than a decade. Increases in advertising levels increase vote shares, even holding constant candidate attributes and even using variation from cost shifters plausibly exogenous to market-level unobservables.

This IV approach yields estimates of advertising effects larger than the corresponding OLS, revealing that candidates’ target ads preferentially to markets where they would otherwise do worse than expected given observables. This is consistent with persuasion of marginal voters or demobilization of the opponent’s supporters, but not with mobilization of core supporters, as the primary goal of advertising.

A shortcoming of existing work that this paper seeks to address is the focus on average effects to the exclusion of discussion of mechanisms of influence. Political scientists’ theoretical interest in advertising derives largely from advertising’s potential effects on voters’ participation in or knowledge of politics. Such considerations underlie, for example, Sulkin’s (2011) study of the relationship between congressional campaign communication and legislative behavior in office, or Geer’s (2006) study of the relationship between ad tone and policy content.

The knowledge that ads affect vote shares does not on its own speak to these fundamental questions about the role of advertising in democratic performance. The results in the literature on advertising effects — which this paper confirms — do not discriminate between two possible mechanisms for advertising influence, one that is social welfare improving and one that is, at best, purely redistributive.

The former possibility is that advertising changes vote shares by helping voters to understand the policy issues at stake in an election, to recognize the candidate whose ideological orientation towards politics is closer to their own, or to identify differences in ability or effectiveness between the candidates. If this mechanism is the dominant one, advertising moves votes, on net, in the direction of candidates who are more effective or more ideologically aligned with the median
voter, improving the performance of elections as a selection mechanism.

A second, unhappier possibility is that advertising adds some transient experiential utility to the act of voting for the advertised candidate, with no influence on voters’ knowledge of politics or policy. A persuasive mechanism of this form severs any necessary connection between candidate attributes and advertising effectiveness, moving votes in the direction of whichever candidate succeeds in raising more contributions. Viewed in strict isolation, this raises the chances of election of some candidates who need not be better at the expense of others who need not be worse, a wash from the perspective of voters. Of course, there is an additional cost to consider, in that electoral advantage may be paid for by the diversion of public resources to the donor groups to whom candidates are indebted.

A direct approach to measuring informational effects of ads is difficult, not least because measuring political informedness among voters is hard. Henderson (2016) measures voters’ ability to infer candidates’ partisanship from advertising content, and Prat et al. (2010) ask whether voters could infer state legislators’ effectiveness from the quantity of campaign messaging they are able to deploy. While direct measurement exercises like these are useful, they are inevitably incomplete in capturing what voters might learn from advertising.

I instead use an indirect approach relying on the theories’ divergent implications for the strategic interactions between candidates’ advertising choices. The results, which use a variety of methods to eliminate possible confounding by common influences on advertising choices, are consistently in line with the prediction of the prestige or goodwill model and in contradiction to the prediction of the informational model.

This is not to say that advertising campaigns cannot inform or never inform voters. Kendall et al. (2015), for example, provide experimental evidence that exposure to advertising containing verifiable information can reduce the uncertainty in voters’ beliefs about candidate types. The two mechanisms are not mutually exclusive, and reality may well consist of a mixture of both.

However, the results presented here suggest that the weights in that mixture are not equal. Candidates behave as if the prestige mechanism is the dominant one. To be sure, this conclusion is based on observed equilibrium outcomes, and thus is local to the level of advertising actually observed in the data. At much lower average levels of advertising, the balance between the informational and prestige mechanisms might well be different.

The fact that informational mechanisms of advertising persuasion appear relatively unimportant means that there is limited scope for a selection-improving benefit of advertising to offset its potential costs in bias to public policy and to the representativeness of the pool of candidates for office. Influences of fundraising incentives on policy outcomes do not depend on the existence a direct quid pro quo relationship of contributions to policy favors. Even if contributions are strictly

---

32 As in, e.g., Becker and Murphy’s (1993) theory of advertising as complementary to consumption.
non-instrumental, as argued by Ansolabehere et al. (2003), wealthier people’s tendency to contribute more can still bias policy outcomes in favor of the wealthy (Campante, 2011): a sufficiently unequal distribution of income can make politicians dependent on a small group of very wealthy people for the financial support necessary to remain competitive in elections. Legislators are more responsive to contact from donors than average citizens (Kalla and Broockman, 2016) and as a result have biased perceptions of their districts’ policy preferences (Broockman and Skovron, 2018). And work on the effects of campaign spending caps by Fouirnaies (2018) and Avis et al. (2017) shows that the need to keep up in the fundraising arms race skews the pool of candidates who are willing to enter elections towards those with a fundraising advantage, e.g., the wealthy.

Given all of the known distortions induced by candidates’ need to fundraise, maintaining a belief that the net welfare effects of paid campaign advertising for the median voter are positive requires evidence that its effects on the quality of voters’ political information are large. I find little support for this hypothesis. At the current laissez-faire status quo in US campaign finance regulation, informational losses from a hypothetical binding statutory limit on campaign expenditures would be unlikely to outweigh the gains from reducing these distortions.

References


Appendices

A Proofs

A.1 Effect of incumbency in ad effectiveness in the informational model

In the informational model, incumbency can be modeled as a reduction in the variance (increase in the precision) of the performance signal that the voter observes before exposure to advertising. Hence, the effect of incumbency on the effectiveness of advertising is determined by the derivative of the marginal effect of advertising on voters’ posteriors with respect to the variance of the performance signal.

Given the signal structure assumed, the posterior after receiving the combination of all signals is equivalent to that obtained by updating sequentially, using the posterior from the addition of one signal as the prior for the next (Gelman et al. (2013), p. 42). Hence the marginal effect of the candidate’s last positive ad airing is proportional to:

\[
\left( \Sigma_1^{-1} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p^2} & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p^2} & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]

Where \( \Sigma_1 = (\Sigma_0^{-1} + \Phi^{-1})^{-1} \). Taking the derivative of this quantity with respect to \( \sigma_\xi^2 \) and applying matrix calculus identities, we get:

\[
\frac{\partial^2 m_{j,k}^\xi}{\partial q_{j,k}^p \partial \sigma_\xi^2} = -\Sigma_1 \begin{bmatrix} -\frac{1}{\sigma_\xi^2} & 0 & 0 \\ 0 & \Sigma_1^1 & 0 \\ 0 & 0 & \Sigma_1^1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p^2} & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]

\[
= \frac{\Sigma_1^{(1,2)}}{\sigma_\xi^2 \sigma_p^2} \begin{bmatrix} 0 & \Sigma_1^{(1,1)} & 0 \\ 0 & \Sigma_1^{(1,2)} & 0 \\ 0 & \Sigma_1^{(1,3)} & 0 \end{bmatrix}
\]

Thus the derivative w.r.t \( \sigma_\xi^2 \) of the marginal effect on voters’ posterior of the candidate’s type is

\[
\frac{\left( \Sigma_1^{(1,2)} \right)^2}{\sigma_\xi^2 \sigma_p^2} > 0.
\]

Hence the effectiveness of advertising is declining, the higher the precision (lower the variance) of the quality signal.
A.2 Cross-partials in the informational model

The cross-partial derivative of candidate $k$’s share in market $j$, $s_{j,k}$, with respect to her own promotional advertising $q_{j,k}^P$ and her opponent’s attack advertising $q_{j,k}^A$ is:

$$\frac{\partial^2}{\partial q_{j,-k}^A \partial q_{j,k}^P} s_{j,k} = \frac{(1 + e^{u_{j,-k}}) e^{u_{j,-k}}}{(1 + e^{u_{j,-k}} + e^{u_{j,k}})^2} \left[ \frac{\partial^2 m_{j,k}^x}{\partial q_{j,-k}^A \partial q_{j,k}^P} + \frac{\partial m_{j,k}^x}{\partial q_{j,-k}^A} \frac{\partial m_{j,k}^x}{\partial q_{j,k}^P} \left( 1 - e^{u_{j,k}} + e^{u_{j,-k}} \right) \right]$$

$$= s_{j,k} \left( 1 - s_{j,k} \right) \left[ \frac{\partial^2 m_{j,k}^x}{\partial q_{j,-k}^A \partial q_{j,k}^P} + \frac{\partial m_{j,k}^x}{\partial q_{j,-k}^A} \frac{\partial m_{j,k}^x}{\partial q_{j,k}^P} (1 - 2s_{j,k}) \right]$$

The second term inside the brackets in the above is negative whenever $s_{j,k} < \frac{1}{2}$ and both candidates are willing to advertise (i.e., $\frac{\partial m_{j,k}^x}{\partial q_{j,k}^P} > 0$ and $\frac{\partial m_{j,k}^x}{\partial q_{j,-k}^A} < 0$). The first term inside the brackets can be expanded as:

$$\frac{\partial^2 m_{j,k}^x}{\partial q_{j,-k}^A \partial q_{j,k}^P} = \frac{\partial}{\partial q_{j,-k}^A} \left( \sum_1 - \left[ \frac{1}{\sigma_x^2} \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 1 + qp & 0 \\ 0 & 0 & q_A \\ \end{array} \right] \right)^{-1} \left[ \begin{array}{ccc} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p} & 0 \\ 0 & 0 & 0 \end{array} \right]$$

$$= \frac{\partial}{\partial q_{j,-k}^A} \left( \sum_1 \begin{array}{ccc} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p} & 0 \\ 0 & 0 & 0 \end{array} \right)$$

$$= -\sum_1 \left[ \begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_x^2} & 0 \end{array} \right] \sum_1 \left[ \begin{array}{ccc} 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_p} & 0 \\ 0 & 0 & 0 \end{array} \right]$$

$$= -\frac{\sum_1^{(2,3)}}{\sigma_x^2 \sigma_p} \left[ \begin{array}{ccc} 0 & 0 & 0 \\ 0 & \sum_1^{(1,3)} & 0 \\ 0 & 0 & \sum_1^{(2,3)} \end{array} \right]$$

Hence the sign of the derivative of $\frac{\partial}{\partial q_{j,k}^A}$ with respect to the opponent’s advertising level $q_{j,-k}^A$ is the opposite of the sign of $\sum_1^{(2,3)} \sum_1^{(1,3)}$. We expect the posterior $\sum_1^{(1,3)}$ to be negative in equilibrium (otherwise it would not make sense for the opponent to run negative ads) and hence the sign depends on the sign of $\sum_1^{(2,3)}$. Some tedious algebra shows that
\[ \Sigma^{(2,3)}_1 = (\Sigma^{(2,3)}_0)(\sigma^2_\xi + \Sigma^{1,1}_0) - \Sigma^{(1,2)}_0 \Sigma^{(1,3)}_0 / M \]

Where \( M = \frac{\sigma^2_0 \sigma^2_\xi \det \Sigma_0}{\det \Sigma_1} \), which is greater than 0 by positive semi-definiteness of \( \Sigma_0 \) and \( \Sigma_1 \). Thus, the second derivative is negative whenever

\[ \Sigma^{(2,3)}_0 < \frac{\Sigma^{(1,2)}_0 \Sigma^{(1,3)}_0}{\sigma^2_\xi + \Sigma^{1,1}_0} < 0 \quad (10) \]

The constraint approaches zero as the sum of the prior variance on candidate quality and the variance of the performance signal become large, e.g. in cases where advertising makes sense.

### A.3 Effect of incumbency on ad effectiveness in the prestige model

In the prestige / goodwill model, incumbency can be modeled as an increase in the candidate’s initial stock of goodwill prior to voters’ exposure to advertising. Hence, the effect of incumbency on the effectiveness of advertising is determined by the derivative of the marginal effect of advertising on vote probabilities with respect to the initial goodwill stock. Given the contest structure assumed, this is simple to calculate:

\[
\frac{\partial}{\partial s^0_{j,k}} \left( \frac{\partial p_{j,k}}{\partial q^p_{j,k}} \right) = \frac{(1 + g(g^0_{j,-k} + q^p_{j,k} - q^A_{j,k})) g'(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})}{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k}) + g(s^0_{j,-k} + q^p_{j,-k} - q^A_{j,-k}))^2} \cdot \frac{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})) g''(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})}{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k}) + g(s^0_{j,-k} + q^p_{j,-k} - q^A_{j,-k}))^2} - \frac{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})) g'(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})}{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k}) + g(s^0_{j,-k} + q^p_{j,-k} - q^A_{j,-k}))^2} \cdot \frac{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})) g''(s^0_{j,k} + q^p_{j,k} - q^A_{j,k})}{(1 + g(s^0_{j,k} + q^p_{j,k} - q^A_{j,k}) + g(s^0_{j,-k} + q^p_{j,-k} - q^A_{j,-k}))^3} < 0
\]

### A.4 Cross-partial derivatives on ad quantity in the prestige model

Similarly, cross-partial derivatives on ad quantities, though cumbersome to write out, have unambiguous signs:
\[
\frac{\partial p_{j,k}}{\partial q_{j,k}} = \frac{(1 + g(g_{j,-k}^0 + q_{j,k}^p - q_{j,k}^A))g'(g_{j,k}^0 + q_{j,k}^p - q_{j,k}^A)}{(1 + g(g_{j,k}^0 + q_{j,k}^p - q_{j,k}^A) + g(g_{j,-k}^0 + q_{j,k}^p - q_{j,k}^A))^2} > 0
\]
\[
\frac{\partial p_{j,k}}{\partial q_{j,-k}} = \frac{-(1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A))g'(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A)}{(1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A) + g(g_{j,-k}^0 + q_{j,-k}^p - q_{j,-k}^A))^2} < 0
\]
\[
\frac{\partial^2 p_{j,k}}{\partial (q_{j,k})^2} = -\frac{2(1 + g(g_{j,k}^0 + q_{j,k}^p - q_{j,k}^A))}{(1 + g(g_{j,k}^0 + q_{j,k}^p - q_{j,k}^A) + g(g_{j,-k}^0 + q_{j,k}^p - q_{j,k}^A))^3} + \frac{g''(g_{j,k}^0 + q_{j,k}^p - q_{j,k}^A)}{1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,k}^A)} < 0
\]
\[
\frac{\partial^2 p_{j,k}}{\partial (q_{j,-k})^2} = -\frac{2(1 + g(g_{j,-k}^0 + q_{j,k}^p - q_{j,-k}^A))}{(1 + g(g_{j,k}^0 + q_{j,k}^p - q_{j,-k}^A) + g(g_{j,-k}^0 + q_{j,k}^p - q_{j,-k}^A))^3} + \frac{g''(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A)}{1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A)} < 0
\]
\[
\frac{\partial^2 p_{j,k}}{\partial q_{j,k} \partial q_{j,-k}} = \frac{2(1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,k}^A))}{(1 + g(g_{j,k}^0 + q_{j, - k}^p - q_{j,-k}^A) + g(g_{j,-k}^0 + q_{j,-k}^p - q_{j,-k}^A))^3} - \frac{g''(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A)}{1 + g(g_{j,k}^0 + q_{j,-k}^p - q_{j,-k}^A)} > 0
\]

\section*{B Data Details}

\subsection*{B.1 Advertising data}

Data on advertising quantity (measured in impressions per 1000 voting-age citizens) come from the Wesleyan Media Project / Wisconsin Advertising Project (Goldstein and Rivlin, 2005, 2007; Fowler et al., 2014, 2015, 2017a), henceforward "WMP", for 2002-2008 and from Nielsen for 2010-2014. Ad tone comes from WMP in all years. I focus on ads run in Senate and gubernatorial general election campaigns by the general election candidates from the two major parties. I identify general election ads using the date of the state’s primary election; any ads run after that date by the general election candidates are considered to be general election ads.

I exclude a handful of races for idiosyncratic reasons: 2002 and 2004 senate races in Louisiana, where multiple candidates from the same major party competed in a “jungle primary” on the general election date, and the 2002 Minnesota Senate race, where the Democratic incumbent, Paul Wellstone, died in a plane crash 11 days prior to the election. Summary statistics for the remaining races are in Table 7 below.

33
There are a few choices that must be made to map the raw advertising data onto the model. I describe these below.

**Contrast ads** The WMP classifies ads according to the main subject of the advertisement: ads may promote the sponsoring candidate, or attack the opposing candidate. The WMP also includes an intermediate third category, “contrast” ads, which fall somewhere in between. WMP also includes a sub-categorization of ads as “more promote than attack,” “about equal promote and attack,” “more attack than promote,” or “only contrasting element is statement of candidate endorsement.”

I apportion these ads to one or the other of the primary two categories, using a simple weighting scheme. I assign .75 of the impression quantity of Contrast ads to the “promote” category and .25 to the “attack” category if the WMP defines them as “more promote than attack,” and the reverse if they are described as “more attack than promote.” Impressions of contrast ads labeled “about equal promote and attack” are split equally between “promote” and “attack” categories. The remaining “only contrasting element….” ads are treated as 100% attack impressions.

**Quantities** Estimating effects of advertising requires a measure of the quantity of advertising that a candidate allocated to a given media market. The obvious measure of quantity - the total number of ad spots purchased - is inappropriate because, in general, spots are not a homogeneous commodity. For instance, a thirty second ad purchased on a highly-rated prime-time sitcom may be seen by an order of magnitude more viewers than the same thirty second ad run on a daytime soap opera.

The quantity that we want to measure is not seconds of advertising but **impressions**: the total number of viewers who can be expected to have seen the ad. Nielsen data (covering 2010-2014) include an estimate of impressions defined at the spot level. WMP data, which I use prior to

<table>
<thead>
<tr>
<th></th>
<th>Governor</th>
<th>Senate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cycles</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Number of Races</td>
<td>117</td>
<td>148</td>
</tr>
<tr>
<td>Number of Candidates</td>
<td>234</td>
<td>296</td>
</tr>
<tr>
<td>Number of Markets</td>
<td>300</td>
<td>323</td>
</tr>
<tr>
<td>Number of Market-Years</td>
<td>741</td>
<td>997</td>
</tr>
</tbody>
</table>

Table 7: Counts of Races, Candidates, Markets in Final Data Set.
2010, does not record impressions directly. For these years impressions must be imputed from the WMP data using the method described by Gordon and Hartmann (2013) and Gordon and Hartmann (2016), which I follow. Because television advertising is sold by the ratings point,\textsuperscript{33} dividing the total cost of the spot (recorded by the WMP) by the per-point price (collected from the Spot Quotations and Data (SQAD) Media Market Guides) and then scaling by the adult population of the media market yields an estimate of the number of viewers the spot reached. Posted prices on a per-1000-impressions basis are available by media market, program type,\textsuperscript{34} and day part,\textsuperscript{35} yielding a reasonably fine-grained match from advertising observations to the price used to compute quantity in units of impressions.

Table 8 shows summary statistics for advertising quantities, split by tone (positive or negative). Ads are measured in 1000’s of impressions per capita; an observation is a candidate-year-race-DMA.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attack</td>
<td>0.000</td>
<td>0.014</td>
<td>0.586</td>
<td>0.035</td>
</tr>
<tr>
<td>Promote</td>
<td>0.000</td>
<td>0.014</td>
<td>0.384</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table 8: Summary Statistics: Ad Quantities.

B.2 Vote data

I collected voting data for the 265 Senate and gubernatorial races in the sample from Congressional Quarterly’s Voting and Elections Collection,\textsuperscript{36} which gathers county-level election results for all federal and gubernatorial races.

I then aggregated voting results to the state-DMA level using a zipcode-to-DMA mapping file from Nielsen. DMA boundaries are usually drawn to coincide with county borders but when, on occasion, DMA boundaries split a county, I allocated the county’s voting results to each DMA in proportion to the fraction of the county’s population residing within each DMA, according to the zip-level DMA assignments.

\textsuperscript{33}One ratings point represents 1\% of the TV-watching population of the media market.

\textsuperscript{34}Different prices prevail for news and non-news programs at the standard evening (5PM) and late (11PM) news slots.

\textsuperscript{35}There are seven day parts, ranging from the relatively cheap daytime (9AM-4PM) to the most expensive prime-time (8PM-11PM) slots.

\textsuperscript{36}http://library.cqpress.com/elections/
The number of DMAs in a state varies with the geographic and population size of the state. The largest and most populous states have the most DMAs within their borders: the state with the most markets is Texas, with 18. Some of the smallest states, like Rhode Island, have only one. Because I rely on within-candidate variation, I limit the dataset to states with at least two media markets. In 2002 and 2004, WMP data include only the top 100 media markets (by population) in the country. In these years I limit to states with at least two media markets among the top 100. In the remaining years, voting and advertising data covers all 210 DMAs and hence the restriction is only to states with 2 markets of any size.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Share of Voting Age Population</td>
<td>0.019</td>
<td>0.164</td>
<td>0.525</td>
<td>0.068</td>
</tr>
<tr>
<td>Share of Votes Cast</td>
<td>0.080</td>
<td>0.483</td>
<td>0.899</td>
<td>0.144</td>
</tr>
<tr>
<td>Two-Party Vote Share</td>
<td>0.082</td>
<td>0.500</td>
<td>0.918</td>
<td>0.148</td>
</tr>
<tr>
<td>Log Vote Share Difference</td>
<td>-3.754</td>
<td>-1.474</td>
<td>0.483</td>
<td>0.566</td>
</tr>
</tbody>
</table>

Note: Log Vote Share Difference is \( \log s_{jk} - \log s_{j0} \), where \( s_{jk} \) is the candidate’s share of voting age population in market \( j \) and \( s_{j0} \) is the share of voting age population who did not cast a vote in market \( j \).

Table 9: Summary Statistics: Vote Shares.

The fundamental variation that my estimator uses to fit the model’s parameters is differences in vote shares across markets within a typical statewide race. Figure 1 shows the density of demeaned vote shares - i.e., each candidate’s vote share in each market within a state, minus that candidate’s average share in the entire state. The typical race has significant cross-market variation: the standard deviation is approximately 3.5 percentage points.

### B.3 Demographic and partisanship data

I collected demographic data at the zip code level from the US Census and American Community Survey using datasets produced by the Minnesota Population Center (2011). These were aggregated to the state-DMA level using population-weighted averages. The set of variables included in models where the “Demographics:” row indicates a Y are the candidate’s party interacted with:

1. Percent of residents who own a home
2. Percent of residents aged 15-34; 35-54; 55-69; 70+
3. Percent of residents who are Black; percent Hispanic
Figure 1: Density of Within-Candidate Deviations from Candidate-Mean Vote Share.

4. Percent of residents with at most a high school diploma; an associate’s degree; some college; a bachelor’s degree; graduate degree

5. Percent of residents with income between $25-49K; $50-99K; $100-199K; $200K+

I interact demographics with candidates’ party affiliation (R or D) to allow for variation in party preference across demographics. E.g., a state-DMA with a large population of wealthy residents has different implications for Democratic than for Republican candidates.

I augmented the census data with information on voters’ stated party identification from a combination of the National Annenberg Election Surveys (NAES) conducted in 2000, 2004, and 2008, and the Cooperative Congressional Election Study (CCES) conducted in 2008-2014.37 Respondents in CCES / NAES were matched to state-DMA by their zip code of registration.

A challenge here is that some of the smallest DMAs, and many state-DMA combinations, have low populations (e.g., about 35,000 Alabama residents are in the Atlanta DMA). These geographies have accordingly low, and sometimes no, residents included in the CCES / NAES, which typically surveyed on the order of 50,000 respondents out of the entire US population.

I deal with this by augmenting the CCES / NAES with individual survey data from the Census Bureau’s Current Population Survey (CPS), which has larger sample size. CPS does not ask about

---

37Because there were no survey data available in 2002, I used the 2000 survey values for the 2002 elections.
voter partisanship, so I project partisanship using a linear probability model of partisan affiliation as a function of demographics and DMA fixed effects estimated in the CCES / NAES. I compute the average value of the combined (actual in NAES / CCES plus predicted in CPS) partisanship variables of all respondents within a given election year, state, and DMA. Models with a ‘Y’ in the “Partisanship:” column include two variables containing the fraction of Democratic party identifiers and Republican party identifiers, respectively, in the year-state-DMA.