A Framework for Dynamic Causal Inference in Political Science [*]

Matthew Blackwell[†]

May 6, 2011

Abstract

Traditional single-shot causal inference models investigate the effect of a single action at a single point in time and are an invaluable tool for political scientists. Often, however, actions unfold over time, with political entities reacting to a shifting environment. Accordingly, single-shot methods leave researchers unable to extract meaningful causal inferences about these dynamic processes. This stems from a fundamental tension: in dynamic settings, regression and matching force a choice between omitted variable bias on the one hand, and post-treatment bias on the other and are unable to simultaneously correct for both. To avoid these problems, I introduce a framework for dynamic causal inference and utilize marginal structural models to estimate dynamic causal effects. The effectiveness of "going negative" serves as a motivating example—an apt illustration since candidates change their strategy as the campaign unfolds. Furthermore, I introduce novel diagnostics and a sensitivity analysis for the model.

[*]The draft in front of you has seen marked improvement from comments by Steve Ansolabehere, Adam Glynn, Justin Grimmer, Luke Keele, Gary King, James Robins, Brian Schaffner, Maya Sen, Elizabeth Stuart, Jonathan Wand, and participants at the Murno Seminar Series at Stanford University. The errors that remain are my own responsibility. Previous versions were presented at the 2010 Annual Meeting of the American Political Science Association (Washington DC) and the 2011 Annual Midwest Political Science Association Conference (Chicago, IL).

[†]Department of Government, Harvard University. 1737 Cambridge St K351, Cambridge, MA 02138. web: http://www.people.fas.harvard.edu/~blackwel/ email: mblackwell@iq.harvard.edu

§1 Introduction

Political science has seen enormous growth in attention to causal inference over the past decade. This heightened focused on causal issues has lead to a more careful and thoughtful treatment of causality in applied research. These advances have, however, heavily focused on snapshots—studies that observe units at one or two points in time. A snapshot, while often a forceful way to tell a story, imposes an artificial and unwarranted stability to politics. As political science finds itself with a growing number of motion pictures—panel data, time-series cross-sectional data—a tension has emerged between substance and method. Indeed, applied to dynamic data, the best practices of these *single-shot* causal inference methods provide conflicting advice and fail to alleviate omitted variable or post-treatment bias.

This paper focuses on a specific dynamic process: negative advertising in U.S. statewide elections. Candidates change their tone over the course of the campaign, reacting to their current environment. A single-shot causal inference method would compare campaigns that are similar on a host of pre-election variables in order to eliminate omitted variable bias. While this is often the best approach with single-shot data, such an approach ignores the fundamentally dynamic nature of campaigns: races that become close over the course of the campaign are more likely to go negative than those that are safe. Attempting to correct for this dynamic selection by controlling for polls leads to post-treatment bias since earlier campaign tone influences polling. The inappropriate application of single-shot causal inference leaves scholars between a rock and hard place, steeped in bias with either approach.

To help scholars out of this awkward position, this paper introduces a formal model of dynamic causal inference along with marginal structural models (MSM) developed by Robins, Hernán and Brumback (2000) to estimate dynamic causal effects. Dynamic causal inference models overcome the above problems by directly modeling dynamic selection. Actions (such as campaign tone) are allowed to vary over time along with any confounding covariates (such as polling). Thus, we can study the effects of the *action history*—a candidate's tone across the entire campaign—as opposed

to a single action—simply "going negative." The marginal structural model dictates the form of the relationship between the large number of possible action histories and the outcome.

A model for the outcome cannot single-handedly overcome the dynamic selection problem, but under certain assumptions inverse probability of treatment weighting (IPTW) removes both of the above biases. With this approach, we weight each unit by the inverse of the estimated probability of its observed action history. This weighting creates a pseudosample where dynamic selection is eliminated, circumventing the dilemmas posed by single-shot causal inference. Since these methods require strong assumptions, this paper also develops a novel diagnostic tool, the history-adjusted balance, and describes a sensitivity analysis framework to catch potential causes of concern.

The paper proceeds as follows. Section 2 reviews traditional causal inference models that focus on a single-shot action. Section 3 introduces the major concepts involved in dynamic causal inference. Section 4 introduces marginal structural models and describes how to use them to estimate causal effects. Section 5 applies the techniques to estimating the effectiveness of "going negative" in campaigns. Section 6 discusses useful diagnostics and a sensitivity analysis framework for marginal structural models. Section 7 concludes with directions for future research.

§2 Single-shot causal inference

By far the most common causal inference models in political science are *single-shot* models. The goal of these approaches is to estimate the effect of a single action, A_i, on an outcome, Y_i, at a single point in time. In a population with a typical unit i, A_i takes the value 1 if unit i takes the action and 0 otherwise.[1] If the unit takes the action we say that it is in the *treatment group*, otherwise it is in the *control group*. With the example of campaigns, we might be interested in the effect of a Democratic candidate i running a negative campaign, $A_i = 1$, or a positive one, $A_i = 0$, on their share of the two-party vote, Y_i.

[1] Actions here are synonymous with *treatments*, a more common term in the causal inference literature. While we use the active voice in this article, the unit may or may not be the decision-maker with regard to the action. Sometimes the action is taken by the unit, other times, others perform the action on the unit.

In the tradition of Rubin (1978) and Holland (1986), we define causal effects in terms of counterfactuals or potential outcomes. Specifically, $Y_i(1)$ is what unit i's outcome would have been if she had taken the action. With negative advertising, this is Democrat i's share of two-party vote if she ran a negative race. Of course, each unit has a corresponding potential outcome for not taking the action, $Y_i(0)$, which is similarly defined. The *individual causal effect* for unit i, then, would simply be $Y_i(1) - Y_i(0)$. It is the difference between what would happen under action and what would happen under inaction. Note that the causal effects are defined in terms of potential outcomes, not observed quantities.

To draw the connections between the observed data and the potential outcomes, we make use of a *consistency* assumption: when we observe a unit taking an action, we observe their potential outcome for that action. Mathematically, we write this as

$$Y_i = Y_i(a) \qquad \text{if } A_i = a, \tag{1}$$

where a is either 1 or 0. It is important to note that the two variables in (1) are distinct. Whereas Y_i is the observed value of the outcome in the data, $Y_i(a)$ is a hypothetical value of the outcome if we had forced unit i to take a. These two need not be the same if the potential outcomes depend on the actions of other units. If two Senate campaigns were occurring at the same time and in the same state, for instance, one candidate's negativity may influence the potential outcomes of the other race since they share a media market and electorate. The consistency assumption directly connects the observed and potential outcomes, ignoring any of these possible spillover effects.[2]

Consistency implies that we observe only *one* of the potential outcomes for any individual. A candidate running for office has either gone negative or she has not. We cannot simultaneously observe a unit's outcome under both action and inaction. This is commonly known as the fundamental

[2] Rubin (1978) refers to this assumption as the stable unit treatment value assumption or SUTVA. It may seem that consistency is a definition rather than an assumption, but a few recent studies have called that assertion into question (Cole and Frangakis, 2009; VanderWeele, 2009; Pearl, 2010).

problem of causal inference and makes individual causal effects difficult to estimate without strong assumptions. We can estimate other quantities such as the average treatment effect,

$$\text{ATE} = \mathrm{E}\big[Y_i(1) - Y_i(0)\big] = \mathrm{E}\big[Y_i(1)\big] - \mathrm{E}\big[Y_i(0)\big], \tag{2}$$

where the expectations are over units. The ATE is the difference between the average outcome if everyone in the population acted and the average outcome if no one in the population acted. These are population-wide questions—for example, what would happen if *every* Democrat went negative?—but the consistency assumption only helps us for subsets of the population—for example, what happened to the Democrats who actually went negative? That is, consistency implies

$$\mathrm{E}\big[Y_i(1)|A_i = 1\big] = \mathrm{E}[Y_i|A_i = 1], \tag{3}$$

where $\mathrm{E}[Y_i|A_i = 1]$ is the observed average outcome among those who have acted. But this implies nothing about the counterfactual quantity $\mathrm{E}[Y_i(1)|A_i = 0]$, which is the what the average outcome would be for non-actors if they had acted. Consistency alone cannot tell us, even on average, what would have happened to negative Democrats if they had run a positive campaign.

To overcome this problem and estimate the ATE, we need the further assumption of *conditional ignorability*. This assumption states that the potential outcomes are independent of the action, conditional on a set of covariates, X_i, which we write as:

$$\big(Y_i(1), Y_i(0)\big) \perp\!\!\!\perp A_i \mid X_i. \tag{4}$$

where $\perp\!\!\!\perp$ indicates conditional independence.[3] Political scientists call this assumption *no omitted variables*, economists call it *no selection on unobservables*, and epidemiologists call it *no unmeasured confounders*. In words, the assumption means that the distribution of the potential outcomes is the

[3] To be precise, $A \perp\!\!\!\perp B|C$ is defined to mean that $f(A|B, C) = f(A|C)$, where f is the conditional density of A (Dawid, 1979).

same for those who have and have not taken the action. Ignorability would be violated, for instance, if incumbent Democrats were more likely to stay positive than challenger Democrats and we failed to control for incumbency status. In this case, the decision to go negative would be correlated with potential outcomes, since incumbents were likely to run positive campaigns and also likely have higher potential election outcomes than challengers. Augmenting consistency with ignorability allows us to fully connect the observed data and the potential outcomes:

$$\mathrm{E}[Y_i|A_i = 1] = \mathrm{E}[Y_i(1)|A_i = 1] = \mathrm{E}[Y_i(1)], \tag{5}$$

where the first equality comes from consistency and the second comes from ignorability. Thus, ignorability allows us to connect observed outcomes in observed subgroups ($\mathrm{E}[Y_i|A_i = 1]$) to counterfactual outcomes about the entire population ($\mathrm{E}[Y_i(1)]$). If we apply the same logic to $Y_i(0)$, then it becomes clear that the ATE can be calculated using observable quantities in the data.

Ignorability is, however, an untestable assumption. Random assignment of an action guarantees it will hold, but in observational studies it must be justified based on subject-matter knowledge. In order to ensure ignorability holds, the conditioning set, X_i, must include any variable that (a) is a cause of, or shares a common cause with, the action, (b) causally affects the outcome, and (c) is not affected by the action. We call these variables relevant omitted variables or confounders. Note that problematic omitted variables are *pre-action variables* in the sense that they are causally prior to the action. That is, they either affect the treatment or they share a common cause with the action, but they are never affected by the action. We avoid controlling for *post-action* variables because doing so can induce bias in estimating the causal effect. This is known in the causal inference literature as *post-teatment bias* (Ho et al., 2006).

There are two, related sources of post-treatment bias. First, conditioning on post-action variables can "block" part of the action's overall effect. For instance, suppose a researcher controlled for polling results from the day of the election when attempting to estimate the effect of incumbency. This will

understate the effect of incumbency since most of the effect flows through the standing of candidates late in the race. Second, conditioning on a post-action variable can induce selection bias even when no bias exist absent the conditioning. For instance, suppose at the start of campaign we randomly assigned high and low budgets to different Democratic candidates for Senate. If we condition on the polls sometime during the campaign, we can seriously bias our estimates of the effect of campaign budgets. Those leading Democrats who had high budgets are likely to differ strongly from leaders with small budgets. For example, if higher budgets help a candidate, then those low-budget leaders are actually much stronger candidates than the high-budget leaders, since they were able to lead in the polls without the additional funding. Thus, comparing high- and low-budget leaders would give a misleading estimate of the causal effect of campaign finance, even though it was randomly assigned.

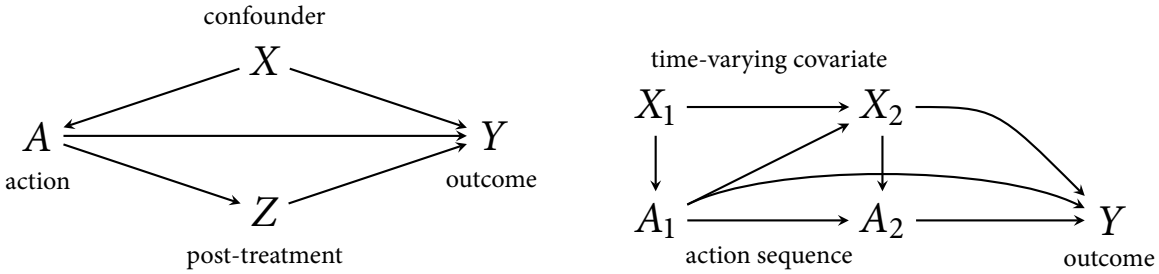
§3 Dynamic causal inference

The above single-shot framework is perhaps useful for experiments and observational studies where the action of interest occurs once. Yet, there are many situations in political science where actions are evolving over time and reacting to the current state of affairs. To handle this rich class of empirical situations, we need a framework that explicitly incorporates time. This section gives a general introduction to such a framework.

3.1 Time-varying actions

Here, the strategy of negative advertising provides a useful illustration. A campaign can "go negative" at multiple points over the course of the campaign. Perhaps a candidate attacks early, before their opponent has a footing, or perhaps she runs negative ads late, responding to smear tactics. These two situations, as far apart as they are, would both register as "going negative" in a single-shot model. Furthermore, it is unclear what the relevant pre-action omitted variables would be for these campaigns. The opponent's negativity and polling numbers appear to be post-action for the early attacker, but would be pre-action for the late responder. This dynamic application fits, at best, awkwardly into the single-shot framework.

In order to sort out these and other problems, it is important to explicitly incorporate the dynamic structure of the data into the causal framework. This generalization flows from a simple subscript, t, which indicates *when* actions take place. Single-shot causal inference models ignore t and implicitly assume that all actions occur at once. This is an acceptable framework for many problems because actions really do occur once. When actions unfold over time, however, the incorporation of t and its implications become necessary. Dynamic causal inference allows the actions to vary over time, A_{it}. Elections are no longer simply positive or negative; they can shift their tone from week to week. Figure 1b shows a simple dynamic causal inference model where actions can vary from period 1 to period 2, as opposed to Figure 1a where there is only a fixed action.



(a) Single-shot causal inference.

(b) Dynamic causal inference.

Figure 1: Directed Acyclic Graphs showing single-shot and dynamic causal inference frameworks. Each arrow represents a causal relationship.

A key point is that allowing for this additional flexibility comes at a price, as we will now have far greater than two groups to compare. In each round, units can either take the action ($A_{it} = 1$) or not ($A_{it} = 0$), resulting in a sequence of action decisions. We call each of these possible histories an *action sequence*. If T is the number of periods, then there would be 2^T possible action sequences. For instance, according to Figure 1b, each campaign would follow one of the following sequences: $\{(0,0),(0,1),(1,0),(1,1)\}$. As with single-shot models, each sequence has an associated potential outcome. The outcome for candidate i if we forced him to follow the "always negative" sequence would be $Y_i(1,1)$. Instead of estimating a single "causal effect," we compare some combination of these counterfactual quantities. One such contrast is the expected difference between "always negative" and

"always positive":

$$\mathrm{E}[Y_i(1,1) - Y_i(0,0)]. \tag{6}$$

This would be the effect on the average Democratic vote share due to a move from all Democrats always going negative to all Democrats always staying positive. These simple differences in means are useful for two time periods, but for campaigns of over ten weeks there will be few candidates that follow any particular action sequence. Below, we build a model for the relationship between the action and the potential outcome to alleviate this curse of dimensionality.

3.2 Time-varying confounders cause both omitted- and included-variable bias

Dynamic settings create feedback loops where past actions affect confounders that themselves affect future actions and the outcome. These variables that change over time, reacting to and affecting the action sequence, are *time-varying confounders*. In Figure 1b, X_2 both affects A_2 and is affected by A_1, while having a separate effect on the outcome, Y. With negative advertising, polls affect the decision to go negative, but going negative affects subsequent polls and, ultimately, the election.

Figure 1b highlights a fundamental problem with the standard statistical advice, discussed in Section 2, that researchers should control for pre-action variables and omit any post-action variables. The time-varying confounder X_2 is, however, both pre-action for A_2 and post-action for A_1. Thus, the traditional rules dictate that we both control for X_2 *and* omit X_2 in any analysis. Thus, with dynamic data, standard single-shot techniques force researchers to choose between omitted variable bias on the one hand and post-treatment bias on the other. One might hope that the effects estimated from each approach would bound the true effect. Unfortunately, this need not be the case and, in fact, the true effect can have the opposite sign as either of the two recommended methods.

Regression and matching, two common causal inference methods, cannot avoid this tension, since they are explicitly conditional (Robins, 2000). That is, in their own ways, they estimate effects within groups of similar units and then average over those subgroup effects. And yet, grouping units by a time-varying confounder actually creates *dissimilar* groups. For the reasons described in

Section 2, if a positive campaign and a negative campaign are both trailing with a month before the election, they likely have very different potential electoral outcomes.

There are techniques to remove the omitted variable bias without inducing any post-treatment bias. Inverse probability of treatment weighting (IPTW) and g-estimation are two methods to adequately control for time-varying confounders (Robins, 2000). While g-estimation and the related "structural nested models" can estimate more flexible causal quantities of interest than IPTW, they require a larger modeling burden. In fact, they require modeling and integrating over the distribution of the time-varying confounders, X_1 and X_2, about which we often have very little knowledge. Thus, this paper takes an IPTW approach to dynamic causal inference.[4]

§4 Marginal structural models

This paper focuses on the example of estimating the effect of campaign advertising tone on electoral outcomes for Democratic candidates, but the methodology developed here generalizes easily to other contexts.[5] Suppose i indexes the campaign, with $i = 1, \ldots, N$. Let t denote the week of the campaign, taking possible values $1, \ldots, T$, where T is the final week before election day. We refer to $t = 1$ as the "baseline" time period; it is the time period before the campaign begins, assumed to be the first week after the primary. In each period, campaigns can either go negative, denoted $A_{it} = 1$ or remain positive, $A_{it} = 0$.

Campaigns face a rapidly evolving environment. To account for this, let X_{it} represent the characteristics of the campaign in week t that affect the Democrat's decision to go negative in week t. This would include recent polling or Republican negativity in the previous weeks. This definition assumes that the decision to go negative occurs "after" the variables in X_{it}, so that they are pre-action for week t.[6] Instead of containing all variables occurring at time t, the set of covariates describes the informa-

[4] For a more detailed discussion of the advantages and disadvantages of IPTW approaches versus g-estimation and structural nested models see Robins (2000).

[5] The following section rests heavily on the potential outcomes-based model of causal inference championed by Rubin (1978) and extended to dynamic settings by Robins (1986, 1997).

[6] The causal ordering here is notationally arbitrary as its reversal would require only a change in subscript. More crucially, researchers must determine what information is pre- and post-action in a given period for the substantive

tion setting for the action decision at time t. Simply put, X_{it} is the most recent set of variables that could possibly affect A_{it}. The baseline covariates, X_{i1}, include background information that remains static over the course of the study. For campaigns, these could be perceived competitiveness of the election, number of ads shown in the primary, incumbency status, or challenger quality. The choice of relevant covariates of course depends on the outcome, Y, which in this case is the Democratic percent of the two-party vote.

Dynamic settings require references to the *history* of a variable. A history is the set of all instances of that variable up to some point in time. In this example, it may be the sequence of campaign tone or poll results in each week. Underlines indicate the history of a variable, so that $\underline{A}_t$ would be the negativity up through time t. Formally, this is:

$$\underline{A}_t \equiv (A_1, A_2, \ldots, A_t) \qquad \forall t \in \left[1, T\right] \tag{7}$$

$$\underline{X}_t \equiv (X_0, X_1, \ldots, X_t) \qquad \forall t \in \left[1, T\right] \tag{8}$$

One possible realization of $\underline{A}_t$ is $\underline{a}_t \equiv (a_1, \ldots, a_t)$, where each a_t can take the values 0 or 1. Furthermore, let $\underline{A} = \underline{A}_T$ be the sequence of negativity over the course of the entire campaign. Let $\underline{a}$ be a representative campaign tone history and $\underline{\mathcal{A}}$ as the set of all possible values of $\underline{a}$; that is, all the possible ways a candidate could go negative over the course of the campaign. Let $\underline{X}$, $\underline{x}_t$, and $\underline{x}$ be defined similarly for the covariate history.

Each possible negativity sequence, $\underline{a}$, has an associated potential electoral outcome. Let $Y_i(\underline{a})$ be the Democratic percent of the two-party vote if we forced candidate i to implement the campaign $\underline{a}$. Note that there are 2^T possible sequences $\underline{a}$. As before, any individual candidate can experience at most one of these potential outcomes, which is the one associated with the their observed action history. The rest of the potential outcomes will be counterfactual; they are *what would have happened* if the unit had followed a different sequence. Suppose campaigns only lasted two weeks. In this world,

question at hand.

$Y_i(0,1)$ would be the Democratic vote-share if candidate i were to remain positive in week one and go negative in week two.

We say that $\underline{X}_t$ contains a time-varying confounder if it (a) affects the election outcome, (b) affects future negativity, and (c) is affected by past negativity. In estimating the effect of Democrats going negative, the advertising tone of the Republican would be a time-dependent confounder. Democrats are more likely to go negative if their opponent has gone negative and their opponent's actions are likely related to the outcome. Note that X_t could include past values of Y, in which case the lagged dependent variable would be a time-dependent confounder.

4.1 A model for the potential outcomes

The goal of dynamic causal inference is identical to that of single-shot causal inference: to estimate the means of the potential outcomes under various action sequences. These are population-based quantities of interest: what would happen if *every* Democrat remained positive? In the single-shot approach, this only involved estimating two quantities: $E\big[Y_i(1)\big]$ and $E\big[Y_i(0)\big]$. In dynamic causal inference, there is one potential outcome for each action sequence. A key consequence is that even with a small number of time periods, there will be an overwhelming number of possible action sequences. With two potential outcomes, we can non-parametrically estimate the mean outcome in the treated and control groups by taking sample means. With just ten periods, however, there would be 1,024 possible action sequences, making it unlikely that there will be even one unit following any particular sequence. Thus, the non-parametric approach of single-shot methods will be useless here.

To overcome this curse of dimensionality, we can use a parametric model to relate the action sequences to the potential outcomes. That is, we will suppose that "similar" action sequences should have "similar" potential outcomes. Imposing this structure on the problem reduces the dimensionality of the problem at the expense of possible model misspecification. Robins, Hernán and Brumback (2000) introduced a parsimonious class of semi-parametric models for this problem called marginal structural models (MSM). In this class of models, we assume a parametric form for the mean of the

potential outcome

$$\mathrm{E}[Y(\underline{a})] = g(\underline{a}; \beta), \tag{9}$$

while leaving the rest of the distribution of $Y(\underline{a})$ unspecified.

The function g defines our assumptions about which action sequences should have similar potential outcomes. We may have, for instance,

$$g(\underline{a}; \beta) = \beta_0 + \beta_1 c(\underline{a}), \tag{10}$$

where $c(\underline{a}) = \Sigma_{t=0}^{T} a_t$ is the cumulative action. This model assumes that units with the same number of total periods acted should have similar potential outcomes, with β_1 as the causal effect of an additional period of the action. In the context of negative campaigning, β_1 is the effect of an additional week of negativity. An assumption here is that going negative for the first five weeks of the campaign is the same as going negative for the last five weeks of the campaign. Depending on the application, this might be a more or less plausible assumption and, in general, these types of modeling assumptions will always produce some amount of bias. The greater flexibility we allow for $g(\underline{a}; \beta)$, however, the more variable our estimates become. The substance of the problem and the amount of data on hand will determine what model makes sense for the potential outcomes.

Supposing that (10) was the correct model for the potential outcomes, we want to estimate its causal parameters. One approach would be to estimate

$$\mathrm{E}[Y|\underline{A} = \underline{a}] = \gamma_0 + \gamma_1 c(\underline{a}), \tag{11}$$

which omits any covariates X_t and simply regresses the outcome on the observed action. This approach replaces the potential outcomes $Y(\underline{a})$ with the observed outcomes Y, holding the model fixed. If X_t affects the action and the outcome, however, the associational parameter, γ_1, will not equal the causal parameter, β_1, due to omitted variable bias. That is, differences in the observed outcomes could

be due to difference in the covariate history, not the action sequence. We could instead condition on X_t by estimating

$$\mathrm{E}[Y|\underline{A} = \underline{a}, X_t] = \delta_0 + \delta_1 c(\underline{a}) + \delta_2 X_t, \tag{12}$$

either through a regression that includes X_t or a matching algorithm which matches on X_t. The key parameter, δ_1, will still fail to equal the causal parameter of interest, β_1, when X_t is a time-varying confounder, since X_t is post-treatment for $\underline{A}_{t-1}$. Thus, X_t is in the difficult position of being both an omitted variable *and* a post-treatment variable for the action history. These traditional methods of estimating β_1 fail in the face of time-varying confounders, whether or not we adjust for them, since either approach leads to bias.

One might think that the two traditional estimation procedures would at least provide bounds on the true causal effect, with β_1 falling between γ_1 and δ_1. When the omitted variable bias and the post-treatment bias have the same sign, however, this bounding will fail to hold. This can occur, for instance, when strategic actors attempt to compensate poor performance with beneficial actions. Suppose that there is a strong, positive effect of negative advertising and that trailing campaigns use it to bolster their positions. The omission of polling in a model would lead to an understatement of the negativity effect, since candidates tend to be trailing when they go negative. Positive campaigns would appear stronger than negative campaigns, even though negativity boosts performance. The inclusion of polling in a model would also lead to an understatement of the effect, since it washes away the increase in polls from past negativity. Thus, the true effect of negativity would be higher than either of the traditional methods would predict.

4.2 The assumptions

The quantity of interest in (10) is a feature of the potential outcomes, not observable quantities. As with single-shot actions, we must make certain assumptions to connect the theoretical constructs with functions of the data.

Assumption 1 (Consistency). *For any unit i, action sequence* $\underline{a}$*, and observed action history* $\underline{A}_i$*, if* $\underline{A}_i = \underline{a}$

for some unit, then for the same unit $Y_i(\underline{a}) = Y_i$.

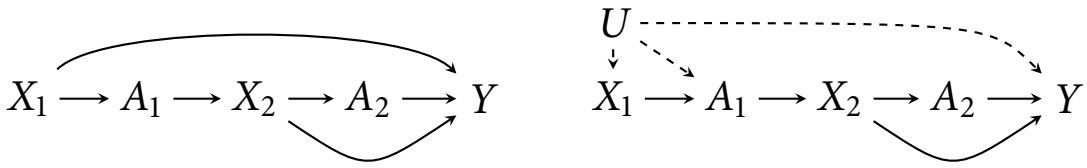
This assumption simply connects the potential outcomes to observed outcomes. Namely, we assume that units who followed an action sequence will observe (a draw from the distribution of) the potential outcomes for that action sequence. For example, the potential outcome under the strategy "always go negative" is observed by those units who do, in fact, always go negative. As the potential outcomes are defined at the level of the unit, we are implicitly accepting SUTVA. With negative advertising, this might be problematic if the analysis included separate units for the Democrat and Republican candidates from the *same election*, since opponent strategy must affect the potential outcome in a race. To avoid these problems, we can take one candidate from each election, the Democrat, as the unit of analysis.

In the sample, however, the candidates that actually went negative always might be different than those who did not. Thus, the sample of units who followed the strategy would be an unrepresentative sample of the potential outcome under that strategy. In order to rid our analysis of the above selection problems, we must be able to identify and measure all possible confounders.

Assumption 2 (Sequential Ignorability)**.** *For any action sequences* $\underline{a}$*, covariate history* $\underline{X}_t$*, and time t, if* $\underline{A}_{t-1} = \underline{a}_{t-1}$*, then* $Y(\underline{a}) \perp\!\!\!\perp A_t | \underline{X}_t, \underline{A}_{t-1} = \underline{a}_{t-1}$.

The assumption of sequential ignorability extends the conditional ignorability assumption to time-varying actions. It states that action decision at time t is independent of the potential outcomes, conditional on the covariate and action histories up to that point. That is, conditional on the past, those who go negative are similar to those who stay positive. Figure 2a shows a causal directed acyclic graph (DAG) in which sequential ignorability holds, while Figure 2b shows a situation where the assumption fails to hold due to an omitted variable U. If decisions are made by a coin flip, then clearly this assumption will hold. If units act based on the covariate history, however, then it will fail to hold unless the analyst can observe all of those covariates. For instance, the assumption would be violated if campaigns made the decision to go negative based on polling data, but the analyst did not have

access to that polling data. The goal for researchers, then, is to collect all the covariates that might influence the decision to go negative in some week. While this is a daunting task in an observational study, it is no harder than satisfying conditional ignorability in the single-shot case and Section 6.1 shows how to relax the assumption in a sensitivity analysis.



(a) Sequential ignorability holds.

(b) Sequential ignorability fails.

Figure 2: Directed Acyclic Graphs representing different assumptions about sequential ignorability, where U is an unobserved variable.

Finally, in order to compare the various action sequences, each must have some positive probability of occurring. It is nonsensical to estimate the effect of a sequence that could never occur.

Assumption 3 (Positivity). *For any sequences* $\underline{a}_t = (\underline{a}_{t-1}, a_t)$ *and* $\underline{x}_t$*, and time t, if* $\Pr\left(\underline{A}_{t-1} = \underline{a}_{t-1}, \underline{X}_t = \underline{x}_t\right) > 0$*, then* $\Pr\left(A_t = a_t | \underline{X}_t = \underline{x}_t, \underline{A}_{t-1} = \underline{a}_{t-1}\right) > 0$.

This assumption outlines the types of strategies we can study. Positivity can break down when some sequences fail to occur in the actual data even though they are theoretically possible. In negative advertising, for instance, candidates with extremely safe seats never go negative, even though nothing is stopping them from doing so. Unfortunately, we will be unable to estimate the effect of going negative for these candidates. These empirical violations of positivity are closely related to the assumption of *common support* often invoked in the matching literature. Section 6 discusses these practical problems with positivity and how to restrict the analysis to the common support.

4.3 The action decision model and inverse-probability of treatment weighting

As shown above, the usual single-shot approaches break down when the actions can vary over time. Fortunately, inverse-probability of treatment weighting (IPTW) can recover unbiased estimates of

causal effects, even in dynamic settings. To see how IPTW works, note that, due to the omitted variables, the distribution of the potential outcomes differs from the distribution of the observed outcomes ($E[Y(\underline{a})] \neq E[Y|\underline{A} = \underline{a}]$). Regression and matching attempt to avoid this problem by finding subsets of the data where those distribution are the same and making comparisons within these subsets. This conditioning removes the omitted variable bias, but it can induce post-treatment bias. Methods that rely on weighting, such as IPTW, avoid these by never explicitly conditioning on the confounders.

Robins, Hernán and Brumback (2000) show that under the above assumptions, a reweighted version of the observed outcomes will have the same distribution as the potential outcomes. In the campaigns context, the reweighted outcomes for always-positive campaigns will look like the outcomes if we forced all Democrats to remain positive. The weights are defined as

$$W_i = \frac{1}{\prod_{t=1}^{T} \Pr(A_{it}|\underline{A}_{it-1}, \underline{X}_{it})}. \tag{13}$$

In words, the denominator of W_i is the probability of observing the action sequence that unit i actually took. To build up this quantity, we take the product of observed action probabilities in each period, conditional on the past.

A simple example helps to explain the construction of the weights. Suppose that there were only two weeks in a campaign, with a poll update in between the weeks. A candidate decides to go negative or stay positive in the first week, sees the outcome of the poll, decides to go negative in the second week, and then observes the election results. A candidate who stays positive in week one, trails in the polls, and then goes negative in week two would have the following weight:

$$W_i = \frac{1}{\Pr(\text{pos}_1) \Pr(\text{neg}_2|\text{trail}, \text{pos}_1)}. \tag{14}$$

The first term in the denominator is simply the probability of being positive in the first week. The second term is the probability she would have gone negative in the second week, conditional on

trailing and having been positive in the first week. The resulting denominator is the probability of observing the campaign $(\text{pos}_1, \text{neg}_2)$, conditional on the time-varying covariate, polls.

Of course, without randomization, the probability of going negative will be unknown, leaving (13) to be estimated. To do so, we must model the decision to go negative in each week, conditional on the past. Since the decision is dichotomous, a common approach is to estimate the probability of going negative with a logit model:

$$\Pr(A_{it} = 1 | \underline{A}_{it-1}, \underline{X}_{it-1}; \alpha) = \left(1 + \exp\left\{-h\left(\underline{A}_{it-1}, \underline{X}_{it}; \alpha\right)\right\}\right)^{-1}, \tag{15}$$

where h is a linear, additive function of the action history, covariate history, and parameters α. For instance, we might have

$$h\left(\underline{A}_{it-1}, \underline{X}_{it}; \alpha\right) = \alpha_0 + \alpha_1 A_{it-1} + \alpha_2 X_{it} + \alpha_3 t, \tag{16}$$

which models the action decision as a function of negativity in the last week (A_{t-1}), the most recent poll results (X_{it}), and the week of the campaign (t).

An estimate of the weights requires an estimate of the parameter vector α from this model. We can obtain these estimates, $\hat{\alpha}$, from a pooled logistic regression, treating each campaign-week as a separate unit. These estimates form the basis for the estimated weights,

$$\widehat{W_i} = \frac{1}{\prod_{t=1}^{T} \Pr\left(A_{it} | \underline{A}_{it-1}, \underline{X}_{it}; \hat{\alpha}\right)}, \tag{17}$$

where the denominator is now the predicted action probability (or fitted value) for unit i at each time period. [7] Note that it is not necessary to estimate the same model for all units. For example, incumbents and non-incumbents might require different models because their approaches to the negativity decision are so distinct.

[7] These values are easily found using a combination of the `glm` and `predict` functions in R (R Development Core Team, 2011).

Each observation i is then weighted by $\widehat{W_i}$ in a weighted generalized linear model for the outcome, with form $g(\underline{a};\beta)$ from (10).[8] Appendix A shows this estimation procedure is consistent for the causal parameters, β, under sequential ignorability, consistency, positivity, and the correct model for the weights. The most straightforward way to estimate standard errors and confidence intervals is to bootstrap the entire estimation procedure, including the weights (Robins, Hernán and Brumback, 2000). For negative campaigning, this means resampling the set of campaigns (not the set of campaign-weeks), re-estimating the weights, and running the weighted outcome model on the resampled data.

4.4 Stabilized Weights

If campaigns have vastly different likelihoods of going negative, then the estimated weights from (17) can have extreme variability, which results in low efficiency. We can use a slightly different version of the weights, called the *stabilized weights*, to decrease this variability and increase efficiency. The stabilized weights take advantage of an interesting fact: the numerator of the weights does not change the consistency of the estimation procedure.[9] While it was natural to use the value 1 as the numerator, we can replace it with other functions of the action history that increase efficiency. The usual choice used in the literature (Robins, Hernán and Brumback, 2000) is

$$SW_i = \frac{\prod_{t=1}^{T} \Pr\left(A_{it}|\underline{A}_{it-1};\delta\right)}{\prod_{t=1}^{T} \Pr\left(A_{it}|\underline{A}_{it-1},\underline{X}_{it};\alpha\right)}, \tag{18}$$

where the numerator is a model for the marginal probability of action, conditional on past action. When actions are randomized, these stabilized weights will be equal to one since the action probability would be unaffected by the covariates in the denominator.

Of course, the numerator of SW_i is unknown, leaving us with the task of estimating δ. All this

[8] The `survey` package in R can implement this weighting for a large class of outcome models (Lumley, 2004).

[9] As shown in Appendix A, the numerator only alters the marginal distribution of the action $\underline{A}$, which does not affect the marginal distribution of the potential outcomes, $Y(\underline{A})$. This is because the marginal distribution of $\underline{A}$ does not affect the distribution of Y conditional on $\underline{A}$, which is the crucial ingredient for the distribution of the potential outcomes.

requires is an additional logit model for the numerator to estimate the probability of going negative without conditioning on the time-varying covariates. If the outcome model will include interactions with baseline covariates, then both the numerator and the denominator should include those variables. To construct these weights, one simply needs to obtain predicted probabilities from each model for every unit-period. Then, for each unit, take the product of those probabilities across time periods and divide to obtain the estimates $\widehat{SW_i}$.

§5 An empirical example: going negative

Pundits and theorists often bemoan the growth in negative campaign advertising in recent decades. Less often do they discuss its effectiveness. An implicit assumption in the air of political discourse is "Of course it works, politicians do it." The prospect of dirtying the waters with such cheap and tawdry tactics is bad enough, being useless would only add insult to injury. A contingent of political scientists have investigated just how useful negativity is for candidates, to varying levels of success.[10]

A common problem for these investigations is that campaign tone is a dynamic process, changing from week to week. Furthermore, there are strong time-varying confounders. For instance, poll numbers affect the decision to go negative, but going negative also affects poll numbers. Thus, polling is both pre- and post-action: a classic time-varying confounder. As shown above, ignoring the polls and conditioning on the polls will both result in biased estimates. We can estimate the effect of time-varying actions, though, using marginal structural models and inverse probability of treatment weighting.

The goal of this application is to estimate the effect of going negative for Democratic candidates in state-wide elections. I use data on campaigns for Senate and Gubernatorial seats in the cycles of 2000, 2002, 2004, and 2006. For each campaign, I code the advertising tone using data from the University of Wisconsin Advertising Project (Goldstein and Rivlin, 2007). To ensure consistency across years, I use a simple measure of negative or contrast ads: does the ad mention the opposing

[10] See Lau, Sigelman and Rovner (2007) for a recent, comprehensive review.

candidate?[11] I use this coding to construct a measure of whether a candidate has "gone negative" in a given week of the campaign based on what percentage of ads are negative.[12] The WiscAds data also provide a proxy for weekly campaign spending: the total number of ads aired in a week. In addition to advertising data, I also collected weekly polling data from various sources,[13] along with baseline covariates, such as predicted competitiveness of the race (as measured by the *Congressional Quarterly* score), incumbency status, number of ads run by each candidate in their primaries, the length in weeks of the campaign, measures of challenger quality and incumbent weakness, and the number of Congressional districts in the state. Much of this data comes from Lau and Pomper (2002) with additional data collection. In this example, baseline is the day after the final primary.

5.1 A model for going negative

In order to estimate the causal parameters from an MSM, we must construct the weights from Section 4.4. The covariates, X_{it}, include covariates that influence the decision to go negative: polling, past negativity of both Democrats and Republicans, and the amount of advertising by both candidates. I constructed the weights from four separate pooled-logistic models: a separate numerator and denominator model for incumbents and non-incumbents.[14]

These models largely fit with the intuition and theory of campaigns, with high-advertising and already-negative races being more likely to be negative. Figure 3 shows that there is a strong relationship between polling and the decision to go negative: non-incumbent Democrats in safe seats rarely go negative, but those who are trailing often do. To construct the weights, I combine predicted probabilities from these models according to (18).[15] Due to empirical violations of positivity, I restricted

[11] The WiscAds project failed to collect data in 2006, so I acquired and computer-coded the data directly from CMAG, the consultant group which provides the data to WiscAds.

[12] For the analysis below, I used a cutoff of 10%. The results appear unaffected by this choice, as weeks tend to be dominated by only a few ads. If there is one ad that is negative, it pushes the percent negativity quite high.

[13] Polling data comes from the *Hotline* daily political briefing for 2000 and 2002 and from `http://www.pollster.com` for 2004 and 2006.

[14] In order to stabilize the weights further, I include all baseline covariates in both the numerator and denominator models. This means that IPTW will only balance the time-varying covariates, leaving any remaining baseline imbalance. Since this imbalance is time-constant, we can remove it through traditional modeling approaches and, thus, I include those covariates in the outcome model below.

[15] The weight models are pooled logistic generalized additive models (GAMs), which is what allows for the flexible

the analysis to common support (see Section 6 for further details).

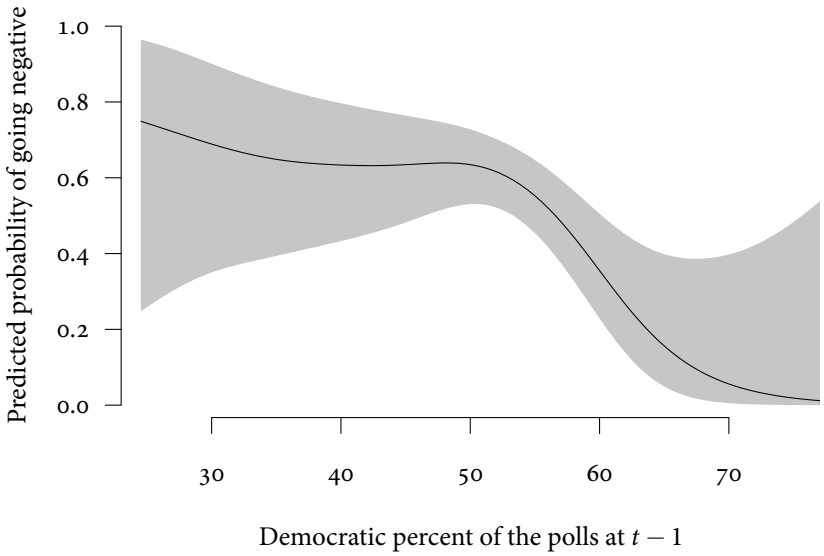


Figure 3: The marginal relationship between lagged polling numbers and going negative for Democratic non-incumbent candidates. All other variables from the model held at their mean, or median, depending on the type of variable. The shaded region is a 95% confidence bands. Intuitively, trailing Democrats are more likely to go negative than leading Democrats.

5.2 The time-varying effects of negativity

The effect of negative advertising is unlikely to be constant across time. Ads closer to election day should have a stronger impact than those earlier in the campaign and marginal structural models allow us to estimate these time-varying effects. I break up the effect into a an early campaign effect (the primary through September) and a late campaign effect (October and November). Vote shares are continuous, so a linear MSM is appropriate for the potential outcomes:

$$E[Y_i(\underline{a})] = \beta_0 + \beta_1 \left(\sum_{T-5}^{T} A_{it} \right) + \beta_2 Z_i + \beta_3 Z_i \left(\sum_{T-5}^{T} A_{it} \right) + \beta_4 \underline{A}_{iT-6} + \beta_5 Z_i \underline{A}_{iT-6} + \beta_6 X_i, \qquad (19)$$

where Z_i is an indicator for being a Democrat incumbent, X_i is a vector of baseline covariates, and T is the week of the election. The summation terms calculate how many of the last five weeks of

modeling of the polling. I used the `mgcv` package to fit this model (Wood, 2011).

Estimator	Democratic Incumbent	Democratic Non-incumbent
Naïve	-0.96 (-1.68, -0.33)	0.49 (-0.18, 1.20)
Control	-0.54 (-1.28, 0.11)	0.63 (-0.02, 1.26)
IPTW	-0.70 (-1.47, 0.01)	0.75 (0.21, 1.27)

Table 1: Estimated effects of an additional week of negative advertising in the last five weeks of the campaign on the Democratic percent of the two-party vote. Bootstrapped 95% confidence intervals are in parentheses, with those crossing zero set in gray. Inverse probability weighting estimates a strong, positive effect for non-incumbents and a strong, negative effect for incumbents. Note that the competing models fail to bound the IPTW-estimated effect.

the campaign the Democrat went negative. This covers October and early November, which is the home-stretch of the campaign. The model separately estimates the direct effect of earlier negativity and allows for incumbent status to modify both early and late effects. Following the IPTW approach, I weight each campaign using weights constructed from the above "going negative" model.

It is instructive to compare estimates from this model with two competing approaches. First, the *naïve estimator* simply ignores all time-varying covariates and fits (19) to the observed data without weights. Second, the *control estimator* attempts to control for the covariates by including them as additional regressors in (19).[16] These represent the two single-shot methods recommended by the statistical literature: the naïve estimator to guard against post-treatment bias and the control estimator to guard against omitted variable bias.

Table 1 shows the estimated effects of late campaign negativity from all three models broken out by incumbent status.[17] The MSM finds that Democratic incumbents are hurt by going negative, while non-incumbents are helped. Non-incumbents see a 0.72 percentage point increase in the Democratic percent of the two-party vote for every additional week of negative advertising in the last five weeks. Incumbents, on the other hand, drop 0.70 percentage points for the same change. As Figure 4 shows, there is no evidence of a direct effect of earlier negativity on the final vote in either group. Note that these results control for polls taken at the beginning of the campaign. It is surprising, then, to see effects that are even this large since these baseline polls are highly predictive of the outcome in Senate

[16] Polls are included as the mean Democrat poll percentage, total number of ads as the average ads per week, and Republican negativity as the overall duration of Republican negativity.

[17] Estimates here are produced using the `svyglm` function in the `survey` package, version 3.22-4 (Lumley, 2010). Standard errors and confidence intervals come from bootstrapping this model.

and Gubernatorial elections.

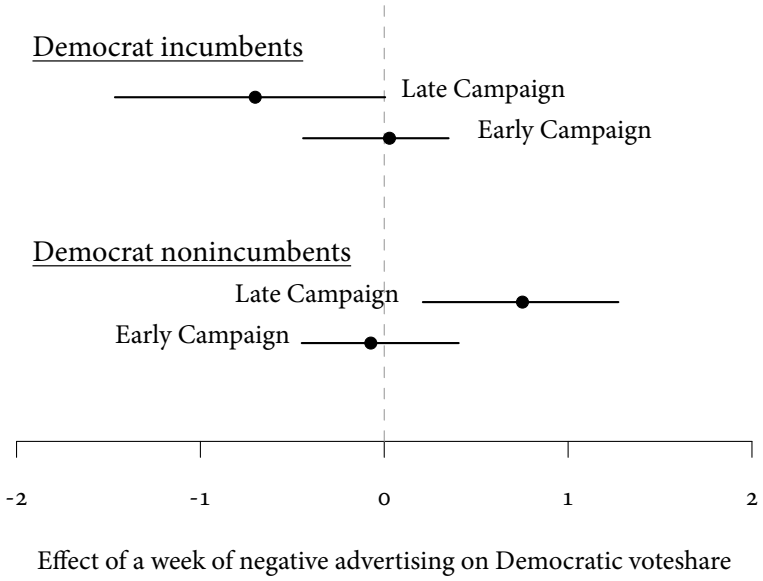


Figure 4: Inverse-probability of treatment weighting estimates of the time-varying effects of negative campaigning with bootstrapped 95% confidence intervals. Negative ads are more potent later in the campaign (October and November) than earlier in the campaign, but the direction of the effect is negative for incumbents and positive for non-incumbents.

Interestingly, the MSM-estimated effect is well outside of the bounds set by the naïve and control estimators. For non-incumbents, the IPTW estimate is over 18% larger in magnitude than either of the other methods.[18] Thus, "trying it both ways" would be an unsuccessful strategy in this case. This could be due to the strategic compensation story from Section 4.1. If negativity does work and this is common knowledge among campaign managers, then candidates who are behind might attempt to use negativity to boost their chances of winning. In this case, both the omitted variable bias and the post-treatment bias would have the same sign, leading to a bounding failure. Indeed, Figure 3 suggests those candidates who are trailing are much more likely to go negative than leading candidates.

[18]For incumbents, the two methods do bound the effect, but there are additional reasons, below, to be skeptical of the results for this group.

§6 Assessing model assumptions

With single-shot causal inference methods such as matching, balance checks are crucial diagnostics (Ho et al., 2006). These checks ensure that the treated and control groups are similar on their background covariates. Usually this takes the form of simple comparisons of covariate means in the treated and control group, though more sophisticated techniques exist. Unfortunately, this simple approach is ill-suited to the dynamic setting since it is unclear what groups to compare. At a given week of the campaign, negative and positive campaigns might differ on a time-varying confounder, but these differences might be due to past negativity.

Under the above assumptions of the IPTW estimator, the decision to go negative is unconfounded in the weighted data, conditional on past negativity. We should expect, then, that the observed actions will be independent of time-varying covariates once we weight by SW_i. This independence is, however, conditional on a unit's action history. For instance, suppose we had two campaigns that had remained positive until week t. Then the decision to go negative in week $t+1$ for these two campaigns should not depend on time-varying covariates, such as polling, in the weighted data.

We can assess balance in the weighted data, then, by checking for associations between the action decision and the time-varying covariates that affect that decision, conditional on the action history. If, after reweighting the data and conditioning on past negativity, the decision to go negative is still predictive of past polling, then there is likely residual confounding of the relationship between the outcome and negativity. Thus, we regress polling at time t on negativity at time t (which occurs after polling) in the weighted data while controlling for tone history and any baseline covariates. That is, we estimate a weighted generalized linear model for $E[X_t|A_t, \underline{A}_{t-1}, X_1]$ and take the coefficient on A_t as the history-adjusted imbalance. The history of negativity and the baseline covariates, X_1, should have the same functional form as in the weighting model. Thus, if the weighting model incorporates the actions of the last five periods, then this same structure should be in the imbalance models. This approach checks that, after weighting, the decision to go negative is unrelated to any other aspects of

the campaign except past negativity. One useful approach to choosing a model for the weights is to minimize the history-adjusted imbalance.

Figure 5 shows how the weights reduce this *history-adjusted imbalance* in the campaign advertising example. It shows the change in the standardized history-adjusted imbalance from the unweighted to the weighted data.[19] In the unweighted data, for instance, Democrats were much more likely to go negative after an attack by Republicans (R Neg$_{t-1}$). Once we apply the weights, however, the differences move much closer to zero because IPTW gives relatively more weight to races that went negative without Republican negativity in the last week.

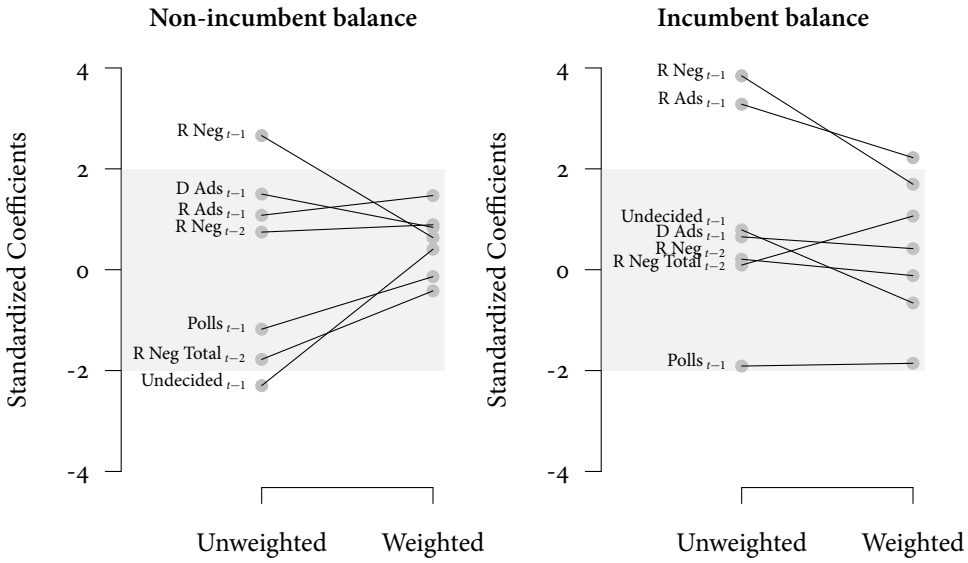


Figure 5: The change in history-adjusted balance between the weighted and unweighted data as measured by standardized differences between those campaign-weeks that went negative versus those that remained positive, conditional on baseline covariates. Note that the differences are, all told, closer to zero in the weighted model. "R Neg" is whether the Republican went negative, "Ads" are the number of ads run by the candidates, "R Neg Total" is the total number of Republican negative weeks in the campaign, and "Polls" is the averaged polling numbers for Democrats.

One stark observation from the diagnostic plot is that confounding exists for incumbents even after weighting for the time-varying confounders. This makes sense for two reasons. First, data quality issues plague incumbents since their safer races attract less polling. Second, incumbents have

[19]These differences come from a unweighted and weighted pooled regression of the time-varying covariate at t on (a) the baseline covariates, (b) Democratic negativity before week t, and (c) Democratic negativity in week t. The coefficient on (c), divided by its standard error, is the standardized difference.

stronger positivity problems with extremely safe seats rarely going negative. Furthermore, incumbent campaign-weeks with high total ad volumes almost always feature negativity. These issues prevent the weights from fully eliminating the confounding in the data and should give us pause when interpreting the estimates for incumbents.

Of course, the history-adjusted balance depends on the correct specification of the weighting model, in addition to the assumptions of sequential ignorability and positivity. Cole and Hernán (2008) propose a series of models checks based on the distribution of the weights, SW_i. They note that if the mean of the stabilized weights strays too far from 1, then there is a strong indication of an empirical positivity violation. To see why this is the case, it is instructive to look at an example. Take Hernandez, a fictional candidate running a positive campaign in week $t-1$, when her opponent, candidate Smith releases a slew of negative ads. Most campaigns that are positive until $t-1$ will remain positive in week t. Thus, the numerator model of SW_i would predict a high probability of staying positive in week t. Among those campaigns positive up to $t-1$, though, an attack from their opponent drives most to respond with negativity in week t. Thus, the denominator model of SW_i, which includes time-varying covariates, would predict a very low probability of staying negative. If Hernandez goes negative in week t, then she will have a weight less than 1, since her observed action is very *unlikely* for her action history overall, and yet very *likely* for her combined action and covariate history. If she remains positive, then her weight will be very high, since her action is surprising given her covariate history.

The confounding of time-varying covariates is what pushes weights away from 1. A mean weight far below 1 indicates that there are relatively few surprise actions—those that are unlikely given the covariate history. This lack of "surprises" indicates that the probability of going negative is close to 0 and 1 in some parts of the covariate space, which is a violation of positivity. In the matching literature, this is called a lack of common support. Cole and Hernán (2008) also point to the extreme values of the weights as potential signs of imbalance and positivity violations, since these values could be due to units isolated in the sample space, with very few good comparison observations. A good first

check for these issues in the weight model is to check the distribution of stabilized weights period to ensure that (a) the means at each point in time are close to 1 and (b) the minimum and maximum values are reasonably close to 1. A boxplot or stem-and-leaf plot of the weights for each time period is a reasonable way to visually inspect these distributions.

In the campaign advertising application, initial weighing models had maximum weights of 15 to 20, a clear indication that there were portions of the covariate space with essentially no common support. For instance, campaigns for extremely safe seats in either direction had almost no negative ads. Thus, I remove any campaign with baseline polling greater than 70% for either side. The restriction to common support alters the quantity of interest to the effect on the sub-population defined by the remaining units. This is fairly innocuous as the restrictions affect less than 10% of the campaign-weeks. Note that weeks in which a candidate runs *no* ads is week where a candidate *cannot* go negative. These weeks receive a stabilized weight of one, meaning they do not contribute to the weight of their overall campaign. A more thorough analysis would treat the number of ads and the tone of those ads a joint treatment. The tone of ad-less weeks, then, would be censored and one could use inverse probability of censoring weights as part of the stabilized weights (see Robins, Hernán and Brumback, 2000, Section 10). The final distributions of the weights by week are in Figure 6. Their means are all very close to one and the upper bounds are fairly low, indicating well-behaved weights.

6.1 Sensitivity Analyses

Causal estimates from an MSM have excellent properties when the assumptions of consistency, positivity, and sequential ignorability hold. Of these, sequential ignorability is the trickiest, as it requires that, conditional on the covariate and action histories, the action decision is unrelated to the potential outcome. Any residual differences in potential outcomes between treated and control groups we call *unmeasured confounding* or omitted variable bias. Unless we conduct an experiment and randomize the action, this assumption must be justified through substantive knowledge. Since it is impossible to test this assumption, it is vital to include as much information as possible and to conduct a sensitivity analysis of any estimated results.

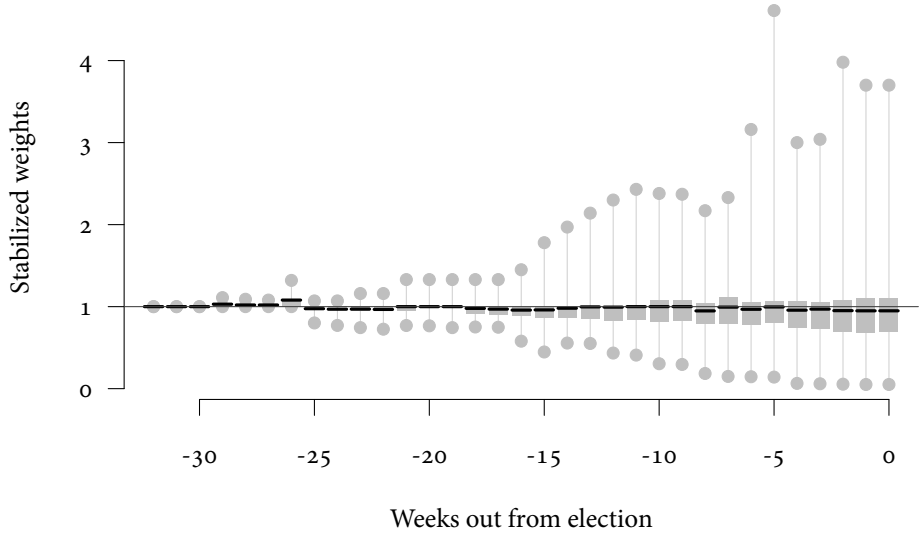


Figure 6: Stabilized weights over the course of the campaigns. The black lines are the weekly means, the gray rectangles are the weekly inter-quartile ranges, and the thin gray lines denote the range of the weights. Note that campaigns begin at various times so that there are very few campaigns at 30 weeks out, but very many at 5 weeks out. These weights appear well-behaved as their means are close to 1.

Robins (1999) proposes a method to investigate the sensitivity of estimates to the presence of unmeasured confounding. Robins quantifies the amount of confounding as

$$q(\underline{x}_t, \underline{a}_{t-1}, a_t^*) = E[Y(\underline{a})|\underline{x}_t, \underline{a}_{t-1}, a_t] - E[Y(\underline{a})|\underline{x}_t, \underline{a}_{t-1}, a_t^*]. \tag{20}$$

This function measures the difference in potential outcomes for some group that shares an action and covariate history. For two campaigns that are observational equivalent up to week t, q measures the structural advantages of those campaigns that go negative in week t compared to those that remain positive in week t. Almost all causal inference models, including MSMs, assume that there are no omitted variables so that the function q is always zero.

Instead of assuming no omitted variables, the sensitivity analysis method of Robins assumes there is some bias, the amount controlled by a parameter α. For instance, the confounding may take the form

$$q(\underline{x}_t, \underline{a}_{t-1}, a_t^*; \alpha) = \alpha[a_t - a_t^*]. \tag{21}$$

This is a simple and symmetric form of omitted variable bias. When α equals zero, then there is no difference between those campaigns that go negative in week t and those that stay positive, given campaign histories. If α is positive, then negative campaigns are instrinsically better than those that remain positive. That is, when α is positive, then $Y(\underline{a})$ is higher for $a_t = 1$ (negative campaign-weeks) than $a_t = 0$ (positive campaigns-weeks). If α is negative, then those candidates who are going negative are worse off. Note that these selection biases are all conditional on the observed covariate history.

The above IPTW estimation procedure assumes that $\alpha = 0$, yet Robins (1999) shows that we can estimate the parameters under any assumption about α. Thus, by setting α to various levels, we can estimate the causal effect under different assumptions about the degree of omitted variable bias. To do so, we have to replace the outcome with a *bias-adjusted* outcome,

$$Y_\alpha \equiv Y - \sum_{k \in (0,1)} \sum_{t=0}^{T} \underbrace{q(\underline{X}_t, \underline{A}_{t-1}, k; \alpha)}_{\text{bias at this history}} \cdot \underbrace{\Pr(A_t = k | \underline{A}_{t-1}, \underline{X}_t)}_{\text{probability of reaching this history}} , \tag{22}$$

where the first term is simply the observed outcome, Y, and the second term is the overall omitted variable bias, built up from the bias in each time period. We can then re-estimate the parameters of the marginal structural model with outcome Y_α instead of Y to get bias-adjusted estimates and bias-adjusted confidence intervals. Note that when $\alpha = 0$, the bias function is zero, so that $Y_0 = Y$ and the usual estimation aligns with the assumption of sequential ignorability. Of course, the probability term is unknown and must be estimated. Fortunately, we have already estimated this function as part of the estimation of the weights, SW_i.

Figure 7 shows how the estimated effect of late-campaign negativity varies across different assumptions about the omitted variable bias, encoded in the parameter α, which runs along the x-axis. The magnitude of α describes how much stronger or weaker the negative campaigns are, on average, in terms of their potential outcomes. This figure also charts the change in the confidence intervals under the various assumptions about bias, with those that cross zero shaded lighter.

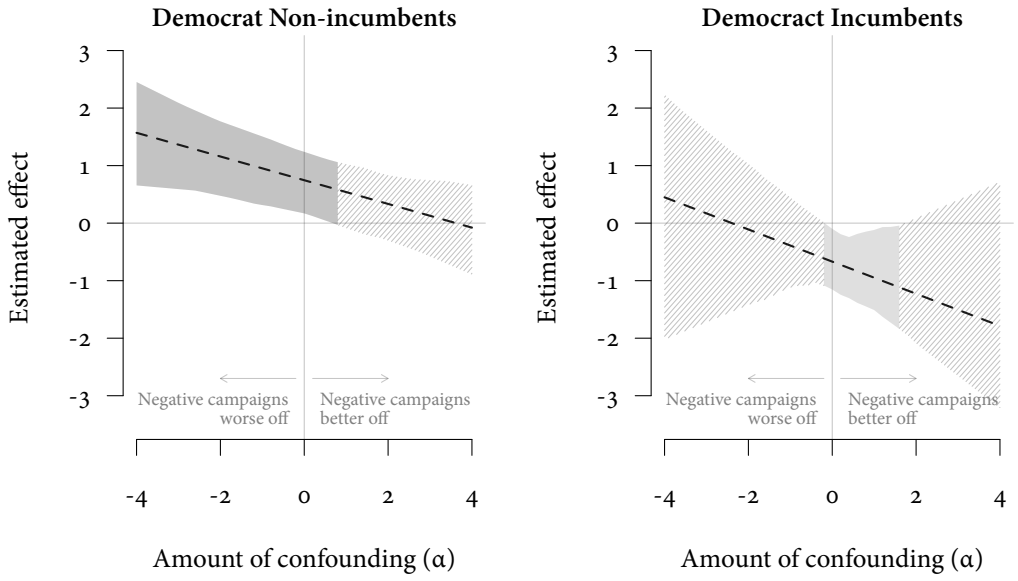


Figure 7: Sensitivity of the results to deviations from the sequential ignorability assumption. The parameter α *indexes assumptions about confounding, where negative values indicate that the observed negative campaigns are inherently weaker than predicted by the observed variables. Positive values assume that those negative campaigns are stronger than predicted.*

When negative campaigns are even 0.5 points stronger than positive campaigns on average, our 95% confidence intervals would overlap with zero for non-incumbents. This might occur if campaigns were attacking their opponents for (unmeasured) behavior such as a scandal or unpopular vote. Note that these imbalances would have to occur within levels of the covariate history and thus exist after conditioning on polls. If the negative campaigns are instead weaker, perhaps because campaigns go negative when they are in trouble, then results only grow stronger for non-incumbents. The results for incumbents highlight the potential violations of ignorability for that group. The results are fairly sensitive to the degree of confounding, both in the point estimates and the confidence intervals.

It is clear that there is some sensitivity to the sequential ignorability assumption and we could attempt to gather more data (quantitative and qualitative) to justify which direction this confounding is likely to lean. Notably, we could measure a larger set of dynamic campaign features such as scandals and endorsements. It would also help to draw on electoral cycles, such as 2008, when incumbents faced strong challengers to overcome the overlap and ignorability issues.

§7 Conclusion

Political actions do not happen all at once. There are sequences of events that unfold over time. As we have seen, this poses strong problems for extant single-shot causal inference methods. This paper brings to bear a framework that explicitly models the dynamic sequences and builds methods to test their effects. The original application of MSMs was to epidemiological data. Robins (1997) develops a set of methods called structural nested models with an application to HIV treatment studies. In that context, the units are patients and doctors change the treatment over time if the patient status worsens. The analogy to politics is suggestive: campaign managers and candidates as doctors, working to save their patient, the election. Of course, candidates face human opposed to viral opponents, yet this changes only the types of variables needed to satisfy sequential ignorability.

The the structural nested model of Robins (1997) provide an alternative approach to dynamic causal inference. These techniques center on modeling the effect of going negative at every possible history which allows effects that interact with time-varying covariates. The estimation methods resemble backwards induction in game theory. Unfortunately, these structural nested models require models for entire set of time-varying covariates and complicated computation to estimate while researchers can easily use off-the-shelf software to implement an MSM.

The focus of this paper has been the effect of action sequences, yet in many political science situations, actors follow dynamic strategies—updating their actions based on changing conditions. It is likely that the optimal action is actually a strategy, since being able to respond to the current state of affairs is more effective than following a predefined sequence of actions. Hernán et al. (2006) demonstrates that marginal structural models and inverse-probability weighting can estimate the effectiveness of strategies with a simple form such as "go negative when polls drop below x%." In addition, structural nested models can estimate the effect of arbitrary strategies. As might be expected, precisely estimating these effects requires larger sample sizes than the effects of simple action sequences.

A crucial path for future research is model development. In this paper, I used a fairly simple model

to estimate different effects for early and late in the campaign. This is a crude division of the data and more fine-grained modeling might help to smooth effects over time. Indeed, we would expect that the effect of negativity in week 5 should be quite similar to effect of negativity in week 6. Better MSMs should be able to handle this type of structure.

Dynamic causal inference is a problem for more than just campaigns. Each subfield of political science analyzes actions that occur over time and have multiple decision points: foreign aid, interest rates, budget allocations, state policies, and even democracy. Indeed, many of the assumptions in this paper (or variations thereof) are implicit in time-series cross-sectional TSCS models, where the counterfactual framework is rarely discussed in explicit terms. Thus, there is a great opportunity for future work that identifies areas with dynamic causal inference problems and attempts to clarify or improve existing results.

References

Cole, Stephen R and Constantine E Frangakis. 2009. "The consistency statement in causal inference: a definition or an assumption?" *Epidemiology* 20(1):3–5. 4

Cole, Stephen R and Miguel A Hernán. 2008. "Constructing inverse probability weights for marginal structural models." *American Journal of Epidemiolology* 168(6):656–64. 25

Dawid, A. Phillip. 1979. "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society. Series B (Methodological)* 41(1):1–31.
URL: *http://www.jstor.org/stable/2984718* 5, 34

Goldstein, Kenneth and Joel Rivlin. 2007. "Congressional and gubernatorial advertising, 2003-2004.". Combined File [dataset]. Final release. 20

Hernán, Miguel A, Emilie Lanoy, Dominique Costagliola and James M Robins. 2006. "Comparison of dynamic treatment regimes via inverse probability weighting." *Basic & Clinical Pharmacology Toxicology* 98(3):237–42. 29

Ho, Daniel E., Kosuke Imai, Gary King and Elizabeth A. Stuart. 2006. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15(3):199. 6, 23

Holland, P. W. 1986. "Statistics and causal inference." *Journal of the American Statistical Association* 81(396):945–960. 4

Lau, Richard R. and Gerald M. Pomper. 2002. *Negative Campaigning: An Analysis of U.S. Senate Elections* . Campaigning American Style Lanham, MD: Rowman & Littlefield Publishers, Inc. 20

Lau, Richard R., Lee Sigelman and Ivy Brown Rovner. 2007. "The effects of negative political campaigns: a meta-analytic reassessment." *The Journal of Politics* 69(04):1176–1209. 19

Lumley, Thomas. 2004. "Analysis of Complex Survey Samples." *Journal of Statistical Software* 9(1):1–19. R package verson 2.2. 18

Lumley, Thomas. 2010. "survey: analysis of complex survey samples.". R package version 3.23-2. 22

Pearl, Judea. 2010. "On the consistency rule in causal inference: axiom, definition, assumption, or theorem?" *Epidemiology* 21(6):872–5. 4

R Development Core Team. 2011. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: . ISBN 3-900051-07-0.
URL: *http://www.R-project.org* 18

Robins, James M. 1986. "A new approach to causal inference in mortality studies with sustained exposure periods-Application to control of the healthy worker survivor effect." *Mathematical Modelling* 7(9-12):1393–1512. 10

Robins, James M. 1997. Causal Inference from Complex Longitudinal Data. In *Latent Variable Modeling and Applications to Causality*, ed. M. Berkane. Vol. 120 of *Lecture Notes in Statistics* New York: Springer-Verlag pp. 69–117. 10, 29, 35

Robins, James M. 1999. "Association, Causation, and Marginal Structural Models." *Synthese* 121(1/2):151–179.
URL: *http://www.jstor.org/stable/20118224* 26, 27

Robins, James M. 2000. Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, ed. M. Elizabeth Halloran and Donald Berry. Vol. 116 of *The IMA Volumes in Mathematics and its Applications* New York: Springer-Verlag pp. 95–134. 9, 10, 37

Robins, James M., Miguel Ángel Hernán and Babette Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11(5):550–560.
URL: *http://www.jstor.org/stable/3703997* 2, 12, 16, 18, 19, 26

Rubin, Donald B. 1978. "Bayesian Inference for Causal Effects: The Role of Randomization." *Annals of Statistics* 6(1):34–58. 4, 10

VanderWeele, Tyler J. 2009. "Concerning the consistency assumption in causal inference." *Epidemiology* 20(6):880–3. 4

Wood, Simon N. 2011. "Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models." *Journal of the Royal Statistical Society (B)* 73(1):3–36. 21

§A Supporting Information (SI): Proofs

In this appendix, I establish a few of the results discussed in body of the text. I assume throughout that consistency, sequential ignorability, and positivity hold. First, it is useful to establish that the action decision is unconfounded in the reweighted data.

Lemma 1. *If the observed and counterfactual data meet assumptions* (1), (2), *and* (3), *then the distribution of the observed data,* $O = (Y(\underline{A}), \underline{X}, \underline{A})$, *weighted by SW is unconfounded. That is, in the reweighted*

data,

$$Y(\underline{a}) \perp\!\!\!\perp_{SW} \underline{A} \tag{23}$$

or equivalently,

$$f^{SW}\big(Y(\underline{a}) \mid \underline{A} = \underline{a}\big) = f^{SW}\big(Y(\underline{a})\big), \tag{24}$$

where f^{SW} is the density of the reweighted data.

Proof. The reweighted data has the following density:

$$f^{SW}(Y, \underline{X}, \underline{A}) = \frac{SW(\underline{X}, \underline{A}) f(Y, \underline{X}, \underline{A})}{\int SW(\underline{X}, \underline{A}) f(Y, \underline{X}, \underline{A}) d\mu(Y, \underline{X}, \underline{A})} \tag{25}$$

where $f(\cdot)$ is the density of the unweighted data and I have written $SW = SW(\underline{X}, \underline{A})$ to emphasize its dependence on the observed data. Due to positivity, we know that this density is well defined. The numerator of (25) can be written

$$\frac{\prod_{t=1}^{T} f(A_t|\underline{A}_{t-1})}{\prod_{t=1}^{T} f(A_t|\underline{A}_{t-1}, \underline{X}_t)} f(Y|\underline{X}, \underline{A}) \prod_{t=1}^{T} f(X_t|\underline{A}_{t-1}, \underline{X}_{t-1}) f(A_t|\underline{A}_{t-1}\underline{X}_t),$$

which reduces to

$$f(Y|\underline{X}, \underline{A}) \prod_{t=1}^{T} f(X_t|\underline{A}_{t-1}, \underline{X}_{t-1}) f(A_t|\underline{A}_{t-1}). \tag{26}$$

The denominator of (25) is 1 since it is simply the integral of (26), which is a proper density. In fact, (26) is simply the distribution of the observed data distribution with the conditional action history distribution, $f(\underline{A}|\underline{X})$ replaced with the marginal action history distribution $f(\underline{A})$. Since both lead to valid joint distributions on the observed data, we know that the denominator of (25) will integrate to 1. Thus, we can rewrite the distribution of the reweighted data as,

$$f^{SW}(Y, \underline{X}, \underline{A}) = f(Y|\underline{X}, \underline{A}) \prod_{t=1}^{T} f(X_t|\underline{A}_{t-1}, \underline{X}_{t-1}) f(A_t|\underline{A}_{t-1}). \tag{27}$$

Note that in this distribution, $\underline{X}_t \perp\!\!\!\perp_{SW} A_t | \underline{A}_{t-1}$, since actions are unaffected by the history of the covariates. Combining this fact with sequential ignorability and the contraction property of conditional independence (see Dawid, 1979, Lemma 4.3), we have

$$(Y(\underline{a}), \underline{X}_t) \perp\!\!\!\perp_{SW} A_t \mid \underline{A}_{t-1}, \tag{28}$$

which implies

$$Y(\underline{a}) \perp\!\!\!\perp_{SW} A_t \mid \underline{A}_{t-1}, \tag{29}$$

by the decomposition property of conditional independence. Note that since t was arbitrary, this must hold for all t, which implies the result $Y(\underline{a}) \perp\!\!\!\perp_{SW} \underline{A}$. □

Note that this lemma applies to any weight $SW = g(\underline{A})/f(\underline{A}|\underline{X})$, where $g(\underline{A})$ is an arbitrary density of $\underline{A}$, not necessarily the true marginal density. This would only change the integration in the denominator of (A) to some other constant than 1. This lemma shows that when we weight each unit by the inverse of the probability of its observed action history, we obtain a pseudo-sample in which the treatment is balanced. Thus, in the reweighted population, we can obtain the mean of the counterfactual outcome from the conditional expectation of the observed outcome.

$$E^{SW}[Y(\underline{a})] = E^{SW}[Y(\underline{a}) \mid \underline{A} = \underline{a}] = E^{SW}[Y \mid \underline{A} = \underline{a}]. \tag{30}$$

Of course, there is no guarantee that the counterfactual potential outcome mean in the reweighted data is the same as in the original data. The following lemma proves that fact.

Lemma 2. *If the observed and counterfactual data meet assumptions* (1), (2), *and* (3), *then*

$$E^{SW}[Y(\underline{a})] = E[Y(\underline{a})]. \tag{31}$$

Proof. This is a result of Theorem 3.2 of Robins (1997). That theorem states that the marginal mean of

the counterfactual outcomes is given by the g-computational formula, given sequential ignorability. The g-computational formula has

$$E[Y(\underline{a})] = \int \cdots \int E(Y|\underline{A} = \underline{a}, \underline{X}) \prod_{t=0}^{T} f(X_t \mid \underline{A}_{t-1} = \underline{a}_{t-1}, \underline{X}_{t-1}) dx_1 \cdots dx_T. \tag{32}$$

Note that this formula does not depend on the distribution of $\underline{A}$, which is the only way in which f and f^{SW} differ. Thus, their marginal distributions must be identical. □

Recall the marginal structural model is the following restriction on the marginal mean of the potential outcomes:

$$E[Y(\underline{a})] = g(\underline{a}; \beta). \tag{33}$$

We now define the marginal structural model estimator $\hat{\beta}$ for the true causal parameters, β_0 as the solution to the estimating equation

$$\sum_i q(\underline{A}_i) \left\{ Y_i - g(\underline{A}_i; \hat{\beta}) \right\} SW_i = 0, \tag{34}$$

where $q(\underline{A}_i)$ is any function with the same dimension as β. For example, if we specified a linear MSM with $g = (1, A_t', A_{t-1}')'\beta$, then $q = (1, A_t', A_{t-1}')'$ would be the function corresponding to weighted least squares, with weights SW_i. In the above analysis of negative campaigning, this is the estimation procedure.

Theorem 4. *In a model characterized by a* MSM $E[Y(\underline{a})] = g(\underline{a}; \beta_0)$*, with data meeting assumptions* (1)*,* (2)*, and* (3)*. Then, subject to regularity conditions, the solution to the estimating equation* (34) *is a consistent and asymptotically normal estimator* $\hat{\beta}$ *for* β_0*.*

Proof. We can combine Lemmas 1 and 2 to find the result. Note that:

$$g(\underline{a};\beta) = E[Y(\underline{a})] \tag{35}$$

$$= E^{SW}[Y(\underline{a})] \tag{36}$$

$$= E^{SW}[Y \mid \underline{A} = \underline{a}]. \tag{37}$$

The first equality is by the definition of the model, the second from Lemma 2, and the third from 1. Using the properties of conditional means, it must be the case that $g(\underline{a};\beta)$ is the unique function $c(\underline{A})$ such that

$$E^{SW}\Big[q(\underline{A})\big\{Y - c(\underline{A})\big\}\Big] = 0,$$

for all functions $q(\underline{A})$. The definition of the weighting implies that this equivalent to the condition:

$$E\Big[q(\underline{A})\big\{Y - g(\underline{A};\beta_0)\big\}SW\Big] = 0. \tag{38}$$

We can expand the estimator in the following way:

$$\begin{aligned}
0 &= \sum_{i=1}^{n} q(\underline{A}_i)\big\{Y_i - g(\underline{A}_i;\hat{\beta})\big\}SW_i \\
&= \sum_{i=1}^{n} q(\underline{A}_i)\big\{Y_i - g(\underline{A}_i;\beta_0)\big\}SW_i + \left[\sum_{i=1}^{n} q(\underline{A}_i)G(\underline{A}_i;\beta^*)SW_i\right](\hat{\beta} - \beta_0) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(\underline{A}_i)\big\{Y_i - g(\underline{A}_i;\beta_0)\big\}SW_i + \left[\frac{1}{n}\sum_{i=1}^{n} q(\underline{A}_i)G(\underline{A}_i;\beta^*)SW_i\right]\sqrt{n}(\hat{\beta} - \beta_0),
\end{aligned} \tag{39}$$

where G is the matrix of partial derivatives of g with respect to the parameters, β:

$$G(\underline{A};\beta) = \frac{\delta g(\underline{A};\beta)}{\delta\beta},$$

and β^* is a value between $\hat{\beta}$ and β_0. Suppose two regularity conditions hold: that $V_{\beta_0} = E[q(\underline{A})G(\underline{A};\beta_0)SW]$

is positive semi-definite and that $|q(\underline{A})G(\underline{A};\beta_0)SW|$ is locally dominated in neighborhood of β_0, which implies

$$\frac{1}{n}\sum_{i=1}^{n} q(\underline{A}_i)G(\underline{A}_i;\beta^*)SW_i \xrightarrow{p} V_{\beta_0}.$$

Combining these conditions with (39), we have

$$\sqrt{n}(\hat{\beta}-\beta_0) \xrightarrow{p} -V_{\beta_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} q(\underline{A}_i)\left\{Y_i - g(\underline{A}_i;\beta_0)\right\}SW_i,$$

where, due to the Central Limit Theorem and (38) the right-hand side is normal with mean 0 and variance

$$V_{\beta_0}^{-1}E\left[\left(\left\{Y-g(\underline{A};\beta_0)\right\}SW\right)^2 q(\underline{A})q(\underline{A})'\right](V_{\beta_0}^{-1})'.$$

□

Robins (2000) shows that this theorem holds when we replace the true stabilized weights with estimated weights based on a correctly specified parametric model. Correct in this case means that $\text{plim}(\widehat{SW}_i) = SW_i$. The asymptotic variance of the estimator changes, though, and can be hard to calculate. In this case, it is easier to simply use a non-parametric bootstrap to estimate the standard errors. Robins (2000) also shows that "robust" standard error estimates based on sandwich estimators will be conservative estimates of the true standard errors.