Spring 5-17-2019

Estimating Controlled Direct Effects through Marginal Structural Models

Silvia Michelle Torres Pacheco
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS
Department of Mathematics and Statistics

Estimating Controlled Direct Effects through Marginal Structural Models

by
Silvia Michelle Torres Pacheco

A thesis presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Master of Arts

May 2019
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

This thesis closes a crucial stage of my professional life. Already at a very early age I fell in love with numbers, formulas, graphs and Math. I have always been amazed by the power of numbers, and I felt a bit conflicted that by choosing political science as my main career, I would have to give up Math. Luckily, the mentors and teachers that have accompanied me since my college days have always made sure that I never stray too far from statistics and Math.

However, there have been several times where I thought I did not have the necessary skills, and I doubted whether my stubbornness and efforts to learn more and be good in statistics were not going to be enough to achieve my goals. Again, I was very lucky to be surrounded by amazing people that not only had the hard task of teaching and training me, but that also took the job of empowering and encouraging me in every single stage of the process.

Thanks to Javier Aparicio, Allyson Benton, Jorge Buendía, Javier Márquez and Lorena Ruano for always believing in me and for not only giving me strong theoretical foundations, but also for nurturing my passion for science.

Special thanks to my professors in Political Science, Jacob Montgomery, Andrew Reeves, Guillermo Rosas, Jon Rogowski, Margit Tavits, Betsy Sinclair, and Jeff Gill, who encouraged me to keep going and that always had the most thoughtful feedback and advice. Also, thanks to those inspiring and brilliant women that have deeply inspired my work and objectives. Thanks for your efforts and work to close gaps and for opening doors to women like me who admire you and aim to one day be as amazing as you are: Lonna Atkeson, Adriana Crespo-Tenorio, Jimin Ding, Erin Hartman, Diana O'Brien, Xun Pang, Molly Roberts, Betsy Sinclair, Rociío Titiunik, Vera Troeger and Hannah Wallach.

This thesis received a lot of feedback from the most amazing groups of scholars. Thanks

to the Political Data Science Lab at WashU, and to people like Jonathan Homola, Erin Hartman, Kosuke Imai, Marc Ratkovic, Tess Wise, and Matt Blackwell. All the potential good parts of this manuscript can be attributed to them.

I could not have progressed through this program without the help and support of my Math professors: Todd Kuffner, José Figueroa-López, Nan Lin, Sid Chib, and Jimin Ding. I learned so much from you and I am eternally grateful for your patience. Your classes made me a stronger scholar.

Thanks to my amazing family. In particular, thanks to my parents, Silvia and Miguel, who have always believed in me regardless of the craziness of my dreams and goals. Thanks for giving me everything you had to help me achieve them. Similarly, thanks to my amazing husband, Jonathan, because without his support and infinite love I could not have finished a single problem set, let alone a thesis (this is quite literal).

Finally, this thesis is for all the women and girls that are out there double doubting themselves but wanting to pursue a career or a degree in STEM. Do not let anybody tell you that you are not good enough, or that you do not belong in this field. The road is far from perfect and easy, but we need you to fight this battle, and a lot of us will be there by your side to help you.

Michelle

Washington University in St. Louis

May 2019

Para Camila, Chio, Sofía, María y Valeria.

Porque estoy segura que ustedes van a cambiar al mundo.

ABSTRACT OF THE THESIS

Estimating Controlled Direct Effects in Panel Data with Marginal Structural Models

by

Silvia Michelle Torres Pacheco

Master of Arts in Statistics

Washington University in St. Louis, 2019

When working with panel data, many researchers wish to estimate the direct effects of time-varying factors on future outcomes. However, when a baseline treatment affects both the confounders of further stages of the treatment and the outcome, the estimation of controlled direct effects using traditional regression methods faces a bias trade-off between confounding bias and post-treatment control. Drawing on research from the field of epidemiology, in this thesis I present a marginal structural modeling (MSM) approach that allows scholars to generate unbiased estimates of controlled direct effects. Further, I detail the characteristics and implementation of MSMs, compare the performance of this approach under different conditions, and discuss and assess practical challenges when conducting them. After presenting the method, I apply MSMs to estimate the effect of wealth in childhood on political participation, highlighting the improvement in terms of bias relative to traditional regression models. The analysis shows that MSMs improve our understanding of causal mechanisms especially when dealing with multi-categorical time-varying treatments and non-continuous outcomes.

Chapter 1

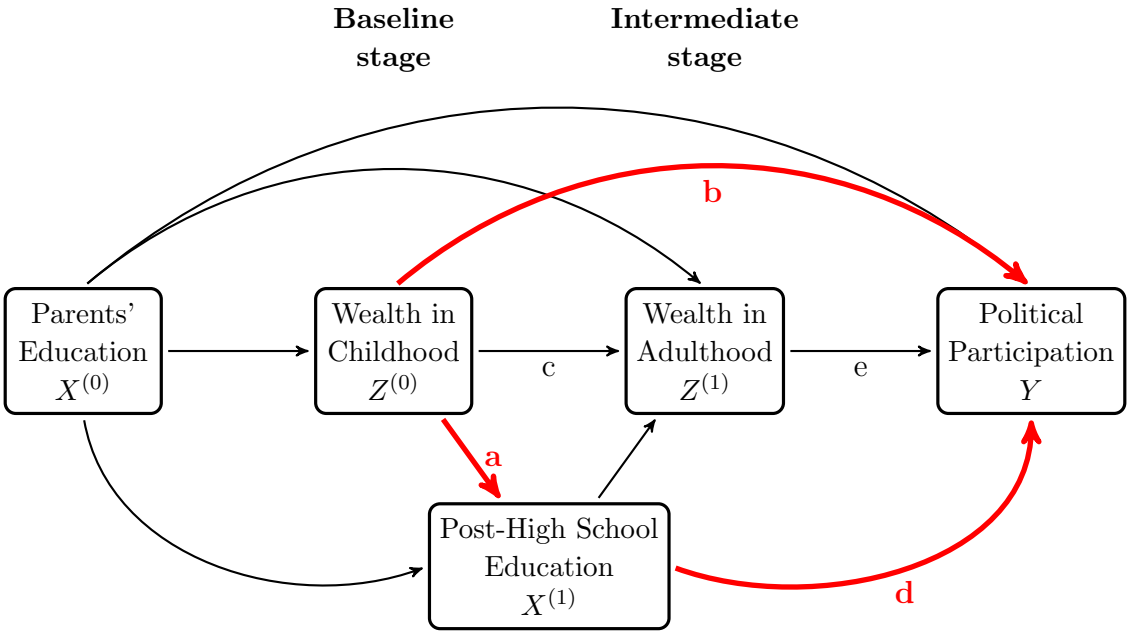
Introduction

In recent years, considerable progress has been made in providing methodological tools that allow political scientists to better estimate the causal effects of treatments on outcomes. However, in many cases, we are interested not in identifying the effect of a variable at one period, but rather in assessing effects in a dynamic setting. We might, for instance, observe units in multiple time periods and wish to estimate the independent effect of treatments at each stage on some future outcome. In estimating these effects, researchers can better understand not only *how* and *why* political phenomena are linked, but also the potential consequences of changing a treatment of interest that varies through time. Yet, standard tools in the literature are often ill suited for making valid causal claims in dynamic settings.

To provide some clarity to this discussion, consider the following example, which is depicted visually in Figure 1.1. Past research on political participation identifies wealth as a key factor that influences citizens' political participation (Almond and Verba 1989; Verba, Nie and Kim 1978). Typically, scholars emphasize the effect of wealth in adulthood (measured through self-reported income of adult respondents) as a provider of resources that ease participation (e.g., a car that helps a citizen to reach a polling station). However, a separate question is how wealth in early stages of life (measured through the income of a citizen's

parents) can affect political participation *independent of the effect of wealth in adulthood*. Finding such an effect would suggest that, for instance, children being raised in wealthier homes receive a lifelong boost in terms of socialization, cognitive skills, or psychological orientations towards politics that affect participation regardless of their own economic success later in life (Beck and Jennings 1982; Brady, Verba and Schlozman 1995; Currie 2008). Figure 1.1 shows a simplified version of this example and highlights this unmediated effect through the bolded red arrows (path b and path a-d).

Figure 1.1: DAG showing the relationship between a time-varying treatment (wealth) and outcome (political participation)



Note: The bold paths represent the controlled direct effect of the baseline treatment (wealth in childhood) on the outcome (political participation).

Although dynamic treatments in panel datasets abound in the political science literature, applied scholars have thus far been given little guidance as to how to proceed in such settings. Indeed, in situations such as the one depicted in Figure 1.1 traditional regression techniques offer no way to consistently estimate causal effects. Specifically, researchers wishing to estimate the effect of treatments at various time points in dynamic processes using traditional

regression confront a bias trade-off between *confounding bias* and *post-treatment bias*. Confounding bias, sometimes termed omitted variable bias, results from failure to control for important common causes of treatment and outcome when estimating causal effects—a confounder. Post-treatment bias arises from controlling for an intermediate variable that has been affected by the treatment—a post-treatment variable.

The key to understanding this tradeoff is that in order to estimate the unmediated effect of a treatment at the baseline stage (e.g., wealth in childhood/parents' income), one must simultaneously correctly estimate the effect of the treatment at intermediate stages (e.g., wealth/income in adulthood). Yet, to estimate the effect at the intermediate stage (path e in Figure 1.1), researchers must either fail to account for important confounders (e.g., attending college) or include such confounders as "control" variables. The former approach will induce confounding bias into our estimates by failing to control for a variable that is causally prior to both income in adulthood and political participation. The latter will introduce post-treatment bias by controlling for a variable that is itself affected by parents' income. In such cases, both controlling for intermediate confounders and failing to control for them will result in biased estimates.

While this problem has certainly not gone unrecognized in the broader statistics literature, these issues have received relatively little attention in political science. Some seemingly plausible approaches, such as mediation analysis (Imai, Keele and Tingley 2010; Imai et al. 2011), are unsuited to handling the dynamic relationship between treatment and confounding variables. Other approaches are more difficult to implement and offer little flexibility. Structural nested mean models, as presented by Acharya, Blackwell and Sen (2016), for instance, are only suitable when the outcomes are continuous.

In this article, I draw on research from the field of epidemiology (Hernán, Brumback and Robins 2000; Robins, Hernán and Brumback 2000), to outline a marginal structural models (MSMs) framework for estimating controlled direct effects of multi-valued treatments at

different time periods that is both easy to implement and suitable for use with several data types. This class of models was introduced to political science by Blackwell (2013) and later discussed by Imai and Ratkovic (2014). Still, their studies mainly focus on the estimation of *cumulative effects* of dynamic treatments, and offer little discussion on the applicability of these models to the estimation of controlled direct effects. Further, I extend previous work by testing and addressing practical challenges of this method, such as the tools for weight estimation, implications and use of weights, and consequences of the violation of the main assumptions.

MSMs overcome the bias-tradeoff dilemma described above by using an inverse probability of treatment weighted (IPTW) estimator. This allows researchers to account for confounders while avoiding directly controlling for post-treatment variables (Blackwell 2013; Blackwell and Glynn 2014; Robins, Hernán and Brumback 2000). By estimating correct weights, researchers are able to create pseudo-samples that are balanced with respect to confounders and therefore allow for consistent estimation of causal quantities of interest. Importantly, unlike previous methods in the political science literature, I present and detail the implementation of MSMs to estimate controlled direct effects with multi-categorical treatments as well as non-continuous outcomes.

In the next section, I define controlled direct effects and discuss the challenges that researchers face when estimating them in a dynamic setting. I then provide an overview of MSMs—and the assumptions that undergird them—and explain how they allow for unbiased estimates of treatment effects. As part of this presentation, I not only detail important elements of the implementation of these models but also provide guidance for the weighting process, and for the cases in which assumptions are not fully fulfilled. The section also includes a discussion of the advantages of MSMs over other alternatives, especially traditional regression models. Finally, I present an application that compares the inferences reached by MSMs and traditional models regarding the causes of political participation. This application

focuses on the estimation of the controlled direct effect of wealth in childhood, as measured by parents' income, on political activism using a panel survey that spans over 30 years.

Chapter 2

Controlled direct effects and bias trade-off

To formally articulate the difficulty of estimating causal effects in a dynamic setting, I return to the example depicted in Figure 1.1. We are interested in calculating the effect of wealth in childhood on political participation that is not mediated by wealth in adulthood. The effect of economic resources on political participation has been widely studied. However, recent studies have recognized and focused on the cumulative and long term effects that economic conditions in childhood may have on participation in later stages of life. For example, Ojeda (Forthcoming) finds that it is possible to identify two participations gaps with different sizes and implications: one that childhood economic history generates, and another caused by income in adulthood.

For the illustration and application presented below, I measure wealth of an individual using her own income, and her parents' income.[1] In Figure 1.1, this effect is represented by the highlighted paths (a-d and b). Substantively, this will allow us to explore the impact of early economic conditions on adult political participation independent of the level of

[1] *Wealth* is the treatment of interest in two different stages: childhood and adulthood. While "wealth" can imply multiple factors, there is a strong correlation between wealth indexes and income (Córdova 2009).

affluence later in life.[2] In other words, if we could fix (or control) respondents' income in adulthood to a specific level, what would be the effect of changes in parents' income on adult political activity? This quantity is known as the controlled direct effect (CDE) which I define formally below (Pearl 2001, 2011; VanderWeele 2009). This estimand is useful to 1) understand the mechanism through which treatments affect the outcome, and 2) explore the different effects that treatment regimes have on an outcome. The estimation of controlled direct effects is relevant to address several social science questions: the analysis of the effects of historical institutions on current economic and political conditions (e.g. "zoning" on political participation), the study of issues related to public policy (e.g. the impact of welfare programs on economic development), or the exploration of early conditions of citizens on their current political attitudes and behavior.

2.1 Defining controlled direct effects

Our goal is to estimate the causal effect of a treatment Z (income) at different "stages" in time. Although the model can be easily extended to allow for multiple stages, I focus on only two stages of treatment: parents' income ($t = 0$) and income in adulthood ($t = 1$). For this discussion I assume that the measurement of wealth of an individual i, income,[3] in both stages can be either low ($Z_i^{(t)} = 0$), middle ($Z_i^{(t)} = 1$) or high ($Z_i^{(t)} = 2$).[4] Finally, I assume that the education of each subject's parents and their level of post-High School education are the sole confounders, which means that these variables are affecting both treatments and

[2]It is important to highlight that while this case considers a "sequence" of conceptually similar treatments, researchers can also use this method for sequences of semantically different variables as long as they hold a clear causal relationship (i.e. one precedes and affects the other).

[3]The categorization of a continuous variable such as income is a common practice in multiple fields. One important reason is measurement. In order to decrease non-response and increase perceptions of privacy, respondents generally choose their income from multiple categorical options defined by the researcher. Also, conceptually, researchers are generally interested in the differences between levels of income rather than in its unitary nature (Córdova 2009; Moore and Welniak 2000).

[4]Under the assumption that $Z_i^{(t)}, X_i^{(t)}$ and Y_i are sampled iid from a population, I treat them as random and therefore avoid the subindex i.

the outcome.

The outcome of interest is an individual's level of political participation, denoted Y. Let $Y_{Z^{(0)}=a}$ be the subject's level of political participation if parents' income ($Z^{(0)}$) is set to a value a. Thus, $Y_{Z^{(0)}=0}$ represents the outcome when the respondent's parents have a low income, whereas $Y_{Z^{(0)}=1}$ and $Y_{Z^{(0)}=2}$ represent the response of the same respondent if her parents' income was medium and high respectively. Since only one of the possible values will be observed for each individual, then two of the values of Y are *potential* outcomes while the other is the *observed* outcome. Similarly, the intermediate treatment stage $Z^{(1)}$, income in adulthood, can also take on three values. Therefore, let $Y_{Z^{(0)}=a,Z^{(1)}=b}$ denote the level of political participation of a subject if her parents' income and income in adulthood were set to values of a and b respectively.

With this notation, we define the CDE by "fixing" the second-stage of treatment to a specific value (Pearl 2001; VanderWeele 2009). It is important to highlight that this "fixing" assumes that the researcher has the capacity to artificially manipulate the intermediate stages. In practice, the estimation of the CDE is especially useful for policy design and experimental settings where researchers have the chance of manipulating the treatment stages. For example, Akee et al. (2018) manipulates the assignment of unconditional money transfers at different stages of life to study the effect that income has on civic participation. Although this option is not easily available for social scientists, and especially for those dealing with observational data, this quantity is still useful to have a better understanding of the potential outcomes that different treatment combinations generate. Controlled direct effects aid with the operationalization and analysis of the core concept of causal inference: the definition and modeling of *counterfactuals*. Therefore, the value of such estimand should not be underrated, even in the cases where the manipulation of any of the treatment stages is not possible.

We formally define the CDE as:

$$CDE = Y_{Z^{(0)}=a,Z^{(1)}=b} - Y_{Z^{(0)}=a',Z^{(1)}=b}. \tag{2.1}$$

Conceptually, the CDE estimand represents the effect of a treatment at a specific time period while controlling the level of treatment at different stages. In this example, we are interested in the CDE for the baseline treatment ($t = 0$). Of course, we cannot directly calculate the CDE since the counterfactual values are not observed. However, with standard regularity assumptions, we can provide an unbiased estimate of the CDE by calculating the *average controlled direct effect* (ACDE)

$$\begin{aligned} ACDE &= \mathbb{E}(Y_{Z^{(0)}=a,Z^{(1)}=b} - Y_{Z^{(0)}=a',Z^{(1)}=b}) \\ &= \mathbb{E}(Y_{Z^{(0)}=a,Z^{(1)}=b}) - \mathbb{E}(Y_{Z^{(0)}=a',Z^{(1)}=b}), \end{aligned} \tag{2.2}$$

where $\mathbb{E}(\cdot)$ refers to the expectation over the individuals in the sample. This is simply the difference between the average outcomes for units that received different treatments (a and a') at stage $t = 0$ while holding the second stage constant at b.

2.2 The bias trade-off

Although in theory the ACDE seems relatively straightforward to calculate, in practice it is not. In fact, there is actually no way to correctly estimate the ACDE using standard regression techniques. The dilemma is the following: since we want to estimate the effect of the treatment at each stage of the treatment sequence separately, we must estimate a coefficient representing the effect of parents' income and another one for the effect of income in adulthood. In order to generate unbiased estimates we must control for all confounders—the set

of variables that affect both the treatment and the outcome—in order to avoid *confounding bias*. However, some of the confounding variables for the intermediate-level treatments are themselves affected by the baseline treatment. Therefore, controlling for these covariates will introduce post-treatment bias into our estimates (Elwert and Winship 2014; Montgomery, Nyhan and Torres Forthcoming; Rosenbaum 1984; Rubin 1977). As a consequence, we have a situation where *both* controlling for and *not* controlling for confounders will result in biased estimates of the ACDE.

To make this tradeoff clearer, I return to the example depicted in Figure 1.1. In this instance, the problematic variable is *post-High School education*. Why is it necessary to include this variable in the model in the first place? The answer is that the assignment of the treatment in observational studies is not random. In this example, both having a high levels of wealth and political participation are dependent on other factors such as levels of educational attainment. The implication of non-random assignment to treatment is that the observed differences in the outcomes between treated and untreated groups cannot only be attributed to the presence of the treatment but potentially also to inherent differences between the two groups. Therefore, once we identify all confounders, a necessary step is to account for this imbalance. In a standard regression, this would be done by including education as a control variable.

However, including post-High School education as a control variable results in a different problem: post-treatment bias. In our example, whether or not respondents seek post-secondary education is itself caused (in part) by the baseline treatment (wealth in childhood). In the language of causal inference, education is therefore a "collider" (Elwert and Winship 2014), and controlling for it in a regression will bias estimates of causal effects.

In summary, when confounders are affected by a baseline treatment we face an inevitable bias trade-off: excluding problematic confounders leads to omitted variable bias, but including them leads to post-treatment control bias. Although not always recognized, this

trade-off and its consequences are frequently encountered in political science research. If we are dealing with panel or longitudinal data, then it is natural to identify treatments varying through time and complex interactions between those treatment stages and confounders that are not static. In the next chapter, I explain how adopting a marginal structural modeling framework allows us to address confounding bias *without* introducing post-treatment bias.

Chapter 3

Estimating CDE using marginal structural models

Marginal structural models (MSMs) are a class of models used to estimate the causal effect of time-varying treatments such as medicine prescription or medical procedures histories (Hernán, Brumback and Robins 2001; Robins 1999*a*; Robins, Hernán and Brumback 2000). Classic applications have focused on estimating the cumulative effects of these time-varying treatments on future outcomes, and previous applications of MSM in political science focused on estimating these cumulative effects (Blackwell 2013). My presentation below builds on more recent work by researchers who have extended the MSM framework to also estimate controlled direct effects and, under certain conditions, natural direct and indirect effects (Nandi et al. 2012; VanderWeele 2009). I cover and detail cases where the treatment is multi-valued and the outcome is non-continuous, to address questions relevant to political science using panel survey data.

In general, MSMs are useful when dealing with cases where (1) the treatment takes few values, (2) there exists a covariate that acts both as determinant of the outcome of interest and as a predictor of an intermediate stage of treatment, and (3) past exposure to baseline

treatment predicts subsequent levels of this covariate. As I reviewed above, the decision to control or not control for these covariates inevitably leads to either confounding or post-treatment bias. However, through an inverse probability of treatment weighted (IPTW) estimator, MSMs provide unbiased estimates once we meet certain assumptions. The core idea of these models is that through the weights estimated via IPTW, we create a "pseudo-population" consisting of copies of each subject in the sample. This pseudo-population has two important features: first, the probability of receiving the second stage of the treatment is unconditional on the confounders affected by the baseline treatment eliminating the necessity of controlling for them in the final model. And second, the potential outcomes are the same as in the true population allowing the estimation of unbiased causal effects (Robins, Hernán and Brumback 2000). The number of replicas in the pseudo-sample is calculated based on the probability of observing a particular sequence of treatment conditional on relevant confounders (Robins 1999*a*).

Before providing the details of the method, it is important to note that several previous scholars have applied models closely related to MSMs in political science. Perhaps the earliest example is Glynn and Quinn (2010), who introduced and extended the IPTW approach for estimating causal effects in a cross-sectional setting. After Blackwell (2013) formally introduced MSMs to political science, Imai and Ratkovic (2015) generalized the covariate balancing propensity score to dynamic settings to achieve a more balanced pseudo-population. More recently, Samii, Paler and Daly (2017) applied the IPTW framework for estimating causal effects using a machine learning approach for assigning treatment weights. However, the method most closely related to the objective described here, the estimation of CDE, are structural nested mean models (SNMMs), which were recently introduced to political science by Acharya, Blackwell and Sen (2016). I provide a brief discussion comparing and contrasting the MSM and SNMM approaches for estimating controlled direct effects in section 3.4 after I detail the model.

3.1 Assumptions

Going back to our example, MSMs allow us to model levels of political activism of individuals receiving each of the potential *Parents' income-Income in adulthood* sequences: low-low, low-middle, low-high, middle-low, middle-middle, middle-high, high-low, high-middle and high-high. However, modeling these unconditional (or marginal) distributions requires the fulfillment of two assumptions.

The first is the *sequential ignorability condition*, which guarantees the necessary statistical exogeneity for the identification of causal effects (Robins 1999*a*).[1] In essence, this assumption is an extension of a general condition for the estimation of causal effects in single-stage settings: controlling for confounders $X^{(t)}$ assures independence ($\coprod$) of the potential outcomes $Y_{Z^{(0)},Z^{(1)},\ldots,Z^{(T)}}$ from the treatment $Z^{(t)}$. For the multi-stage setting, we need to meet this same condition for each treatment stage. In our example, this would mean controlling for education of the parents (denoted here as $X^{(0)}$) to avoid confounding of parents' income—the first treatment stage $Z^{(0)}$. Formally,

$$Y_{Z^{(0)},Z^{(1)}} \coprod Z^{(0)}|X^{(0)}. \tag{3.1}$$

For the second stage, it is necessary not only to control for education of the parents, $X^{(0)}$, and post-High School education, $X^{(1)}$, to avoid confounding bias, but also to include parents' income, $Z^{(0)}$, as another confounder of wealth in adulthood, $Z^{(1)}$, and participation, Y. In other words, the outcome needs to be independent of any stage in the treatment sequence,

[1]Note that if there are multiple confounders, the values of $X^{(0)}$ and $X^{(1)}$ are going to be in matrix form rather than vectors.

conditional on past confounders and treatments,[2]

$$Y_{Z^{(0)},Z^{(1)}} \coprod Z^{(1)} | Z^{(0)}, X^{(0)}, X^{(1)}. \tag{3.2}$$

The second assumption is the *positivity assumption* which states that a treatment value should not be limited to a single level l of the control variables. Intuitively, this means that all subjects in the sample must have a non-zero probability of getting exposure to the different levels of treatment. In our example, the assumption implies that an individual that did not attend college and whose parents had a low income should still have a non-zero chance of receiving a middle or high income as an adult.[3] Formally,

$$\text{If } \Pr(Z^{(0)} = z^{(0)}, (X^{(0)}, X^{(1)}) = (x^{(0)}, x^{(1)})) > 0, \text{ then} \tag{3.3}$$

$$\Pr(Z^{(1)} = z^{(1)} | (X^{(0)}, X^{(1)}) = (x^{(0)}, x^{(1)}), Z^{(0)} = z^{(0)}) > 0. \tag{3.4}$$

Once we meet these assumptions, we can use MSMs to estimate the ACDE.[4][5]

[2]We can define this assumption more generally as $Y_{Z^{\overrightarrow{(t)}}} \coprod Z^{\overrightarrow{(t)}} | Z^{\overrightarrow{(t-1)}}, X^{\overrightarrow{(t)}}$. Where $\rightarrow$ indicates the treatment or covariate regime up to the time indicated in parentheses.

[3]More generally, if $\Pr(Z^{\overrightarrow{(t-1)}} = z^{\overrightarrow{(t-1)}}, X^{\overrightarrow{(t)}} = x^{\overrightarrow{(t)}}) > 0$, then $\Pr(Z^{\overrightarrow{(t)}} = z^{\overrightarrow{(t)}} | X^{\overrightarrow{(t)}} = x^{\overrightarrow{(t)}}, Z^{\overrightarrow{(t-1)}} = z^{\overrightarrow{(t-1)}}) > 0$.

[4]On the one hand, the fulfillment of the first condition can be difficult given that there is no technique that allows us to diagnose the degree to which it is met. However, this is a classic (and necessary) assumption in any causal analysis. *Naïve* regression estimators are not exempt from meeting the ignorability assumption either. Furthermore, previous work by Blackwell (2013) and VanderWeele (2010) includes the development of sensitivity analyses that allow us to assess the strength of the inferences made from MSMs. On the other hand, the fulfillment of the positivity assumption can be difficult in cases where there is a continuous treatment and confounders, and then alternatives like SNMMs are preferred (VanderWeele 2009).

[5]The simulations in the Appendix show how the bias and variance of the ACDE change depending on mild to strong violations of these assumptions. See discussion below.

3.2 Benefits of the pseudo-sample

MSMs aim to model the potential outcomes for the different sequences of treatment. This strategy allows for the estimation of controlled direct effects. For example, consider the following model:

$$\mathbb{E}[Y_{Z^{(0)},Z^{(1)}}] = \alpha_0 + \alpha_1 Z^{(0)} + \alpha_2 Z^{(1)}. \tag{3.5}$$

The ACDE in model 3.5 is the expected value of the differences in Y when $Z^{(0)}$ is 1 and when $Z^{(0)}$ is 0, while fixing $Z^{(1)}$ to b. Then,

$$\begin{aligned}\mathbb{E}[Y_{Z^{(0)}=1,Z^{(1)}=b} - Y_{Z^{(0)}=0,Z^{(1)}=b}] &= \alpha_0 + \alpha_1 \cdot 1 + \alpha_2 \cdot b - (\alpha_0 + \alpha_1 \cdot 0 + \alpha_2 \cdot b) \\ &= \alpha_1(1-0) = \alpha_1. \end{aligned} \tag{3.6}$$

In other words, when the second treatment stage is set to b, the baseline stage has a causal effect of α_1 on the outcome. This estimation only holds if the differences we observe in Y are only related to the treatment and not to other confounders. From previous chapters, we know that in our example, as in all observational studies, this is not true. Wealth in each of the two stages is not randomized: the levels of this "treatment" are not independent from past economic conditions or education. The implication is that each income sequence has different probabilities of being observed given the values of the confounding factors (e.g., a subject with a college degree is more likely to have a higher income than one that only completed High School). MSMs use these probabilities to build weights that balance the sample across treatment groups. The weights are the product of two components, one per treatment stage, defined as follows:

$$\mathcal{W}(t) = W_{Z^{(0)}} \times W_{Z^{(1)}} = \frac{f(Z^{(0)}|X^{(0)})}{f(Z^{(0)})} \times \frac{f(Z^{(1)}|Z^{(0)}, X^{(0)}, X^{(1)})}{f(Z^{(1)}|Z^{(0)})}. \tag{3.7}$$

The numerator in each of the components of the $\mathcal{W}(t)$ term is the probability that an individual received his own observed treatment at time t, $Z^{(t)}$, given his own past treatment (up to $t-1$) and covariate history up to point t (Robins 1999*a*). For example, in the income example, the numerator of $W_{Z^{(1)}}$ is simply the probability that an individual has her own observed income in adulthood conditional on her observed parents' income, and educational attainment after High School.[6] At the same time, the denominator is the probability that a subject received her observed treatment at time t but only conditional on her treatment sequence until $t-1$. In the example, the denominator of $W_{Z^{(1)}}$ is the probability of observing the actual income in adulthood but only conditional on parents' income.[7]

Once we obtain these weights, we estimate the parameters in Equation 3.6 using a weighted least squares regression in which each subject is given as a weight the inverse of her corresponding $\mathcal{W}(t)$. For illustrative purposes we implement a weighted regression, however the researcher has full flexibility to model the outcome as long as it is applied to the weighted sample. Thus, the model can range from a simple weighted mean to a complex non-parametric weighted model.[8] The weighted model handles confounding while avoiding explicit conditioning and post-treatment bias. How does weighting achieve this? Recall that treatment sequences have different probabilities of being observed given the values of the confounders. By weighting, we are "leveling the field" and breaking the link between the second treatment stage and its confounders: the problematic variables affected by the baseline treatment. To have a more intuitive understanding of this, we can view the pseudo-population

[6]Note that if it is the beginning of the sequence, $t = 0$, then the numerator would only be conditional on the confounders of $Z^{(0)}$ and Y. That is, $f(Z^{(0)}|X^{(0)})$.

[7]The denominator of this quantity can be replaced with another function of treatment history. This would not affect the consistency or unbiasedness of the estimator. The numerator is introduced as a "stabilizer" of weights in order to avoid extreme values. The efficiency of the estimator can be influenced by the decision for the numerator. However, the selected function should not include the intermediate or confounding variables in the model.

[8]The comparison of weighting methods used to model the outcome escapes the scope of this thesis. However, as in any other study, researchers should select the appropriate modeling technique based on a deep understanding of the data and full awareness of the assumptions and trade-offs that the different methods convey.

as a sample composed of each individual in the original population plus $(\mathcal{W}(t)^{-1} - 1)$ copies of themselves.

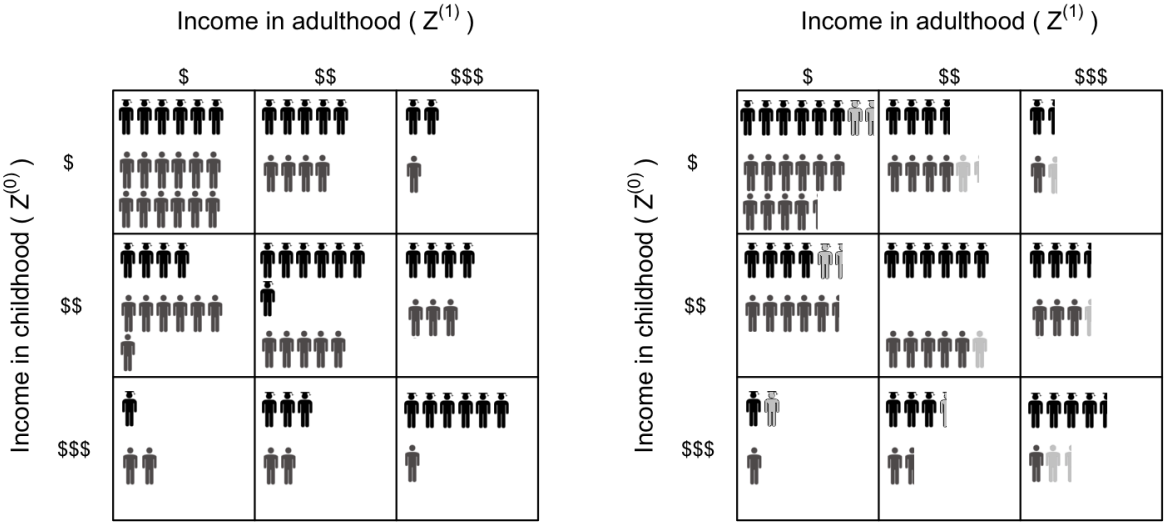
Consider the following hypothetical example based on the income case.[9] The first panel of Figure 3.1 shows the distribution of subjects in the original sample across the different levels of parents' income and adulthood income as well as education. Each human figure represents 1,000 individuals. Each cell represents a potential combination of income in childhood and adulthood: low ($), middle ($$) or high ($$$). Furthermore, the level of post-High School education is indicated by the hat and color of the figures: black symbols wearing a hat attended college while gray figures did not. In this example, we assume that sequential ignorability holds and, as the picture shows, the positivity assumption is met (there is at least one human figure in all possible combinations of parents' income, income in adulthood and education).

Just by visual inspection, it is clear from the figure that the probability of, for example, receiving a high income in adulthood is strongly determined by both levels of parents' income and education. Table A.1 presents the information by stratum and actual probabilities of receiving a particular income in adulthood $Z^{(1)}$ given parents' income and education.[10] For example, the probability of having a high income in adulthood if a subject has a high income in childhood but does *not* attend college is 1,000/5,000=0.2 (bold cell over sum of light-shaded cells in Column 7 of Table A.1). However, the probability of having a high income in adulthood when parents' income is high but the subject *also* attends college is much higher, 6,000/10,000 = 0.6. In other words, we have unbalance across levels of educational attainment. From Column 3 of Table 3.2 (labeled as "Original"), where we can see a summary of these probabilities for all strata, we can conclude that income in adulthood is not independent of levels of education, but that it acts as a confounder of this variable

[9]This example is based on one designed by Robins (1997).

[10]For sake of space, the table with the full set of strata is presented in the Appendix.

Figure 3.1: Distribution of individuals based on treatment sequence (parents' income and income in adulthood) and confounder affected by treatment (post high-school education)



(a) Original population

(b) Pseudo-population

Note: Each figure represents 1,000 individuals. Black figures with hat indicate that those subjects attended college, while gray figures only completed high-school. The panels show the distribution of respondents across treatment conditions.

and political participation.[11]

However, we can eliminate this unbalance by creating a pseudo-population based on copies (or reductions) of the subjects in the original sample using the inverse of the weights $\mathcal{W}(t)$. The column labeled as $\mathcal{W}(t)^{-1}$ in Table 3.2 presents this quantity and all the information necessary to construct it. Based on this information we can build the pseudo-sample shown in the second panel of Figure 3.1. We can now repeat the same exercise of calculating the probabilities of having a high income in adulthood for the individuals in the new sample. Column 4 of Table 3.2 (labeled as "Pseudo") presents these new estimated probabilities. It is important to highlight that while the calculation of these probabilities involved a simple

[11]If we ignore this confounder, the potential differences that we could observe in levels of political participation between groups defined by the different levels of income could not be attributed to the effect of this variable but to the differences in education levels.

Table 3.1: Calculation of weights for each stratum in sample

$Z^{(0)}$	$Z^{(1)}$	$X^{(1)}$	$f(Z^{(1)}\|Z^{(0)})$	$f(Z^{(1)}\|Z^{(0)}, X^{(1)})$	$\mathcal{W}(t)^{-1}$	Original-pop N	Pseudo-pop N
0	0	0	0.6	0.706	0.85	12000	10200
0	0	1	0.6	0.462	1.3	6000	7800
0	1	0	0.3	0.235	1.275	4000	5100
0	1	1	0.3	0.385	0.78	5000	3900
0	2	0	0.1	0.059	1.7	1000	1700
0	2	1	0.1	0.154	0.65	2000	1300
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2	0	0	0.2	0.4	0.5	2000	1000
2	0	1	0.2	0.1	2	1000	2000
2	1	0	0.333	0.4	0.833	2000	1667
2	1	1	0.333	0.3	1.111	3000	3333
2	2	0	0.467	0.2	2.333	**1000**	2333
2	2	1	0.467	0.6	0.778	6000	4667

Note: $Z^{(0)}$ is parents' income where 0 is low, 1 is middle and 2 is high. $Z^{(1)}$ is income in adulthood where 0 is low, 1 is middle and 2 is high. $X^{(1)}$ is post-High School education where 0 is no college and 1 is college. The full table with the probabilities and weights for the full set of treatment and covariate combinations can be found in the Appendix.

Table 3.2: Probabilities of having a high income in adulthood

		Original	Pseudo
$Z^{(0)}$	$X^{(1)}$	$\Pr(Z^{(1)} = 2\|X^{(1)}, Z^{(0)})$	$\Pr(Z_p^{(1)} = 2\|X_p^{(1)}, Z_p^{(0)})$
Low income	No college	0.059	0.1
Low income	College	0.154	0.1
Middle income	No college	0.2	0.233
Middle income	College	0.267	0.233
High income	No college	0.2	0.467
High income	College	0.6	0.467

Note: $Z^{(0)}$ is parents' income, $Z^{(1)}$ is income in adulthood and $X^{(1)}$ is post-high education.

stratification approach, cases with multiple confounders will require more intensive modeling techniques.[12]

Once we weight the sample, the probability of having a high income in adulthood is equal for both levels of education within each parents' income strata—the second treatment stage is

[12]Given that the unbiasedness of MSMs rely on an accurate estimation of the weights, different models will lead to different estimations of ACDEs. A brief comparison of different modeling tools for the estimation of weights is presented in Section 2.3.

balanced within parents' income and education groups. For example, a subject who did not attend college and whose parents had a high income has a probability of having a high income in adulthood of 2,333.33/5,000=0.467. Similarly, a subject that reports that her parents had a high income but that attended college has a probability of 4,667/10,000=0.467 of having a high income in adulthood. Thus, in the pseudo-population, the confounder $X^{(1)}$ does not predict the treatment at $t = 1$ given the baseline treatment. Post-High School education is no longer a confounder and we can assess the controlled direct effect of early income $Z^{(0)}$ on political participation.

The last step of this process consists of fitting a weighted regression of the outcome variable on both the baseline and intermediate treatments using the vector of weights $\mathcal{W}(t)^{-1}$. Other covariates can be included in this regression but these have to be strictly pre-treatment.[13]

3.3 Weighting: methods and implications

3.3.1 Estimation of weights

As I reviewed in the previous section, creating a balanced pseudo-sample involves an accurate estimation of the probabilities of observing the multiple treatment sequences conditional on covariate history. This implies an appropriate model specification of treatment assignment, and a suitable method to estimate probabilities. A proper specification of both treatment assignment and the relationship between confounders and treatment history is fundamental for the fulfillment of the sequential ignorability assumption.

[13]This decision should be strongly motivated by substantive and theoretical knowledge of the question under analysis, as well as by a deep understanding of the data. This is, if there are pre-treatment confounders, they should be included as part of the weight estimation as in any other model aiming to support causal claims. If fulfilled, the sequential ignorability assumption and subsequent weighting guarantee a pseudo-random assignment of the treatment stages. Therefore, covariate adjustment is not necessary. However, it tends to improve precision and reduce standard errors *if* the covariates are predictive of the outcome (Miratrix, Sekhon and Yu 2013).

In order to achieve this, as in any other study, researchers must have a deep understanding of their data, and solid knowledge of the theoretical framework regarding the relationship between their variables of interest. Further, the decision of how to model the multiple components of the dynamic framework under analysis should consider elements such as the number and nature of the variables under analysis, and the assumptions and trade-offs that each modeling tool implies.

In general, the most common alternative to estimate the assignment of a particular multi-category treatment sequence is a generalized linear model for categorical data. The simplicity of the model is its most attractive feature, but the trade-off between parsimony and strong predictive power that could potentially reduce bias has not been fully explored. Therefore, in this section I present a comparison of three different approaches to estimate weights, and an analysis that each of them yields when used in a MSM framework.

The main objective of this exercise is to compare the magnitude of the mean bias of the estimates of the ACDEs that come from four different models: a *naïve*, or saturated model that includes post-treatment covariates, and three MSMs using weights that were obtained using three different methods – an ordered logistic regression (ologit), a generalized additive model (GAM), and a random forest (RF). First, I simulate a dataset with $n = 1,000$ where the main outcome of interest is attendance of a rally (0=No, 1=Yes). The data includes two relevant sequences of covariates: parents' income and income in adulthood (the treatment sequence), and whether parents and respondent attended college (the confounders sequence). In this setup, college attendance of a subject acts as a confounder of income (second stage of the treatment) and rally attendance, but is also affected by parents' income (baseline treatment). The parameters are tuned to purposely allow for the possibility of observing samples in which the positivity assumption is not fulfilled. This is, there are combinations of the sequence treatment and college attendance that do not have any observations.[14] The

[14]The parameters and specification of the simulation are in the Appendix.

idea behind this specification is to illustrate the advantages of MSMs over traditional models even when one of the main assumptions that the former requires are mildly violated. Second, I use this data to estimate and record the ACDEs from four different models: the saturated or *naïve* model, and three MSMs that use weights calculated using ologit, GAM and RF models. The specification of the outcome model is the following:

$$Pr(Y = 1|\overrightarrow{\boldsymbol{Z}}) = \text{logit}^{-1}(\alpha_0 + \boldsymbol{Z^{(0)}\beta} + \boldsymbol{Z^{(1)}\gamma} + (\boldsymbol{Z^{(0)}} \times \boldsymbol{Z^{(1)}})\boldsymbol{\delta}) \tag{3.8}$$

Using this setting, I simulate 500 datasets and for each of the four models I record the differences between the estimated ACDEs and the *true* ones to obtain measures of bias. Figure 3.2 presents the average bias for each model of the nine potential ACDEs. In this figure, each corner of the polygon represents the mean difference between the true and estimated average controlled direct effect of the baseline treatment on the outcome, when fixing the intermediate treatment to a certain level. For example, CDE 1 represents the difference in probabilities of attending a rally between subjects that had a middle income in childhood ($Z^{(0)} = 1$) and others that had low income in childhood ($Z^{(0)} = 0$), when the income in adulthood is fixed to low ($Z^{(1)} = 0$). Further, the colored lines represent each of the four different models: the naïve model, and the three MSMs.

The analysis confirms that all of the MSMs perform significantly better than the saturated model (in blue), and provides useful information about the weighting methods and MSMs in general. First, there are no substantive differences between the weighing methods. The Random Forest shows a slightly better performance than the GAM or the ordered logit, but it does not seem to be a substantive difference. This is due to the simplicity of the example, where both the treatment assignment and outcome model are not complex. However, it is worth noting that these differences might be higher in cases with larger sets of confounders, more complex interactions and relationships between the variables, or where distributional

assumptions are harder to meet. For example Montgomery and Olivella (Forthcoming) show that regression trees yield better estimates of probability of treatment sequences in cases with multiple confounders, which in turn improve the pseudo-sample balance and the overall performance of the MSM. Second, although the mean bias for all treatment sequences is very close to zero, there are instances where this is not the case. Recall that the simulation setting is purposely designed to allow for samples where the positivity assumption is not fulfilled. In this case, although the expected bias is not zero, it 1) is small even under settings where the positivity assumption is violated, and 2) performs better than traditional regressions regardless of the method used for the weights estimation.

3.3.2 Practical considerations about weights

The act of weighting motivates multiple questions with potentially strong implications for the estimation of ACDEs. How should we proceed if the weights are too large (or small)? How do we account for uncertainty when estimating the weights? What implications does weighting have in terms of variance? What is the correct "modeling approach" when dealing with weighted samples? While all of these are important questions which merit thoughtful answers, their discussion escape the scope of this piece. However, this section aims to serve as a brief guide for researchers interested in the implementation of MSMs and a starting point for further exploration of these topics.

First, it is common to encounter cases where the pseudo-sample is constructed using very extreme weights. This occurs when the treatment and covariate combinations have very few observations. Since the weighting process aims to "level and balance" the different treatment and covariate sequences, then those with few individuals will be compensated with higher weights for its members to "represent" those that we cannot observe. Extreme weights may result in unstable estimators with high variance (Kang and Schafer 2007). To account for this issue, researchers should consider trimming or truncating the weights, as well as assessing

Figure 3.2: Mean bias of predicted probabilities by model

Note: Each corner shows one of the nine possible treatment sequences. The axes show the difference between true ACDE and the estimated ACDE by a given model:
CDE 1 = $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=0})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=0})$
CDE 2 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=0})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=0})$
CDE 3 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=0})$ - $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=0})$
CDE 4 = $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=1})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=1})$
CDE 5 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=1})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=1})$
CDE 6 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=1})$ - $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=1})$
CDE 7 = $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=2})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=2})$
CDE 8 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=2})$ - $\mathrm{P}(Y_{Z^{(0)}=0,Z^{(1)}=2})$
CDE 9 = $\mathrm{P}(Y_{Z^{(0)}=2,Z^{(1)}=2})$ - $\mathrm{P}(Y_{Z^{(0)}=1,Z^{(1)}=2})$

the sensitivity of the estimates to this alternative (mainly by exploring the changes in the distribution of weights). Although these alternatives do not completely eliminate the bias, they help to reduce it (Platt, Delaney and Suissa 2012) and also improve the variance of the estimator. Another strategy is to restrict the analysis to cases with moderate weights. While this will not lead to an unbiased estimate of the ATE or ACDE in the *full* sample, it provides information about these effects among the population exposed to the treatment combinations which in practice may be more realistic to observe (Platt, Delaney and Suissa 2012).

Second, it is important to consider that a consequence of the use of weights, regardless of the method used to calculate them, is that it induces within-subject correlation (by "duplicating" individuals), and therefore the standard error estimates reported by standard programs may be invalid. To account for this issue users should use bootstrap methods when assessing the reliability of the estimates (Hernán, Brumback and Robins 2000). It is crucial that in order for the weights to remain useful, they must be estimated in each bootstrapped sample. This will not only help to improve the estimation of standard errors, but will partially ameliorate concerns related to the inclusion of uncertainty in the estimation of weights. Current applications of MSMs do not include information on the uncertainty of the predicted probabilities used for the derivation of the weights. Further studies should address this issue in order to reach better inferences of the object under analysis.

3.4 Advantages and disadvantages of MSMs

MSMs overcome limitations that other tools like mediation analysis and structural nested mean models have. For example, causal mediation analysis (Imai, Keele and Tingley 2010; Pearl 2001) decomposes the total effect of a treatment on an outcome into direct and indirect effects (Imai et al. 2011). However, when one of the confounders is affected by the baseline

treatment, mediation analysis is not appropriate because its procedure requires modeling the outcome as a function of treatment history and those problematic confounders affected by the treatment. Therefore, by explicitly conditioning on them we induce post-treatment control bias as explained above (Montgomery, Nyhan and Torres Forthcoming). More specifically, this method estimates the values of the mediator (the intermediate treatment stage) based on a model that includes relevant confounders and a baseline treatment. Then, the fitted probabilities for each of the values of the treatment are used to predict the outcome. However, for this second step, the model of the outcome includes all treatment stages and all relevant confounders.

Another alternative for the estimation of the ACDE in dynamic settings is the structural nested mean models (SNMMs) approach (Acharya, Blackwell and Sen 2016; Robins 1997, 1999*b*).[15] SNMMs are a powerful alternative for the estimation of treatment effects especially when the treatments are continuous or comprise a large number of categories (Vansteelandt, Joffe et al. 2014). However, even though SNMMs have the great advantage of working for any type of treatments and confounders, they cannot handle any type of *outcome*. Most SNMMs cannot impose restrictions on the finite support of the outcome (Robins 1999*b*) and are therefore unsuitable for the study of ordinal, multinomial, and count variables.[16] Furthermore, SNMMs are less intuitive and accessible than MSMs and its core concept of "balancing" the sample (Vansteelandt, Joffe et al. 2014). As Acharya, Blackwell and Sen (2016) indicate, "when the treatment and mediator are binary or only take on a few values, nonparametric or semi-parametric approaches exist to estimating the ACDE, reducing the need for parametric models." In summary, MSMs are accessible, straightforward and often more suitable for the estimation of controlled direct effects when the treatment has few

[15]For this purpose, these models decompose the overall treatment effect into components that allow for the identification of "demediated" effects.

[16]In their paper Acharya, Blackwell and Sen (2016) present the implementation of SNMMs for continuous variables. Vansteelandt (2010) extends and elaborates on the application of SNMMs to dichotomous outcomes.

values.

MSMs have multiple strengths and advantages but also some weaknesses. First, even though MSMs can theoretically handle any type of outcome and treatment variables, their use is mainly restricted to categorical or binary treatments. The reason is that a large number of values complicates the fulfillment of the positivity assumption. In cases where the treatment is continuous SNMMs should be favored (Acharya, Blackwell and Sen 2016).

Further, MSMs estimates are sensitive to misspecification of the treatment assignment model. This is due to the reliability of IPTW on the calculation of probabilities of treatment sequences. However, there are alternatives that help to alleviate and diagnose this issue. As reviewed in the previous section, there are multiple methods that might aid to achieve more accurate weights in the presence of multiple covariates (Watkins et al. 2013). Further, Imai and Ratkovic (2015) generalize the covariate balancing propensity score (CPBS) methodology (Imai and Ratkovic 2014) to time-varying treatments and confounder settings such that the covariate balance is improved in each stage. In addition, authors like Blackwell (2013); Robins (1997); VanderWeele (2010) have developed and implemented tools to conduct sensitivity analysis in order to assess the robustness of the estimates of ACDEs in multiple scenarios where the sequential ignorability assumption is violated.

Finally, although IPTW estimators remain unbiased even in cases with small samples, the standard errors tend to be larger than in naïve models. Weights induce higher variance and higher standard errors of the estimates under study (see simulations in Appendix). Depending on each particular case and data, researchers should consider the bias-efficiency trade-off when using MSMs (Westreich et al. 2012).

Chapter 4

MSMs and controlled direct effects in practice: the effect of parents' income on political participation

In this chapter I present an extension of the example outlined above regarding the CDE of income in youth on political participation that is not mediated by income in adulthood. This application illustrates the differences between MSMs and traditional regression models in terms of inferences, and second, it extends the analysis of CDE to non-binary treatments and non-continuous outcomes.

To illustrate the consequences that confounding and post-treatment bias have on results and inferences, I compare the estimates from MSMs to two *naïve* models: the *over control* or saturated model and the *under control* model. A common approach is to include all relevant confounders in a regression regardless of whether these are affected by the treatment—the *over control* model. A less common but still plausible practice is to avoid problematic confounders and limit the analysis to the baseline and intermediate treatments—the *under*

control model.[1]

Wealth is assumed to affect several factors in early stages of life, such as motivation, abilities, skills, and favorable social environments. However, the causal effect of these early economic conditions on political participation that is not mediated by economic status in adulthood is understudied. Beck and Jennings (1982) attempted to estimate this effect through traditional regression methods but, as I note above, this approach leads to biased estimates.[2]

I aim to provide evidence that the CDE of parents' income on participation is positive. That is, if we set income in adulthood to a certain level (for example, by providing subsidies or policies to "level" income), there would *still* be an effect of early economic conditions on political participation. However, the magnitude and reliability of this effect varies depending on the specific type of political activity that a subject pursues. While some activities require actual monetary resources, others are more likely to require skills developed in early stages of life (Verba and Nie 1972; Verba, Nie and Kim 1978). For example, Lipset (1960) finds that middle-life practices contribute to the development of democratic political orientations and these, in turn, are associated with engagement in activities such as rallies or protests. However, other activities, such as donating to a campaign, are more likely to be influenced solely by the availability of resources associated with income at the moment of the event (e.g. money, time, transportation means, context).

The data to test these effects comes from the Youth-Parent Socialization Panel Study (Jennings et al. 2005). This is a panel study in which a sample of students and their parents were interviewed for four waves in 1965, 1973, 1982 and 1997. I use the models below to

[1]This model also includes strictly pre-intermediate treatment and outcome confounders. In the Appendix, I illustrate the pernicious consequences of this practice and the benefits that MSMs imply via a simulation exercise.

[2]Beck and Jennings (1982) use causal path analysis and find that there is a strong direct effect of parents' SES (as measured by school achievement) on their child's political participation in early adulthood. However, they use an overall index of political participation and do not distinguish between activities.

study two different outcomes measured in 1982: attending a political rally, and giving money to a candidate or campaign.[3] For the treatment sequence, I measure parents' income, the first stage, as the family income reported by each student's parents in 1965. Income in adulthood, the second stage, is the family income reported by the student in 1982. The treatment variable is a 4-category variable (based on income quartiles) that ranges from low to high income. The confounders of participation and income included in the model were selected based on findings in the previous literature.[4]

I estimate the ACDE of income in youth on political participation using a stabilized inverse-probability-weighted marginal structural model as described above. In brief, I fit a weighted logistic regression model of the form:

$$\Pr(\text{Political event}_{i,1982} = 1) = \text{logit}(\alpha_0 + \alpha_1 \text{ Income}_{i,1965} + \alpha_2 \text{ Income}_{i,1982} + \alpha_3 \text{ Race} + \alpha_4 \text{ Gender}) \quad (4.1)$$

for the events *attending a rally* and *giving money to a campaign*, where 1 indicates that the respondent engaged in that activity in the period between 1973 and 1982.

I account for potential confounding of time-varying items by fitting the earlier models with stabilized inverse probability weights of the form:

$$\mathcal{W}_t^{-1} = w_{(1965)}^{-1} \times w_{(1982)}^{-1} = \frac{f(Z^{(1965)}|X^{(1965)})}{f(Z^{(1965)})} \times \frac{f(Z^{(1982)}|Z^{(1965)}, X^{(1965)}, X^{(1982)})}{f(Z^{(1982)}|Z^{(1965)})}, \quad (4.2)$$

where $f(\cdot)$ is the inverse of the ordered categorical logistic regression to estimate probabilities.

The predicted probabilities for the numerator and denominator were assigned based on the income category that each panelist reported.[5] For example, the weight of an individual

[3] In 1982 respondents are in the "peak" of their adulthood. Therefore, we can obtain better measurements and reach more generalizable conclusions than at other stages of adulthood.

[4] For the first treatment stage, these covariates include education of both mother and father, and race and level of interest in politics of the head of the household. For the second stage, the confounders are the student's characteristics such as education, political interest, political efficacy and political knowledge as indicators of political skills, motivations and self-confidence. For the full model, I include gender and race of the student as "non-problematic" confounders given that income cannot affect these variables.

[5] For the weight estimation, model and bootstrapping I used my own code in `R`. Existing packages do not handle treatments with multiple categories. The functions are available upon request. The weights were

that belongs to the 3^{rd} quartile of income in adulthood is based on the estimated probabilities of belonging to that particular quartile conditional on different combinations of treatments and covariates as specified in the numerator and denominator of each of the components of Equation 12.

The weights estimated from Equation 12 aim to balance the second stage of the treatment, income in adulthood, across confounders. Figure 4.1 shows that the weights lead to a more balanced sample. This figure illustrates the difference in the standardized coefficients of the confounders on income in adulthood in the original population (left side) and the pseudo-population (right side). The figure shows that while in the original population all covariates significantly predict levels of income in adulthood, in the pseudo-population, almost all of these are no longer significantly associated with the latter. In other words, we successfully "broke" the link between post-treatment confounders and treatment.

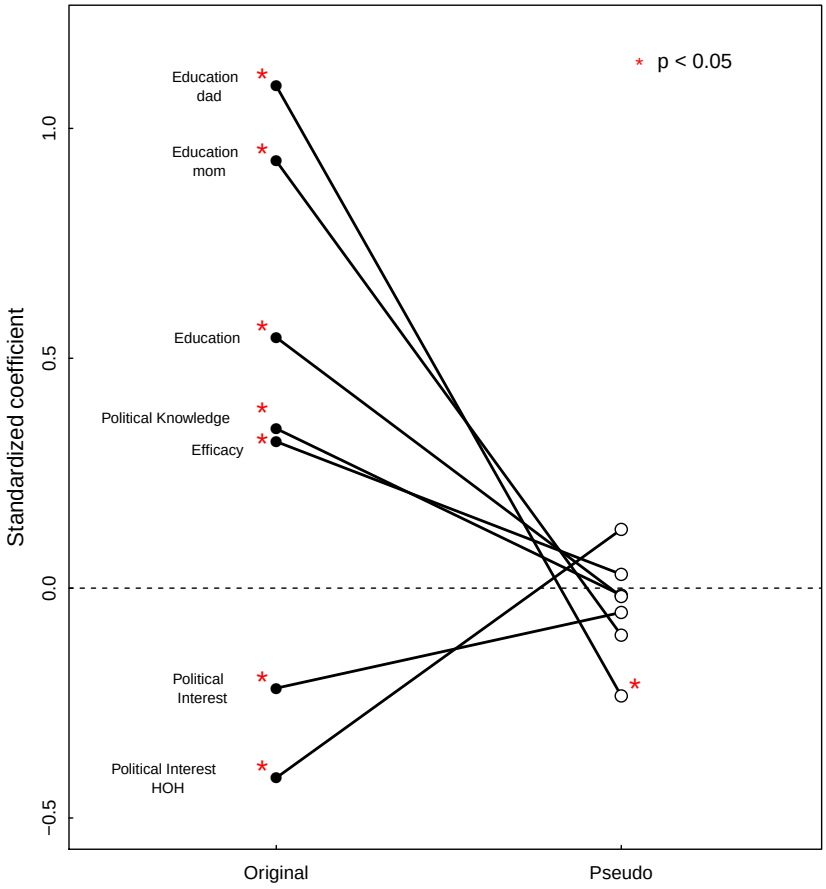
Table 4.1 presents the results for each of the two main outcomes of interest: attending a rally and donating money. For comparison purposes, I implement three modeling strategies: a weighted MSM model, an *over control* model that explicitly controls for all covariates regardless of whether these are post-treatment, and a third *under control* model that excludes relevant confounders.

There are significant differences in the magnitude and reliability of the coefficients. First, the results from the MSM in Column 1 indicate that parents' income has a reliable and positive impact on the propensity of an individual to attend a rally once she is an adult.[6] Being in the third and fourth quartiles of parents' income increases participation in a rally independent of the effect of income in adulthood. However, the results from models 2 and 3 do not support this finding. The over control and under control models indicate that belonging to the fourth quartile of income in adulthood increases the probability of attending a rally, but they do

re-estimated in each of the 500 bootstrapped samples.

[6]The baseline category is the lowest income quartile (Quartile 1), so all interpretations of coefficients are made with respect to this category.

Figure 4.1: Balancing covariates



not provide evidence for a similar effect of parents' income.

Recall that MSMs, assuming that sequential ignorability holds, provide unbiased estimates in contrast to the *naïve* models. That is, no matter the specification, if 1) there is a confounder of treatment and outcome, and 2) at least one confounder is affected by the treatment, then the results from models that over or under control will be incorrect. Furthermore, intuitively and substantively, the results from the MSM are in line with theoretical expectations: activities like a rally are less dependent on economic resources acquired in late stages of life and more likely to be affected by other traits such as group consciousness (Miller et al. 1981) or cross-cutting networks (Mutz 2002) that are influenced by socio-economic con-

Table 4.1: Controlled direct effects of early income on participation

	Attend a rally			Donate money		
	MSM	Over control	Under control	MSM	Over control	Under control
	(1)	(2)	(3)	(4)	(5)	(6)
Parents' income (1965)						
Quartile 2	0.076	−0.020	−0.0005	0.011	−0.037	0.001
	(0.046)	(0.036)	(0.036)	(0.054)	(0.037)	(0.037)
Quartile 3	**0.116**	0.023	0.032	0.054	0.002	0.047
	(0.053)	(0.043)	(0.042)	(0.066)	(0.042)	(0.041)
Quartile 4	**0.169**	0.047	0.055	0.033	0.000	0.077
	(0.086)	(0.055)	(0.049)	(0.081)	(0.053)	(0.048)
Income Adulthood (1982)						
Quartile 2	0.012	0.0002	0.047	−0.002	0.005	0.054
	(0.055)	(0.042)	(0.041)	(0.072)	(0.038)	(0.040)
Quartile 3	−0.051	−0.035	0.027	0.013	0.007	0.077
	(0.062)	(0.047)	(0.051)	(0.075)	(0.045)	(0.047)
Quartile 4	0.072	**0.089**	**0.189**	**0.156**	**0.186**	**0.285**
	(0.068)	(0.043)	(0.043)	(0.076)	(0.045)	(0.044)
N	966	966	966	964	964	964
Post-treatment controls?		✓			✓	

Note: Coefficient estimates for covariates/controls omitted. **Bolded** coefficients reliable at more than 95%. Controls include education of both mother and father, political interest and race of the head of the household, student's education, political efficacy, political interest and knowledge, gender, and race. Regressions include gender and race of the student as strictly pre-treatment covariates. Cut-points and constant terms omitted.

ditions in childhood. The results support the idea that even if all adults manage to close the income gap, there would still be a pervasive effect of inequality on the propensity to participate in rallies. However, the results of the traditional regression models fail to recover this effect.

The results for the "Donate money" outcome are also consistent with theoretical expectations. The effect of parents' income that is not mediated by income in adulthood cannot be distinguished from zero in any of the models considered. However, the effect of income in adulthood is positive and reliable for the fourth income quartile. This suggests that the contemporaneous effect of income in adulthood is more relevant in determining monetary contributions to a candidate or a campaign than any other resources acquired in early stages.

The factual resources that income provides, as well as other determinants such as the network that the professional environment and higher income in adulthood might affect, impact the likelihood of engaging in this activity.

Altogether, the results confirm the hypothesis that for certain activities, there is a positive effect of parents' income on political participation. Although the association of income in adulthood and political participation has been widely supported by many authors, it is important to isolate its effect from that of early economic conditions. Even though for certain activities, such as donating money, the effect of income in adulthood proves to be stronger (probably due to the necessity of resources provided by income in later stages to complete the task), there are other activities such as attending a rally that are more influenced by other traits and characteristics that are highly likely to be developed in (and shaped by) early economic conditions. These effects are accurately captured and estimated through MSMs in contrast to regression techniques that may lead us to substantively different conclusions.

Chapter 5

Concluding remarks

The estimation of direct effects is increasingly receiving attention from scholars in multiple fields. Disentangling causal paths is a strategy that has the potential to improve our understanding of a wide variety of political phenomena. Moreover, the analysis of complex structures, such as those in which there are time-varying treatments and confounders, motivates several research questions in multiple fields that can be answered with the estimation of controlled direct effects. More generally, we can explore the effect of a baseline treatment on an outcome, when we assume that the intermediate treatment stage is set to a particular level. For example, we may be interested in assessing the effect of zoning criteria on political and community engagement that is not mediated by the area's subsequent capital gain, evaluating the effect of welfare support on approval ratings before and after a policy reform, exploring contemporaneous implications of historical variables, or examining similar dynamic relationships.

The analysis of these cases is challenging in methodological terms. The estimation of the average controlled direct effect (ACDE) is complicated when there are time-varying treatments and time-varying confounders affected by the treatment. In dynamic social settings, we have reasons to believe that this is the rule rather than the exception. In this thesis, I

explain the two sources of bias that we are likely to encounter when there are confounders affected by the treatment: first, when these confounders are omitted from a regular regression, the causal effect of an intermediate stage cannot be identified due to confounding bias. Nevertheless, controlling for those confounders may induce bias in the estimation of early treatment stages due to post-treatment control. Under these settings, controlled direct effects cannot be estimated using conventional regression approaches because they do not solve the trade-off between confounding and post-treatment control biases.

In order to solve this bias trade-off, I have used MSMs and IPTW estimators as an alternative for the estimation of the ACDE. Through the calculation of weights that "balance" the marginal distributions of potential outcomes, MSMs account for confounding variables while avoiding post-treatment control bias. I described MSMs' characteristics and presented a detailed description of their implementation especially when dealing with multi-valued treatments. I also illustrated some of the differences in terms of bias between distinct methods for the weights estimation. After the application of this class of models to the analysis of the effects of inequality on political outcomes, I examined the different estimates that we can get from the MSM approach and other common *naïve* models that either under or over control for problematic confounders. More specifically, the results show that the estimates of the effect of parents' income on participation from regular regression techniques differ from those yielded by MSMs. This in turn has an impact on the inferences that we make. I found that there is a positive effect of early income on activities such as attending a rally, that is not mediated by income in adulthood. However, these conclusions are only reached through the use of MSMs, the *naïve* model does not provide enough evidence to sustain these findings and leads to substantively different inferences on this matter. Furthermore, I showed how *naïve* models might lead to conclusions not supported by the literature such as a null effect of economic conditions (both in childhood and adulthood) on turnout.

Despite the wide applicability and accessibility of MSMs, there are issues related to these

models that motivate several further questions. Possibilities for future research include the implementation of inverse probability of treatment and censored weighting estimators in samples as a way of accounting for panel attrition. This would improve the efficiency and accuracy of the estimates by taking into account a problem that is likely to affect the variables under analysis: attrition and non-response. In summary, MSMs are a feasible alternative when dealing with panel/time-series structures and time varying treatments. They offer a straightforward method for the estimation of controlled direct effects, under a small number of assumptions and can be implemented using off-the-shelf software. The application of this method to political questions will certainly lead to a better understanding of the causal associations that exist in the complex systems in which we live.

Bibliography

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings Without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110(3):512–29.

Akee, Randall, William Copeland, E. Jane Costello, John B. Holbein and Emilia Simeonova. 2018. "Family Income and the Intergenerational Transmission of Voting Behavior: Evidence from an Income Intervention." *NBER Working Paper 24770* .

Almond, Gabriel and Sidney Verba. 1989. *The civic culture: political attitudes and democracy in five nations*. Newbury Park, CA: Sage.

Beck, Paul Allen and M Kent Jennings. 1982. "Pathways to participation." *American Political Science Review* 76(1):94–108.

Blackwell, Matthew. 2013. "A framework for dynamic causal inference in political science." *American Journal of Political Science* 57(2):504–520.

Blackwell, Matthew and Adam Glynn. 2014. How to make causal inferences with time-series cross-sectional data. Technical report Working Paper, Harvard University.

Brady, Henry E, Sidney Verba and Kay Lehman Schlozman. 1995. "Beyond SES: A resource model of political participation." *American Political Science Review* 89(2):271–294.

Córdova, Abby. 2009. "Methodological note: Measuring relative wealth using household asset indicators." *Americas Barometer Insights* 6:1–9.

Currie, Janet. 2008. Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. Technical report National Bureau of Economic Research.

Elwert, Felix and Christopher Winship. 2014. "Endogenous selection bias: The problem of conditioning on a collider variable." *Annual Review of Sociology* 40:31–53.

Glynn, Adam N and Kevin M Quinn. 2010. "An introduction to the augmented inverse propensity weighted estimator." *Political Analysis* 18(1):36–56.

Hernán, Miguel Ángel, Babette Brumback and James M Robins. 2000. "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men." *Epidemiology* 11(5):561–570.

Hernán, Miguel Ángel, Babette Brumback and James M Robins. 2001. "Marginal structural models to estimate the joint causal effect of nonrandomized treatments." *Journal of the American Statistical Association* 96(454):440–448.

Imai, Kosuke, Luke Keele and Dustin Tingley. 2010. "A general approach to causal mediation analysis." *Psychological methods* 15(4):309–334.

Imai, Kosuke, Luke Keele, Dustin Tingley and Teppei Yamamoto. 2011. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* 105(4):765–789.

Imai, Kosuke and Marc Ratkovic. 2014. "Covariate balancing propensity score." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 76(1):243–263.

Imai, Kosuke and Marc Ratkovic. 2015. "Robust estimation of inverse probability weights for marginal structural models." *Journal of the American Statistical Association* 110(511):1013–1023.

Jennings, M Kent, Gregory B Markus, Richard G Niemi and Laura Stoker. 2005. "Youth-parent socialization panel study, 1965-1997: four waves combined.".

Kang, Joseph DY and Joseph L Schafer. 2007. "Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data." *Statistical science* pp. 523–539.

Miller, Arthur H, Patricia Gurin, Gerald Gurin and Oksana Malanchuk. 1981. "Group consciousness and political participation." *American Journal of Political Science* 25(3):494–511.

Miratrix, Luke W, Jasjeet S Sekhon and Bin Yu. 2013. "Adjusting treatment effect estimates by post-stratification in randomized experiments." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75(2):369–396.

Montgomery, Jacob M, Brendan Nyhan and Michelle Torres. Forthcoming. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* .

Montgomery, Jacob M. and Santiago Olivella. Forthcoming. "Tree-Based Models for Political Science Data." *American Journal of political Science* .

Moore, Jeffrey C and Edward J Welniak. 2000. "Income measurement error in surveys: A review." *Journal of official statistics* 16(4):331.

Mutz, Diana C. 2002. "The consequences of cross-cutting networks for political participation." *American Journal of Political Science* 46(4):838–855.

Nandi, Arijit, M Maria Glymour, Ichiro Kawachi and Tyler J VanderWeele. 2012. "Using marginal structural models to estimate the direct effect of adverse childhood social conditions on onset of heart disease, diabetes, and stroke." *Epidemiology* 23(2):223–232.

Ojeda, Christopher. Forthcoming. "The Two Income-Participation Gaps." *American Journal of Political Science* .

Pearl, Judea. 2001. Direct and indirect effects. In *Proceedings of the seventeenth conference on uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. pp. 411–420.

Pearl, Judea. 2011. The mediation formula: A guide to the assessment of causal pathways in nonlinear models. Technical report DTIC Document.

Platt, Robert William, Joseph Austin Christopher Delaney and Samy Suissa. 2012. "The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding." *European journal of epidemiology* 27(2):77–83.

Robins, James M. 1997. Causal inference from complex longitudinal data. In *Latent variable modeling and applications to causality*. Springer pp. 69–117.

Robins, James M. 1999*a*. "Association, causation, and marginal structural models." *Synthese* 121(1):151–179.

Robins, James M. 1999*b*. Marginal Structural Models versus Structural Nested Models as tools for causal inference. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*. Vol. 116 New York: Springer-Verlag pp. 95–133.

Robins, James M, Miguel Ángel Hernán and Babette Brumback. 2000. "Marginal structural models and causal inference in epidemiology." *Epidemiology* 11(5):550–560.

Rosenbaum, Paul R. 1984. "The consequences of adjustment for a concomitant variable that has been affected by the treatment." *Journal of the Royal Statistical Society. Series A (General)* 147(5):656–666.

Rubin, Donald B. 1977. "Assignment to Treatment Group on the Basis of a Covariate." *Journal of Educational and Behavioral statistics* 2(1):1–26.

Samii, Cyrus, Laura Paler and Sarah Daly. 2017. "Retrospective Causal Inference with Machine Learning Ensembles: An Application to Anti-Recidivism Policies in Colombia." *Political Analysis* 24(4):434–456.

VanderWeele, Tyler J. 2009. "Marginal structural models for the estimation of direct and indirect effects." *Epidemiology* 20(1):18–26.

VanderWeele, Tyler J. 2010. "Bias formulas for sensitivity analysis for direct and indirect effects." *Epidemiology* 21(4):540–551.

Vansteelandt, Stijn. 2010. "Estimation of controlled direct effects on a dichotomous outcome using logistic structural direct effect models." *Biometrika* pp. 1–14.

Vansteelandt, Stijn, Marshall Joffe et al. 2014. "Structural nested models and G-estimation: The partially realized promise." *Statistical Science* 29(4):707–731.

Verba, Sidney and Norman H Nie. 1972. *Participation in America*. Harper & Row.

Verba, Sidney, Norman H Nie and Jae-on Kim. 1978. *Participation and political equality: A seven-nation comparison*. University of Chicago Press.

Watkins, Stephanie, Michele Jonsson-Funk, M Alan Brookhart, Steven A Rosenberg, T Michael O'shea and Julie Daniels. 2013. "An Empirical Comparison of Tree-Based Methods for Propensity Score Estimation." *Health services research* 48(5):1798–1817.

Westreich, Daniel, Stephen R Cole, Enrique F Schisterman and Robert W Platt. 2012. "A simulation study of finite-sample properties of marginal structural Cox proportional hazards models." *Statistics in medicine* 31(19):2098–2109.

Appendix A

Table A.1: From Table 1: Calculation of weights for each stratum in sample (full-table)

$Z^{(0)}$	$Z^{(1)}$	$X^{(1)}$	$f(Z^{(1)}\|Z^{(0)})$	$f(Z^{(1)}\|Z^{(0)}, X^{(1)})$	$\mathcal{W}(t)^{-1}$	Original-pop N	Pseudo-pop N
0	0	0	0.6	0.706	0.85	12000	10200
0	0	1	0.6	0.462	1.3	6000	7800
0	1	0	0.3	0.235	1.275	4000	5100
0	1	1	0.3	0.385	0.78	5000	3900
0	2	0	0.1	0.059	1.7	1000	1700
0	2	1	0.1	0.154	0.65	2000	1300
1	0	0	0.367	0.467	0.786	7000	5500
1	0	1	0.367	0.267	1.375	4000	5500
1	1	0	0.4	0.333	1.2	5000	6000
1	1	1	0.4	0.467	0.857	7000	6000
1	2	0	0.233	0.2	1.167	3000	3500
1	2	1	0.233	0.267	0.875	4000	3500
2	0	0	0.2	0.4	0.5	2000	1000
2	0	1	0.2	0.1	2	1000	2000
2	1	0	0.333	0.4	0.833	2000	1667
2	1	1	0.333	0.3	1.111	3000	3333
2	2	0	0.467	0.2	2.333	1000	2333
2	2	1	0.467	0.6	0.778	6000	4667

Note: $Z^{(0)}$ is parents' income where 0 is low, 1 is middle and 2 is high. $Z^{(1)}$ is income in adulthood where 0 is low, 1 is middle and 2 is high. $X^{(1)}$ is post-High School education where 0 is no college and 1 is college.

Simulation 1: comparison of bias between weight estimation methods (Section 2.3)

To illustrate the average bias in the estimation of controlled direct effects by model, I simulated 500 datasets, each of size $n = 1,000$, with the following specification:

Baseline covariate: parents' college attendance

$$X_i^{(0)} \sim \mathcal{B}ernoulli(1, 0.5)$$

Baseline treatment: parents' income

$$\begin{aligned}
Z_i^{(0)} &\sim \mathcal{C}ategorical(3, \mathbf{p}_i) \\
\mathbf{p}_i &= (F(\eta_{i1}), F(\eta_{i2}) - F(\eta_{i1}), 1 - F(\eta_{i2})) \\
F(\eta_{ik}) &= \frac{\exp(\theta_k - \eta_{ik})}{1 + \exp(\theta_k - \eta_{ik})} \\
\eta_{ik} &= -2.5 + 1.5X_i^{(0)} + u_i \\
u_i &\sim \mathcal{N}(0, 0.5) \\
(\theta_1, \theta_2) &= (-1.25, 0.45)
\end{aligned}$$

Covariate affected by baseline treatment: subject's college attendance

$$
\begin{aligned}
X_i^{(1)} &\sim \mathcal{B}ernoulli(1, p_i^{\dagger}) \\
p_i^{\dagger} &= \frac{\exp(-2.5 + 0.5X_i^{(0)} + 1.1Z_i^{(0)} + u_i^{\dagger})}{1 + \exp(-2.5 + 0.5X^{(0)} + 1.1Z^{(0)} + u_i^{\dagger})} \\
u_i^{\dagger} &\sim \mathcal{N}(0, 0.35)
\end{aligned}
$$

Second-stage treatment: subject's income

$$
\begin{aligned}
Z_i^{(1)} &\sim \mathcal{C}ategorical(3, \mathbf{p}_i^*) \\
\mathbf{p}_i^* &= (F(\eta_{i1}^*), F(\eta_{i2}^*) - F(\eta_{i1}^*), 1 - F(\eta_{i2}^*)) \\
F(\eta_{ik}^*) &= \frac{\exp(\theta_k^* - \eta_{ik}^*)}{1 + \exp(\theta_k^* - \eta_{ik}^*)} \\
\eta_{ik}^* &= -3.5 + 0.2X_i^{(0)} + 1Z_i^{(0)} + 0.6X_i^{(1)} + u_i^* \\
u_i^* &\sim \mathcal{N}(0, 0.5) \\
(\theta_1^*, \theta_2^*) &= (-1.05, 0.65)
\end{aligned}
$$

Outcome: participation in a rally

$$
\begin{aligned}
\mathbf{Y}_i &\sim \mathcal{B}ernoulli(p_i^{+}) \\
p_i^{+} &= \frac{\exp(\eta_i^{+})}{1 + \exp(\eta_i^{+})} \\
\eta_i^{+} &= -3 + 0.2X_i^{(0)} + 1.5Z_i^{(0)} + 0.4X_i^{(1)} + 0.2Z^{(1)} + u_i^{+} \\
u_i^{+} &\sim \mathcal{N}(0, 0.4)
\end{aligned}
$$

For this exercise, I calculated nine potential outcomes according to the multiple combinations of the baseline and second-stage treatment values. The *true* controlled direct effects of the baseline treatment are calculated for each individual using this framework. The results are presented in Figure 3 in the manuscript.

Simulation 2

The following simulations illustrate the advantages that marginal structural models have over saturated models that control for post-treatment confounders under several conditions.

I conduct three sets of simulations, each changing the value of one of the following parameters while keeping the others constant: number of observations (n), the effect of a covariate on the treatment sequence, $X^{(0)}$, and the effect of a confounder of treatment and outcome affected by the baseline treatment. In each set I generate a set of variables with the structure presented below: a baseline treatment stage with three values $Z^{(0)}$, a binary covariate $X^{(1)}$ affected by the baseline treatment, an intermediate treatment stage affected by both $X^{(0)}$ and $Z^{(0)}$, an outcome Y generated by all of these variables. The third simulation also includes another binary covariate $W^{(1)}$ affected by the baseline treatment. Covariates $X^{(0)}$, $W^{(1)}$ and $X^{(1)}$ confound the relationship between the outcome and the treatment stages.

Baseline covariate

$$X_i^{(0)} \sim \mathcal{B}ernoulli(1, 0.4)$$

Baseline treatment

$$
\begin{aligned}
Z_i^{(0)} &\sim \mathcal{C}ategorical(3, \mathbf{p}_i) \\
\mathbf{p}_i &= (F(\eta_{i1}), F(\eta_{i2}) - F(\eta_{i1}), 1 - F(\eta_{i2})) \\
F(\eta_{ik}) &= \frac{\exp(\theta_k - \eta_{ik})}{1 + \exp(\theta_k - \eta_{ik})} \\
\eta_{ik} &= -2.5 + \beta_1 X_i^{(0)} + u_i \\
u_i &\sim \mathcal{N}(0, 0.5) \\
(\theta_1, \theta_2) &= (-1.25, 0.05)
\end{aligned}
$$

When held constant $\beta_1 = 1$.

Covariates affected by baseline treatment

$$
\begin{aligned}
X_i^{(1)} &\sim \mathcal{B}ernoulli(1, p_i^{\dagger}) \\
p_i^{\dagger} &= \frac{\exp(\eta_i^{\dagger})}{1 + \exp(\eta_i^{\dagger})} \\
\eta_i^{\dagger} &= -2.5 + 0.5 X_i^{(0)} + 1.1 Z_i^{(0)} + u_i^{\dagger} \\
u_i^{\dagger} &\sim \mathcal{N}(0, 0.035)
\end{aligned}
$$

$$
\begin{aligned}
W_i^{(1)} &\sim \mathcal{B}ernoulli(1, p_i^{\ddagger}) \\
p_i^{\ddagger} &= \frac{\exp(\eta_i^{\ddagger})}{1 + \exp(\eta_i^{\ddagger})} \\
\eta_i^{\ddagger} &= -2.5 + 0.5 X_i^{(0)} - 1.5 Z_i^{(0)} + u_i^{\ddagger} \\
u_i^{\ddagger} &\sim \mathcal{N}(0, 0.035)
\end{aligned}
$$

Second-stage treatment

$$
\begin{aligned}
Z_i^{(1)} &\sim \mathcal{C}ategorical(3, \mathbf{p}_i^*) \\
\mathbf{p}_i^* &= (F(\eta_{i1}^*), F(\eta_{i2}^*) - F(\eta_{i1}^*), 1 - F(\eta_{i2}^*)) \\
F(\eta_{ik}^*) &= \frac{\exp(\theta_k^* - \eta_{ik}^*)}{1 + \exp(\theta_k^* - \eta_{ik}^*)} \\
\eta_{ik}^* &= -3.5 + \beta_2 X_i^{(0)} + 1 Z_i^{(0)} + 0.6 X_i^{(1)} + \gamma_1 W_i^{(1)} + u_i^* \\
u_i^* &\sim \mathcal{N}(0, 0.5) \\
(\theta_1^*, \theta_2^*) &= (-1.05, 0.65)
\end{aligned}
$$

When held constant, $\beta_2 = 0.5$ and $\gamma_1 = 0.6$.

Outcome

$$
\begin{aligned}
\mathbf{Y}_i &\sim \mathcal{B}ernoulli(p_i^+) \\
p_i^+ &= \frac{\exp(\eta_i^+)}{1 + \exp(\eta_i^+)} \\
\eta_i^+ &= -3 + 0.2 X_i^{(0)} + \gamma_2 W_i^{(1)} + 1.5 Z_i^{(0)} + 0.4 X_i^{(1)} + 0.2 Z^{(1)} + u_i^+ \\
u_i^+ &\sim \mathcal{N}(0, 0.4)
\end{aligned}
$$

When held constant, $\gamma_2 = 0.3$.

The three sets of simulations have the main objective of comparing and analyzing the biases in the estimation of controlled direct effects in situations where the sequential ignorability and positivity assumptions required by MSMs are violated.

The first simulation varies the number of observations in the simulated datasets (from 60 to 2,000). The objective of this exercise is to explore the sample size properties of the

IPTW estimator while also setting scenarios, such as those with very few observations, where the positivity assumption is likely to be violated. When held constant in the rest of the simulations, $n = 1,000$.

The second simulation varies the effect of a confounder $X^(0)$ on the treatment assignment (both the effect of $X^{(0)}$ on $Z^{(0)}$ [denoted by β_1] and $Z^{(1)}$ [denoted by β_2]). This exercise also helps to illustrate the bias that arises in cases where the treatment assignment is heavily unbalanced and therefore causing 1) certain covariate and treatment histories to be empty and/or 2) to obtain extreme weights for some combinations of such variables.

Finally, the third simulation explores the magnitude and variance of the bias when the researcher omits a confounder of the second stage of the treatment and the outcome. In the simulation, I increase the importance of such confounder by varying the impact of $W^{(1)}$ on the treatment $Z^{(1)}$ (denoted by γ_1) and on the outcome Y (denoted by γ_2).

In order to assess and compute the bias for each case, I generate a set of *potential outcomes* to calculate the "true" controlled direct effects of $Z^{(0)}$ on Y. There are nine CDEs which I present in Table A.2. Then, for each set of simulations I estimated the CDEs based on the *observed outcomes* using two modeling strategies: a marginal structural model (MSM) which implies a weighted regression of the outcome on the two treatments using weights estimated through IPTW, and a saturated model which includes confounders affected by the treatment. For illustrative purposes, weights were estimated using categorical logistic regressions.

For each of the varying values of the parameters of interest, I simulate 400 datasets. After collecting the relevant estimates of ACDEs from a MSM and a saturated model in each dataset, I take the difference between such estimates and the true controlled direct effects. These values represent a measure of bias. Figure A.1 below shows three panels with the distributions of bias when estimating CDE number 1 using either a MSM or a saturated model under different conditions. In each panel, the y-axis indicates the magnitude of the

Table A.2: Simulated controlled direct effects

CDE	$Y_{Z^{(0)}=a,Z^{(1)}=b} - Y_{Z^{(0)}=a',Z^{(1)}=b}$
1	$Y_{10} - Y_{00}$
2	$Y_{20} - Y_{00}$
3	$Y_{20} - Y_{10}$
4	$Y_{11} - Y_{01}$
5	$Y_{21} - Y_{01}$
6	$Y_{21} - Y_{11}$
7	$Y_{12} - Y_{02}$
8	$Y_{22} - Y_{02}$
9	$Y_{22} - Y_{12}$

bias in the estimation of CDE 1, while the upper and lower x-axes show the different values that the parameter of interests take in each set of simulated datasets. The lines indicate the mean bias: red and dashed for the saturated model, and bold and black for the MSM. The gray areas indicate the 5^{th} and 95^{th} percentiles in the distribution of bias.

The results for the first simulation varying the sample size show that although MSMs slightly overestimate the real CDE 1 in small samples (with around 60 to 100 observations), the bias quickly converges to 0 and stays steady for large sample sizes. However, the estimate of ACDE 1 using the saturated model remains biased even when sample size increases. Controlling for the covariate affected by the baseline treatment originates this bias. Although the average bias is close to 0 when implementing a MSM, it is important to consider that its variance is 1) higher than the variance of bias from the saturated model, and 2) decreases at the same time as sample size increases. In general, saturated models perform better in terms of standard errors. A result that does not come as a surprise given the weighting process involved in the estimation of ACDEs in a MSM framework.

The second simulation increases the importance of one of the pre-treatment covariates affecting the assignment of both stages of treatment. Increasing the effect of $X^{(0)}$ on $Z^{(0)}$ and $Z^{(1)}$ yields the following results. Under this setting, we find that, as expected, the estimates of ACDE 1 using MSMs start unbiased when the effect of the covariate is 0 or

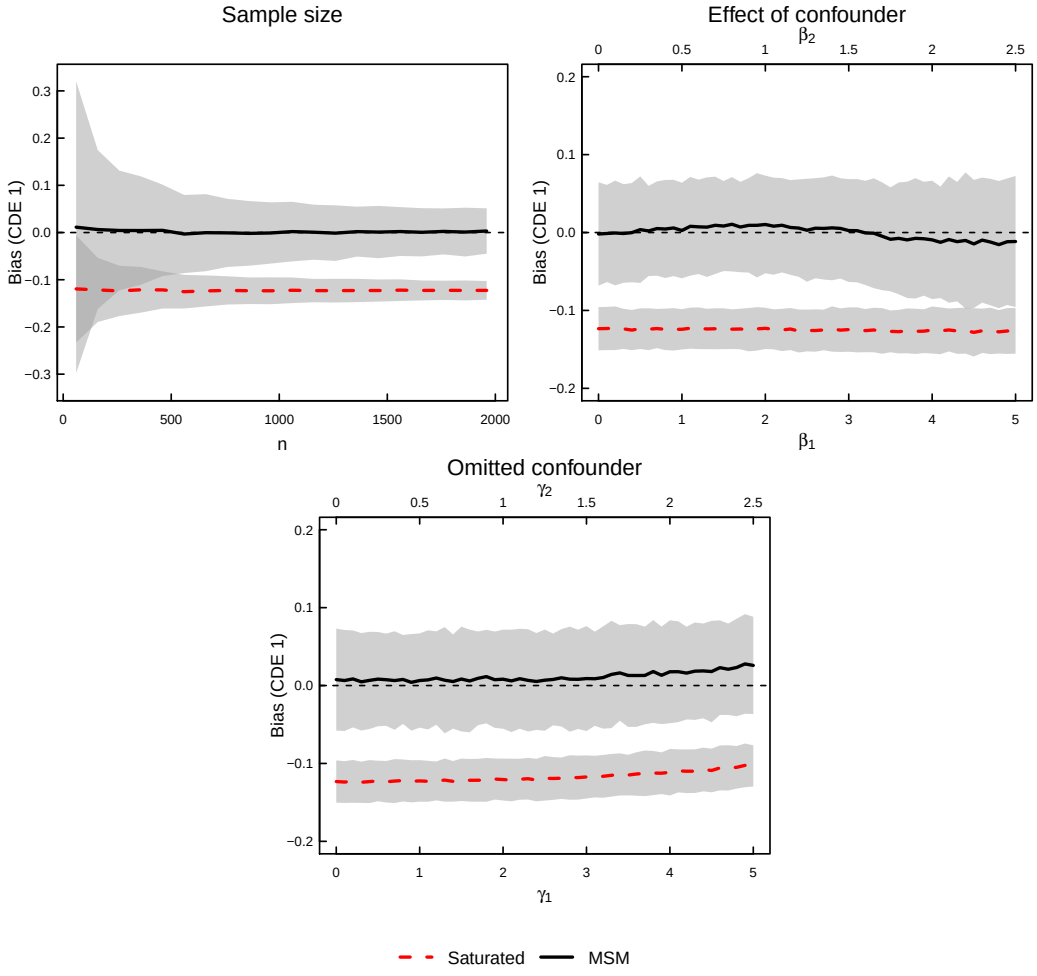
about 0.5 (which once plugged in a the link function it represents a substantive effect in terms of probabilities). When the effects increase, we observe that the average bias departs from zero in both negative and positive directions. This bias, however, is still smaller than the almost constant and negative bias in the estimates that a saturated model yield. Further, we observe an increase in the variance of the bias distribution of MSMs as the effect of $X^{(0)}$ becomes more important. And even in the case where the effect is zero, the variance of the MSM bias is significantly larger than the one for the saturated model.

Finally, the third simulation shows the distributions of bias when a covariate affected by the baseline treatment gets stronger AND when it is omitted from both the MSM and saturated models. In this case, we observe that the estimator is nearly unbiased when this effect is zero, but has a positive trend departing from zero. However, this bias is consistently lower than the one yielded by the saturated model who shows a decreasing trend in terms of bias. While it might appear counterintuitive, this trend can be explained by the "accumulation" of different biases and the bias trade-off that was explained in Section 1 of this text: while ignoring an increasingly strong confounder can have pernicious consequences as the bold line in this simulation shows, including it may also be problematic. In this cases, it seems that the bias generated by post-treatment control is higher than the confounding bias, and therefore we observe a still biased but improving trend in the results.

The main conclusions that we derive from this exercise is that 1) MSMs perform significantly better than saturated models in terms of bias, but 2) the variance of the bias of MSMs suggests a less efficient estimator. This is consistent with the results found by Westreich et al. 2012 in which they also conduct a set of simulations to compare bias, standard errors and mean squared errors. The evaluation of whether the increased in variance that comes from weighting is outweighed by the reduction in bias that MSMs offer heavily depends on the particular characteristics of the study: the distribution of treatments, the effect of the confounders, etc. For example, if the post-treatment confounder has a very small effect on

the treatment sequence and outcome, then a small bias is preferred to large variances. However, the simulations above, conducted under different settings, suggest that the increased variance associated with weighting versus over-adjusting is not too costly. As the graphs show, the MSMs provide a much better coverage of the real estimates than saturated models even at the tail of the bias distribution.

Figure A.1: Distribution of bias: MSMs vs saturated models



Note: Based on 400 simulated datasets per value.

Application

Data description

The sample framework of the Youth-Parent Socialization Panel is composed of senior High School students in 1965. For this wave, the data comes from a nationally representative sample of 1,669 students distributed across 97 public and nonpublic schools selected with probability proportional to size. In the 1965 wave the parents of the students were also interviewed. For the majority of the students, either one or both parents were interviewed. However, for a small number of cases, no parent was interviewed. For the 1973, 1982 and 1997 waves, students were recontacted and resurveyed. Although most of the surveys were completed face-to-face, a number of them in the follow-up waves were completed through mail interviews and computer-assisted telephone interviews (CATI).

The confounders of political participation and income in youth are employment status of the father, his occupation, education levels and place of birth, education of the mother, foreign status and race of the parents, political interest and political efficacy of the father. These measures were collected from the parents in 1965. Only in those cases were there was no information available either from the mother or the father, I use the student's answers to those questions. For the confounders of the treatment and political outcome at $t \neq 0$, I include occupation of the student, education level, labor force status, political interest, level of confidence, political efficacy, gender and race. The last two are the only ones that do not vary over time. All of these variables were collected for each of the waves in the study.

Wording

Outcome variables

- Attend a rally

  - Question: Have you gone to any political meetings, rallies, dinners, or other things like that since 1973?
  - Answers: Yes, No

- Donate money
  - Question: Have you given any money or bought any stickers to help a particular party, candidate, or group pay campaign expenses since 1973?
  - Answer: Yes, No

Treatment variables

- Parents' income: Quartiles based on the categories answered by student's parents in 1965
  - Question: About what do you think your total income will be this year for yourself and your immediate family?

- Income in adulthood: Quartiles based on the categories answered by student in 1973
  - Question: Please look at this page and tell me the letter of the income group that includes the income of all members of your family living here in 1981 before taxes. This figure should include salaries, wages, pensions, dividends, interest, and all other income. If uncertain: what would be your best guess?