How to Make Causal Inferences with Time-Series Cross-Sectional Data*

Matthew Blackwell†
Adam Glynn‡

May 19, 2014
Word Count: 7558

Abstract

Repeated measurements of the same countries, people, or groups over time form the foundation of many fields of quantitative political science. These measurements, called time-series cross-sectional (TSCS) data, can help researchers answer a variety of causal questions. Repeated measurements, however, can also lead to confusion about what causal question scholars are answering and what methods, data, and assumptions they need to do so. In this paper, we clarify these matters and demonstrate that a weighting approach to causal inference can estimate a wide variety of causal quantities of interest. By comparison, approaches such as regression or matching struggle to estimate the effect of histories. Using data on the relationship between trade openness and tax rates, we demonstrate how our weighting technique can overturn results that appear robust.

*We are grateful to Neal Beck, Jake Bowers, Patrick Brandt, Simo Goshev, and Cyrus Samii for helpful advice and feedback. Any remaining errors are our own.

†Department of Political Science, University of Rochester, Harkness Hall, Rochester, NY 14627. web: http://www.mattblackwell.org email: m.blackwell@rochester.edu

‡Department of Government, Harvard University, 1737 Cambridge St., Cambridge, MA 02138 email: aglynn@gov.harvard.edu

1 Introduction

Repeated measurements of the same countries, people, or groups at several points in time form the basis of time-series cross-sectional (TSCS) data. Large swaths of political science collect, use, and even consider the methodological implications of such data. But many fail to realize that TSCS data give researchers the power to ask a richer set of causal questions than data with a single measurement for each unit. We can move past the narrowest contemporaneous questions—what are the effects of a single event—and instead ask how the *history* of a process affects our political world. Furthermore, TSCS data allow researchers to draw on a larger pool of information when estimating these causal effects. This variety of options and information, however, can lead to confusion regarding what causal questions we are answering, what methods we need to do so, and what assumptions justify their use.

Beyond its scientific significance, the choice of contemporaneous versus historical causal questions has consequences for our methodological approach to TSCS data. Regression and matching are popular and valuable tools for estimating the effects of a single action, but they fail for the effects of a sequence of events. In this paper, we demonstrate that a weighting approach to causal inference unifies these questions—contemporaneous and historical—into a single framework. Under assumptions we discuss below, this method, called inverse probability of treatment weighting or IPTW, creates a reweighted dataset in which the historical sequence of events is a sequentially randomized experiment. The lack of confounding in the reweighted data makes analysis straightforward—it's just a randomized experiment. Unfortunately, it's a complicated randomized experiment.

As the complexity of causal questions expands, parametric models become increasingly necessary for two reasons. First, the weights themselves typically require models. Second, nonparametric estimation in the reweighted data with simple sample means is unstable. With a large number of time periods, the number of units that follow any particular sequence shrinks. Thus, historical analyses often require modeling. We discuss the array of modeling choices that accompany these analyses and give guidance on how to select among these choices.

To connect empirical data to causal quantities, our approach, like all of causal inference, requires assumptions. In this paper, we consider the sequential ignora-

bility assumption, which is the TSCS analogue of selection on the observables. The statistical literature has shown that, even under this strong assumption, regression and matching methods cannot recover the cumulative effects of treatment over time (Robins, Hernán, and Brumback, 2000; Blackwell 2013). We illustrate this point by replicating the results from Swank and Steinmo (2002) on the relationship between trade openness and tax rates in advanced democracies. Naively analyzing the cumulative effect of trade openness leads to the conclusion that trade openness has no effect on labor tax rates. However, when properly adjusted for dynamic confounding with inverse probability of treatment weighting, the effect of trade becomes positive and is both statistically and substantively significant.

This paper proceeds as follows. Section 2 clarifies the causal quantities of interest available with TSCS data. In Section 3 we describe the assumptions necessary to identify these causal effects. Section 4 discusses the estimation of these quantities for a given time period, and in Section 5 we highlight the modeling choices necessary when we generalize to multiple time periods. We present the replication of Swank and Steinmo (2002) in Section 6. Finally, Section 7 concludes with thoughts on both the limitations of the weighting approach and avenues for future research.

2 Causal quantities of interest in TSCS data

Repeated measurements greatly expand the range of causal questions available to researchers. At their most basic, TSCS data consists of a treatment, an outcome, and some covariates measured for the same units at various points in time. By treatment, we mean the main explanatory variable of interest—the cause of the main effect of interest. In cross-sectional data with a binary treatment, there are a limited number of counterfactual comparisons to make. Imagine, for instance, a political economy dataset of countries with economic policies as outcomes and internationalization of trade as a binary treatment. With one time period, only one comparison exists: a country has either an open or a closed trade regime. As we gather data on these countries over time, more possibilities arise. How does the history of trade openness in these countries affect tax or budget outcomes? Does their trade regime *today* only affect their policies today or does the recent history

matter as well? The variation over time provides the opportunity and the challenge of answering these more complex questions.

To fix ideas, let A_{it} be the treatment or independent variable of interest for unit i in time period t. For simplicity, we focus on the case of a binary treatment so that $A_{it} = 1$ if the unit is treated in period t and $A_{it} = 0$ if the unit is untreated in period t. We collect all of the treatments for a given unit into a *treatment history*, $\underline{A}_i = (A_{i1}, \ldots, A_{iT})$, where T is the number of time periods in the study. For example, we might have an *always treated* unit with history $(1, 1, \ldots, 1)$ or a *never treated* unit with history $(0, 0, \ldots, 0)$ or any combination of these. In addition, we define $\underline{A}_{it} = (A_{i1}, \ldots, A_{it})$ to be the partial treatment history up through time t. We define X_{it}, $\underline{X}_{it}$, and $\underline{x}_t$ similarly for a set of time-varying covariates that are causally prior to the treatment at time t.

The goal is to estimate causal effects of the treatment on an outcome, Y_{it}, that also varies over time. We take a counterfactual approach (Rubin, 1978) and define potential outcomes for each time period, $Y_{it}(\underline{a}_t)$, where $\underline{a}_t$ is a representative treatment history up through time t.[1] This potential outcome represents the value that the outcome would take in period t if country i had followed history $\underline{a}_t$. Obviously, for any country in any time period, we only observe one of these potential outcomes since a country cannot follow multiple histories at the same time. To connect the potential outcomes to the observed outcomes, we make the standard *consistency asssumption*. Namely, we assume that the observed outcome and the potential outcome are the same for the observed history: $Y_{it} = Y_{it}(\underline{a}_t)$ when $\underline{A}_{it} = \underline{a}_t$.

With these potential outcomes in hand, we can define the causal quantities of interest available with TSCS data.[2] The most basic quantity is the average treatment history effect, or ATHE:

$$\tau(\underline{a}_t, \underline{a}'_t) = E[Y_{it}(\underline{a}_t) - Y_{it}(\underline{a}'_t)]. \tag{1}$$

[1] The definition of potential outcomes in this manner requires the Stable Unit Treatment Value Assumption (SUTVA), (Rubin, 1978). This assumption is questionable for the many comparative politics and international relations applications, but we avoid discussing this complication in this paper in order to focus on the issues regarding TSCS data.

[2] For each of the quantities we present here, there are parallel estimands that condition on baseline (that is, time-invariant) covariates.

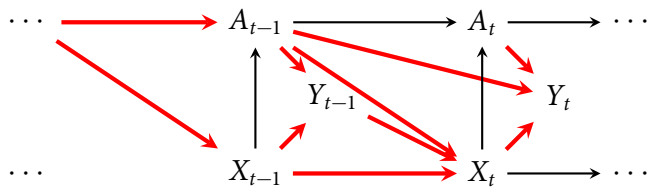

Figure 1: Average Treatment History Effect at Time t

This quantity is the average difference between the world where all units had history $\underline{a}_t$ and the world where all units had history $\underline{a}'_t$. For example, we might be interested in the effect of a country having always traded openly versus a country always having a closed economy. Thus, the ATHE considers the effect of treatment at time t, but also the effect of all lagged values of the treatment as well. A graphical depiction of an ATHE is presented in Figure 1, where the red arrows correspond to components of the effect. These arrows represent all of the effects of A_t, A_{t-1}, A_{t-2}, etc. that end up at Y_t. Note that many of these effects flow through the time-varying covariates, X_t. This point complicates the estimation of ATHEs and we return to it below.

While the ATHE is the most basic effect with TSCS data, it allows a dynamic complexity that makes it quite flexible. It is clear from the definition that there are, in fact, many different ATHEs: one for each pair of treatment histories. As the length of time under study grows, so does the number of possible comparisons. In fact, there are 2^t different values of the ATHE for the outcome in period t. This large number of comparisons allows for a host of causal questions: does the stability of trade openness over time matter for the impact of trade internationalization on economic policies? Is there a cumulative impact of trade openness or is it only the current institutions that matter?

We can define other causal quantities beyond the general ATHE. For instance, one specific class of treatment history effects is the *blip effect* for two histories that agree up to time t:

$$\tau_b(\underline{a}_{t-1}) = E[Y_{it}(\underline{a}^1_t) - Y_{it}(\underline{a}^0_t)|\underline{A}_{i,t-1} = \underline{a}_{t-1}], \tag{2}$$

where $\underline{a}^1_t = (\underline{a}_{t-1}, 1)$ and $\underline{a}^0_t = (\underline{a}_{t-1}, 0)$, so that τ_b represents the effect of a treat-

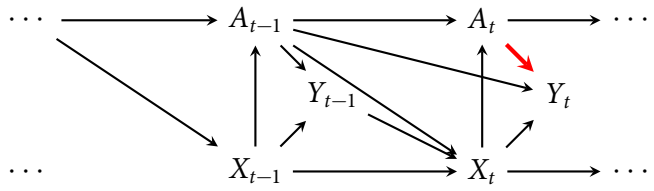

Figure 2: Contemporaneous Effect at Time t

ment "blip" in the last period. For example, this might be the effect of a country opening its trade to international markets at time t after being closed for their entire history. This is obviously a special case of the more general ATHE, where we restrict the two histories to agree up to time t. We refer to the set of all possible treatment histories at $t-1$ as $\underline{\mathcal{A}}_{t-1}$, and every treatment history in this set has its own blip effect for time t. We can average across these blip effects to define the *contemporaneous effect of treatment* (CET) in period t:

$$\begin{aligned}
\tau_t &= \sum_{\underline{m} \in \underline{\mathcal{A}}_{t-1}} E[Y_{it}(\underline{m}, 1) - Y_{it}(\underline{m}, 0) | \underline{A}_{i,t-1} = \underline{m}] \Pr[\underline{A}_{i,t-1} = \underline{m}], \\
&= \sum_{\underline{m} \in \underline{\mathcal{A}}_{t-1}} E[Y_{it}(1) - Y_{it}(0) | \underline{A}_{i,t-1} = \underline{m}] \Pr[\underline{A}_{i,t-1} = \underline{m}], \qquad (3) \\
&= E[Y_{it}(1) - Y_{it}(0)],
\end{aligned}$$

where $Y_{it}(1)$ is the potential outcome under treatment in time t and the second line holds due to the consistency assumption. Here we have switched from potential outcomes that depend on the entire history to potential outcomes that only depend on time t. The CET reflects the effect of treatment in period t on the outcome in period t, averaging across all of the treatment histories. Thus, it would be the expected effect of switching a random country from a closed to an open trade regime in period t. A graphical depiction of a CET is presented in Figure 2, where the red arrow corresponds to component of the effect. Clearly, this quantity narrows the scope of the effect compared to the ATHE. It is common to assume that this effect is constant over time so that $\tau_t = \tau$, but we could alternatively attempt to estimate the average of this effect over time.

It is crucial to distinguish between these different estimands because the var-

ious approaches to causal inference in TSCS data will identify some of these and not others. In general, we will see that blip effects and, thus, the CET will be easier to estimate because they require no fundamental changes to current TSCS estimation. Furthermore, some quantities may be more or less useful for testing theories in political science. In particular, it is important to distinguish whether the ATHE or the CET more directly addresses a particular theory.

3 Causal assumptions in TSCS data

One approach to causal inference in TSCS data relies on a "selection on the observables" assumption, similar to those commonly invoked for cross-sectional data. Beginning with Robins (1986), scholars in epidemiology have extended the usual cross-sectional causal inference framework to handle treatments that can vary over time. These approaches rely on *sequential ignorability*, which is an assumption that weakens the usual cross-sectional ignorability assumption to allow for time dependence. At its core, sequential ignorability describes the relationship between the treatment history and a set of time-varying confounders, X_{it}, and their history, $\underline{X}_{it}$. It allows for feedback between these histories so that the covariates can affect and be affected by the treatment. The assumption states that, conditional on the covariate and treatment histories up to time t, the treatment at time t is independent of the potential outcomes at time t:

Assumption 1 (Sequential Ignorability). *For every action sequences $\underline{a}_t$, covariate history $\underline{X}_{it}$, and time t, if $\underline{A}_{i,t-1} = \underline{a}_{t-1}$, then $Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1} = \underline{a}_{t-1}$.*

This assumption is weaker than the so-called *strict ignorability* assumption, which requires the entire treatment history to be independent of the potential outcomes only conditional on baseline (or cross-sectional) variables. Strict ignorability rules out the possibility of feedback between the time-varying covariates and the treatment. For example, this implies that tax policies and any other time-varying covariates at time t have no effect on trade openness in the future. For these reasons, the strict ignorability assumption is typically unsuitable for use in TSCS applications in political science.

Sequential ignorability, on the other hand, allows for feedback between the treatment status and the time-varying covariates, including the outcome. For instance, sequential ignorability allows for the trade liberalization of a country to impact future economic policy and for policies to affect future trade openness. Thus, in this dynamic case, treatments can affect the covariates and so the covariates also have potential responses: $X_{it}(\underline{a}_{t-1})$. This dynamic feedback is what complicates the estimation of ATHEs. Because the treatment can affect these time-varying confounders, the total effect of treatment becomes the amalgam of effects we see in Figure 1. We must consider not only the direct effect of the treatment history on the outcome, but also its indirect effects through the covariates. For example, trade openness might affect the effective tax rate directly through its effect on capital mobility but also indirectly through its effect on the overall levels of trade.

An important type of time-varying covariate in TSCS data is the lagged dependent variable, or LDV. Usually, scholars worry whether or not the LDV has an effect on the current value of the dependent variable and, if so, how to model that relationship. But if we view the LDV as potentially a member of $\underline{X}_{it}$, then sequential ignorability requires us to know if the LDV has an effect on the treatment history as well. For instance, it may be the case that while trade openness has a strong effect on effect tax rates, these same tax rates might have an effect on future trade openness. If this type of feedback exists, then a lagged dependent variable must be in the conditioning set $\underline{X}_{it}$ and strict ignorability must be violated. This structure is common in TSCS data and implies that sequential ignorability may be the weakest possible assumption for many applications.

4 How to estimate causal effects for a single time period

4.1 Contemporaneous effects

To highlight the relevant issues with the identification and estimation of the above causal quantities, we first investigate the effects for a specific time period—namely, the effect on the outcome in the final period, $Y_i = Y_{iT}$. This choice is for exposition

and we will generalize it in the next section. Under sequential ignorability, the last period of treatment, a_T, is exactly the same as a typical single-shot treatment, with the past history of treatment and the covariate history as the usual confounders. And it is clear that sequential ignorability simplifies to the more usual conditional ignorability assumption if we evaluate it at time period T. Thus, under sequential ignorability, the blip effect of the final period, $\tau_b(\underline{a}_{T-1})$, and the CET for that period, τ_T, are identified by the standard, single-shot causal inference framework. Of course there is nothing special about the final time period; we could repeat the same analysis for the effect of A_t on any Y_t, conditional on the covariate and treatment history.

The above discussion shows that the CET, τ_T, is just the average treatment effect if we regard a_T, the treatment in the final period, as the only treatment of interest. Thus, we can use standard approaches to estimate τ_T. One could use a nonparametric estimator such as a matching analysis on both the treatment history, $\underline{A}_{T-1}$, and the covariate history, $\underline{X}_{T-1}$ (Ho et al. 2007). An alternative to matching would be to run a regression of the final treatment on the final outcome, conditional on the past:

$$Y_i = \beta_0 + \beta_1 A_{iT} + \underline{A}'_{i,T-1}\beta_2 + \underline{X}'_{i,T-1}\beta_3 + \varepsilon_{iT}, \tag{4}$$

where each unit contributes only their last observation to the regression. Note that the coefficient on the final treatment, β_1, in this regression will not, in general, equal the CET, τ_T, or even the blip effect, $\tau_b(\underline{a}_{T-1})$, unless these effects are constant across units due to what we call the *regression weighting problem* or RWP. The RWP is that the regression coefficient averages over a different distribution of the covariates than the average treatment effect. Rather, coefficients are averages weighted by the a unit's treatment variance conditional on the covariates, while the ATE averages over the marginal distribution of the covariates. Unless there are constant treatment effects across units, these two weighting schemes will produce distinct estimates. This is a standard result that applies to regression estimates of the average treatment effect as well (see, for example Angrist and Krueger 1999; Aronow and Samii 2013; Angrist and Pischke 2008). It is possible to use these regression estimators to estimate the CET, though, just as regression estimators can estimate the ATE in the single-shot framework (Imbens 2004). This approach estimates the CET by averaging the predicted effect estimates over the empirical distribution of

the covariates ($\underline{A}_{i,T-1}$ and $\underline{X}_{i,T-1}$ in this case) instead of simply interpreting β_1 as an estimate of the CET.

4.2 Treatment history effects and IPTW

When we move from contemporaneous effects to cumulative effects, the standard regression and matching techniques from the last section break down. This is because the conditioning on the covariate history, $\underline{X}_{i,T-1}$, in (4) induces post-treatment bias for the effect of the lagged treatment values. Intuitively, this conditioning blocks the indirect causal pathways from treatment history to outcome, such as $A_{i,t-1} \rightarrow X_{it} \rightarrow Y_{it}$. Thus, using a regression or matching model from the last section and interpreting the coefficients on lagged values of the treatment as causal will be very misleading. But if we remove the time-varying covariates to avoid the post-treatment bias, we will surely induce omitted variable bias because the time-varying covariates are important confounders. In general, methods that condition on time-varying covariates such as regression and matching will be inappropriate for estimating the effect of the treatment history on the outcome.[3] Note that this bias is present even if there are constant treatment effects across units so that the regression weighting problem is not an issue. Furthermore, not even fixed-effect models, which allow for linear between-country heterogeneity, alleviate these issues of time-varying confounders (Sobel 2012).

There are several methods for estimating ATHEs, though they are quite rare in political science. The easiest to implement of these is to write a model for the marginal mean of the potential outcomes, called a marginal structural model or MSM (Robins, Hernán, and Brumback, 2000; Blackwell 2013). Again, focusing on the effect on the final outcome, the MSM would be

$$E[Y_i(\underline{a})] = g(\underline{a}; \beta), \tag{5}$$

where the function g operates similarly to a link function in a generalized linear model.[4] For instance, we might take g to be linear for a continuous outcome and

[3] This includes lags of the treatment, but also any function of the lags, such as the cumulative sum of treatment or the number periods since the last treated period.

[4] These marginal structural models are similar in spirit to *transfer functions* the context of pure time-series data (Box, Jenkins, and Reinsel 2013).

depend only on an additive combination of the treatment for the current period and the first two lags,

$$g(\underline{a}; \beta) = \beta_0 + \beta_1 a_T + \beta_2 a_{T-1} + \beta_3 a_{T-2}, \tag{6}$$

or we might take g to have a logistic form for a binary outcome,

$$g(\underline{a}; \beta) = \frac{\exp(\beta_0 + \beta_1 a_T + \beta_2 a_{T-1} + \beta_3 a_{T-2})}{1 + \exp(\beta_0 + \beta_1 a_T + \beta_2 a_{T-1} + \beta_3 a_{T-2})}. \tag{7}$$

In both of these cases, we have made restrictions on how the history of treatment affects the outcome. In particular, treatments more than 2 periods before the final outcome are assumed to have no impact on that outcome. There are other ways to map the treatment history to the outcome, such as the cumulative number of treated periods, $\text{sum}(\underline{a}) = \sum_{t=1}^{T} a_{it}$. This allows for the entire history of treatment to affect the outcome in a structured, low-dimensional way. Under any of these models, an ATHE becomes:

$$\tau(\underline{a}, \underline{a}') = g(\underline{a}; \beta) - g(\underline{a}'; \beta). \tag{8}$$

Of course, the choice of the MSM will place restrictions on the ATHEs that we can estimate. A MSM that is a function of only the cumulative treatment, for instance, implies that $\tau(\underline{a}, \underline{a}') = 0$ if $\underline{a}$ and $\underline{a}'$ have the same number of treated periods, even if their sequence differs.

These MSMs lack any reference to time-varying covariates, $\underline{X}_i$. Thus, if one simply estimates these models with observed data, there will be omitted variable bias in the estimated effects. Fortunately, the causal parameters of these models are estimable using a extension of the propensity score weighting approach (Robins, Hernán, and Brumback, 2000; Blackwell 2013). In this MSM approach, we adjust for time-varying covariates using the propensity score weights, not the outcome model itself because, as described above, including such covariates in that model induces post-treatment bias. The weighting removes imbalances on the time-varying covariates across values of the treatments, so that omitting these variables in the reweighted data produces no omitted variable bias. This approach works because, similar to nonparametric matching, weighting ensures that the there is balance in the time-varying covariates across different treatment histories.

Of course, this inverse probability of treatment weighting (IPTW) approach to estimating marginal structural models depends on a number of assumptions, which may be quite strong in some applications. First, sequential ignorability must hold for an observed set of covariates, $\underline{X}_i$. Second, we must assume that *positivity* holds, here defined to mean that

$$0 < \Pr[A_{it} = 1 | \underline{X}_{it} = \underline{x}_t, \underline{A}_{i,t-1} = \underline{a}_{t-1}] < 1 \quad \forall t, \underline{x}_t, \underline{a}_{t-1}, \tag{9}$$

so that it is possible for units to receive treatment at every time period and every possible combination of covariate and treatment histories. This assumption is similar to the common support and overlap conditions in the matching literature. Third, we assume that we have a consistent model for the probability of treatment, conditional on the past:

$$\widehat{\Pr}[A_{it} = 1 | \underline{X}_{it}, \underline{A}_{i,t-1}; \hat{\alpha}_N] \rightarrow_p \Pr[A_{it} = 1 | \underline{X}_{it}, \underline{A}_{i,t-1}]. \tag{10}$$

Here $\hat{\alpha}_N$ is an estimator for the coefficients of a model for the probability of A_{it} conditional on the covariate and treatment histories. This might be simply a pooled logit model, a generalized additive model with a flexible functional form, a boosted regression (McCaffrey, Ridgeway, and Morral 2004), or a covariate-balancing propensity score (CBPS) model (Imai and Ratkovic 2013). To establish consistency of the estimator, we need a model that is correct in the sense that its predicted values converge to the true propensity scores. In spite of this requirement, some methods for propensity score estimation such as CBPS have good finite-sample properties in the face of model misspecification (Imai and Ratkovic 2013).

We use these predicted probabilities to construct weights for each unit-period:

$$\widehat{SW}_i = \prod_{t=1}^{T} \frac{\widehat{\Pr}[A_{it} | \underline{A}_{i,t-1}; \hat{\gamma}]}{\widehat{\Pr}[A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1}; \hat{\alpha}]}. \tag{11}$$

The denominator of each term in the product is the predicted probability of observing unit i's observed treatment status in time t (A_{it}), conditional on that unit's observed treatment and covariate histories. When we multiply this over time, it is the probability of seeing this unit's treatment history conditional on the time-varying covariates. This feature of the IPTW—weighting by the inverse of the probability

of the observed treatment—is what inspires its name. The numerators here stabilize the weights to make sure they are not too variable, which can lead to poor finite sample performance. The numerator is of the same form as the denominator, but with time-varying covariates omitted from the propensity score model. Note that we must build up these weights over time even though we are focusing on the ATHEs for the last outcome.

Under these assumptions, the expectation of Y_i conditional on $\underline{A}_i$ in the reweighted data is equal to the MSM:

$$E_{SW}[Y_i|\underline{A}_i = \underline{a}, X_{i0}] = E[Y_i(\underline{a})|X_{i0}]. \tag{12}$$

Here $E_{SW}[\cdot]$ is the expectation in the reweighted data and X_{i0} are baseline covariates that don't vary over time. This implies that we can estimate ATHEs by simply running a weighted least squares regression of the outcome on the treatment history and any baseline covariates with $\widehat{SW}_i$ as the weights. The coefficients on the components of $\underline{A}_i$ from this regression will have a causal interpretation (Robins, Hernán, and Brumback, 2000). For this case of a single time-period outcome, the usual standard errors estimates, ignoring the estimation of the weights, will be conservative estimates of the true standard errors. This holds because the estimation of the weights actually increases the efficiency of the MSM estimator so that ignoring their estimation only increases the standard error estimates (Robins, 2000).

IPTW and MSMs are not the only way to estimate historical effects. Other estimation strategies rely explicitly on the *g*-computational formula, which uses the entire joint distribution of the data, outcomes and time-varying covariates, to estimate any causal effect (Robins, Greenland, and Hu, 1999). There are a number of ways to implement the *g*-computational formula, including structurally nested models and Bayesian simulation. This approach is very flexible, but it generally requires a model for the distribution of the covariates over time, which can be a large burden for empirical researchers who view these covariates as simply a tool to control for potential bias. Moreover, the dimension of these covariates can be quite large. Robins (2000) and Robins, Greenland, and Hu (1999) discuss some of the tradeoffs involved in choosing among these methods. Appendix A presents a side-by-side stylized demonstration of g-computation and IPTW with MSMs in a simplified setting. Note that all of these approaches assume sequential ignorability.

5 The promises and pitfalls of repeated outcome measurements

Of course, most TSCS studies use more than just the outcome for the last time period when estimating causal effects. There are a number of reasons for this, but on a basic level, the choice of the last period outcome as *the* outcome is arbitrary. Nothing would stop us from choosing the outcome from any other year as the outcome variable, and this along with other considerations may lead the analyst toward a model with repeated outcomes. However, the use of a repeated outcomes creates a number of complications for the estimation of CETs and ATHEs, the most immediate of which is that these quantities are no longer uniquely defined: there is a CET and an ATHE defined for the outcome variable at each point in time. In this section, we clarify exactly how to use and explore the existence of repeated outcomes to strengthen the estimation of causal effects.

5.1 Contemporaneous effects with repeated measurements

The estimation of CETs with repeated outcomes has been covered extensively in the political science literature (see Beck and Katz (2011) for a review), but we review certain aspects here that are relevant to the estimation ATHEs with repeated outcomes. First, as noted above, one could estimate different CETs for the outcome in each year, although usually an assumption is made that allows some amount of pooling across years. Most simply, we might assume that the CET is the same for all years, although this can be relaxed. Second, even if we assumed that the CET were the same for all years, when estimating this time-constant CET, we would have to account for the dependence in the outcome variable across time. To some extent, this concern may be ameliorated by the control variables that are included in order to justify sequential ignorability. For example, we might assume that the lagged dependent variable must be included in the model in order to identify the causal effect and the inclusion of this variable would also alleviate some of the concerns regarding dependence over time (Beck and Katz, 1996). Further concerns about independence can often be addressed using a robust approach such as generalized estimating equations or GEE (Liang and Zeger 1986; Zeger and Liang 1986).

5.2 Treatment history effects with repeated measurement

The treatment regimes of a CET share a history up to time $t-1$ and this simplifies the estimation of CETs with repeated outcomes. Although the CET for time t will be associated with a longer treatment regime than the CET for time $t-1$, both CETs will be concerned with a change in the treatment only in the final period, and hence it is interpretable to pool these CETs across time. In contrast, ATHEs will not be generally comparable across time, because the ATHE for time t will be associated with a longer treatment regime than the CET for time $t-1$. For example, when $t = 2$ we might consider the ATHE $E[Y_{i2}(1,0) - Y_{i2}(0,1)]$, but it is not possible to construct an analogous ATHE for $t = 1$. In the first period, there is only one potential outcome with one cumulative period of treatment, $Y_{i1}(1)$, whereas there are two histories with this cumulative sum in the second period $Y_{i2}(1,0)$ and $Y_{i2}(0,1)$. Without additional structure, there is no way to connect comparisons across these periods. Therefore, when we decide to pool ATHEs across time, we restrict the space of possible contrasts and MSMs that we might consider.

It is helpful to consider two basic MSMs (and extensions) that allow pooling across time, and to consider what assumptions these MSMs imply about the ATHEs. For simplicity we present only linear models here, but the issues are analogous for non-linear models. Recall the linear MSM considered above,

$$E[Y_{it}(\underline{a})] = \beta_0 + \beta_1 a_t + \beta_2 a_{t-1} + \beta_3 a_{t-2}, \tag{13}$$

This MSM assumes that treatments more than two periods prior do not affect the mean outcome in time t. Therefore, as long as the outcomes from $t = 1$ and $t = 2$ are not used in the analysis, this MSM will be comparable across repeated outcomes.

In contrast, consider the cumulative MSM,

$$E[Y_{it}(\underline{a})] = \beta_0 + \beta_1 \sum_{k=1}^{t} a_{ik}, \tag{14}$$

This MSM allows treatments from more than two periods ago to affect the mean outcome by assuming that only the sum of treatments will affect the mean outcome in time t. This MSM is also comparable for different t, although the sum will have

more terms as t grows, and therefore it is typical to include a time term in this model. For example, we might include time additively in the following manner,

$$E[Y_{it}(\underline{a})] = \beta_0 + \beta_1 \sum_{k=1}^{t} a_{it} + \beta_2 t, \tag{15}$$

This allows some heterogeneity of the MSM across time periods, although it still assumes a constant ATHE across time.

Finally, we might want to allow the current treatment to be modeled differently than past treatments, so we could use hybrids of these MSMs.

$$E[Y_{it}(\underline{a})] = \beta_0 + \beta_1 a_{it} + \beta_2 \sum_{k=1}^{(t-1)} a_{ik} + \beta_2 t, \tag{16}$$

This MSM allows the treatment in the current period to have an effect separate from the cumulative effect for the other periods.

Finally, note that we can allow much more heterogeneity in time for the MSM. It is possible to include splines or time-fixed effects for time in order to allow non-linear heterogeneity in the MSM. It is also possible to include interactions between the treatment terms and the time terms in order to allow heterogeneity in the ATHEs across time.

6 The effect of trade on taxation in OECD countries

Should the details of ATHEs burden the applied researcher? Or are they mere technicalities? In this section we address the peril of ignoring the subtleties of causality in TSCS data with the analysis of political economy data on the OECD nations. The above weighting approach overturns results from the conventional TSCS approach when we apply each to the data from Swank and Steinmo (2002). These scholars estimate the effects of domestic economic policies on tax rates in advanced industrialized democracies. Here we focus on one of their explanatory variables, trade openness, and its effect on one of their outcomes, the effective tax rate on labor. In their models, Swank and Steinmo find trade openness to have no statistically significant effect on these tax rates, but they only considered the effect of trade openness in the previous year. While Swank and Steinmo discuss

the long-run effects of economic policies, they only estimate the contemporaneous effect of this trade policy, leaving aside any effects of history.

Swank and Steinmo adhere to the guidance of previous methodological research on TSCS data (Beck and Katz, 1996). The authors regress the tax rate in a given year on economic and political features of each country from the previous year. In addition to trade openness (A_{t-1}), these attributes include liberalization of capital controls, unemployment, leftist share of the government, and importantly, a lagged measure of the dependent variable. We refer to the lagged dependent variable as Y_{t-1} and the set of attributes (excluding trade openness) as X_{t-1}. Thus we can write the main estimating equation of SS as:

$$Y_{it} = \beta_0 + \beta_1 A_{i,t-1} + \beta_2 Y_{i,t-1} + \beta_3 X_{i,t-1} + \varepsilon_{it}. \tag{17}$$

Keep in mind that β_1 only has a causal interpretation as the CET when sequential ignorability holds and when the effect of A_{t-1} is constant across the covariates, Y_{t-1} and X_{t-1}. When the effect size varies across levels of these covariates, we can always use an imputation estimator to avoid the regression weighting problem (Imbens 2004).

To uncover any historical effects of trade openness on the labor tax rate, we expand the model of Swank and Steinmo beyond a single lag. We instead take the cumulative years of trade openness as our main independent variable:[5]

$$Y_{it} = \beta_0 + \beta_1 \left(\sum_{k=1}^{t-1} A_{i,k} \right) + \beta_2 Y_{i,t-1} + \beta_3 X_{i,t-1} + \nu_{it}. \tag{18}$$

Unfortunately, post-treatment bias ruins the causal interpretation of the coefficient on our new measure, β_1. Earlier values of trade openness, such as A_{t-2}, might affect the lagged tax rate, for instance. To avoid this difficulty, we can take a second approach—omitting the time-varying confounders, Y_{t-1} and X_{t-1}, from our model. Here we would estimate the effect of trade openness only conditioning on a time trend:

$$Y_{it} = \beta_0 + \beta_1 \left(\sum_{k=1}^{t-1} A_{i,k} \right) + \beta_2 t + \eta_{it}. \tag{19}$$

[5] Here we need trade openness as a binary treatment, so we create a new trade openness variable which is 1 if the county-year had a score at or above the median of the entire sample. The results are substantively unchanged if we use continuous measures, though IPTW in those situations has much poorer properties (Goetgeluk, Vansteelandt, and Goetghebeur 2008).

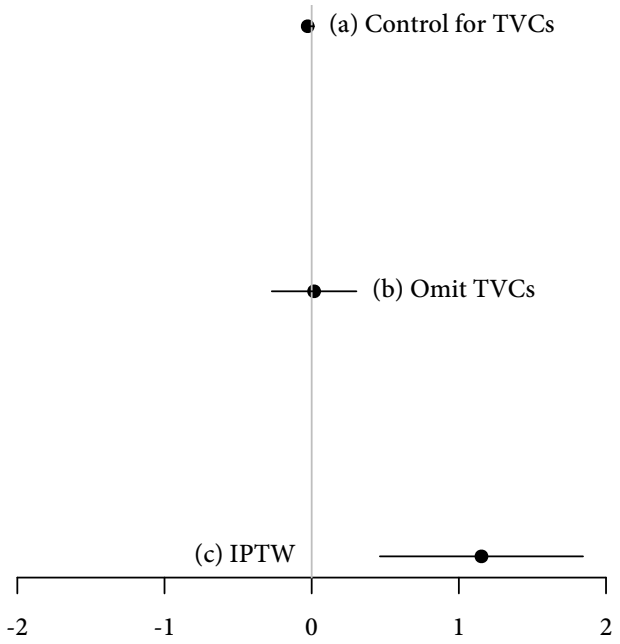

Figure 3: Estimated effect on labor tax rates of cumulative trade openness using three models. They represent the estimated effect and 95% confidence interval (a) when controlling for variables that trade openness affects, (b) when omitting those variables from the model, and (c) using the recommended IPTW approach.

While this method avoids the issue of post-treatment bias entirely, it admits the possibility of omitted variable bias. If past values of the tax rate affect future trade openness, for instance, then excluding these lags of the dependent variable will produce bias in our estimated effects. Each approach has its drawbacks, but we can learn a great deal by comparing their results to our preferred weighting method.

What do these approaches discover about the effects of trade openness? As Figure 3 shows, nothing more than the original contemporaneous effect estimates. Both methods—omitting and controlling for time-varying confounders—lead to

the same basic conclusion: there is no statistically significant effect of trade openness on tax policy.[6] These results are consistent with the findings of Swank and Steinmo (2002). This agreement tempts us to confirm the approximate validity of their results; after all, a natural intuition would be that the true effect must be between these two estimates. Unfortunately, this intuition, while natural, is incorrect. The biases of both approaches can be in the same direction, negating their usefulness as bounds (Blackwell 2013).

An alternative to both of these approaches is our weighting method. To implement IPTW in this case, we omit the time-varying confounders from the tax rate model and instead include those in a logistic propensity score model to create weights as in (11). We then use those weights in a weighted GEE model. Instead of controlling for the time-varying confounders in our regression model, these weights adjust for the confounding in the time-varying covariates without inducing post-treatment bias. Figure 3 shows the IPTW estimates are not only significant and positive, but also far larger in magnitude than either of the other approaches. Thus, the particulars of ATHEs and IPTW are more than mere technicality—they can transform seemingly robust substantive results.

7 Conclusions, drawbacks, and future research

Repeated measurements over time of countries, people, or governments expand the scope of causal inference methods. TSCS data allow us to estimate both contemporaneous effects and treatment history effects. But with an expanded scope comes complications. The usual regression and matching methods break down for historical effects. Nevertheless, we have shown that weighting approaches can overcome these difficulties and recover effect estimates across a wide variety of settings.

IPTW methods have their own drawbacks, of course. The weighting approach, for instance, relies on a comparability assumption, which we call sequential ignorability. Regression and matching estimators of contemporaneous effects, though, rely on the same assumption for their causal interpretation. Yet it is important to consider how our estimates change when this assumption is incorrect. The formal

[6] We estimate both of these models using a GEE approach with robust standard errors.

sensitivity analysis of Blackwell (2014) can quantify the dependence of our estimates on this assumption and applies to TSCS data. The weighting approach also requires an analyst to choose the correct propensity score model. In this paper we have focused on parametric models for the propensity score, but methods such as nonparametric methods (Hirano, Imbens, and Ridder, 2003) or balance restrictions (Imai and Ratkovic 2013) would provide robustness against these modeling choices.

In this paper, we focused on the usual sequential ignorability assumption as commonly invoked in epidemiology. Many TSCS applications in political science rely on a "fixed effects" assumption that there is time-constant, unmeasured heterogeneity in units. Linear models can easily handle these types of assumptions, though nonlinear fixed effects models pose greater difficulties. Estimating the above causal quantities with these models, however, remains elusive except under strong assumptions like strict (not sequential) ignorability (Chernozhukov et al. 2009; Sobel 2012). Future work should investigate how causal estimation could proceed under a unit-specific version of sequential ignorability, where the key assumption only holds within units. Estimators that exploit this assumption could bring together the rich set of causal quantities with the attractive within-unit variation of TSCS data.

References

Angrist, Joshua D., and Alan B Krueger. 1999. "Empirical strategies in labor economics." In *Handbook of labor economics,* 1277–1366. Elsevier. (Cited on page 9).

Angrist, Joshua D., and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics.* An Empiricist's Companion. Princeton University Press, December. (Cited on page 9).

Aronow, Peter M., and Cyrus Samii. 2013. "Does Regression Produce Representative Estimates of Causal Effects?" Presented at the Annual Meeting of the European Political Science Assocation, 2013, Barcelona, Spain. (Cited on page 9).

Beck, Nathaniel, and Jonathan N. Katz. 1996. "Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models." *Political Analysis* 6 (1): 1–36. (Cited on pages 14, 17).

Beck, Nathaniel, and Jonathan N. Katz. 2011. "Modeling Dynamics in Time-Series–Cross-Section Political Economy Data." *Annual Review of Political Science* 14, no. 1 (June): 331–352. (Cited on page 14).

Blackwell, Matthew. 2013. "A Framwork for Dynamic Causal Inference in Political Science." *American Journal of Political Science* 57 (2): 504–520. (Cited on pages 3, 10, 11, 19).

———. 2014. "A Selection Bias Approach to Sensitivity Analysis for Causal Effects." *Political Analysis* 22 (2): 169–182. (Cited on page 20).

Box, George EP, Gwilym M Jenkins, and Gregory C Reinsel. 2013. *Time series analysis: forecasting and control.* Wiley. (Cited on page 10).

Chernozhukov, Victor, Ivan Fernandez-Val, Jinyong Hahn, and Whitney Newey. 2009. "Average and Quantile Effects in Nonseparable Panel Models." *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence* stat.ME (April). (Cited on page 20).

Goetgeluk, Sylvie, Sijn Vansteelandt, and Els Goetghebeur. 2008. "Estimation of Controlled Direct Effects." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 70, no. 5 (November): 1049–1066. (Cited on page 17).

Hirano, Keisuke, Guido W. Imbens, and Geert Ridder. 2003. "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score." *Econometrica* 71 (4): pages. (Cited on page 20).

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth Stuart. 2007. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 15:199–236. (Cited on page 9).

Imai, Kosuke, and Marc Ratkovic. 2013. "Covariate balancing propensity score." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*:n/a–n/a. (Cited on pages 12, 20).

Imbens, Guido W. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review." *Review of Economics and Statistics* 86, no. 1 (February): 4–29. (Cited on pages 9, 17).

Liang, Kung-Yee, and Scott L Zeger. 1986. "Longitudinal data analysis using generalized linear models." *Biometrika* 73 (1): 13–22. (Cited on page 14).

McCaffrey, Daniel F, Greg Ridgeway, and Andrew R Morral. 2004. "Propensity score estimation with boosted regression for evaluating causal effects in observational studies." *Psychol Methods* 9 (4): 403–425. (Cited on page 12).

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

———. 2000. "Marginal Structural Models versus Structural Nested Models as Tools for Causal Inference." In *Statistical Models in Epidemiology, the Environment, and Clinical Trials,* edited by M. Elizabeth Halloran and Donald Berry, 95–134. Vol. 116. The IMA Volumes in Mathematics and its Applications. New York: Springer-Verlag. (Cited on page 13).

Robins, James M., Sander Greenland, and Fu-Chang Hu. 1999. "Estimation of the Causal Effect of a Time-Varying Exposure on the Marginal Mean of a Repeated Binary Outcome." *Journal of the American Statistical Association* 94 (447): pages. (Cited on page 13).

Robins, James M., Miguel A. Hernán, and Babette A. Brumback. 2000. "Marginal Structural Models and Causal Inference in Epidemiology." *Epidemiology* 11 (5): 550–560. (Cited on pages 3, 10, 11, 13).

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

Sobel, Michael E. 2012. "Does Marriage Boost Men's Wages?: Identification of Treatment Effects in Fixed Effects Regression Models for Panel Data." *Journal of the American Statistical Association* 107, no. 498 (June): 521–529. (Cited on pages 10, 20).

Swank, Duane, and Sven Steinmo. 2002. "The New Political Economy of Taxation in Advanced Capitalist Democracies" [in English]. *American Journal of Political Science* 46 (3): pages. (Cited on pages 3, 16, 17, 19).

Zeger, Scott L, and Kung-Yee Liang. 1986. "Longitudinal data analysis for discrete and continuous outcomes." *Biometrics:*121–130. (Cited on page 14).

Appendix A Stylized Demonstration of G-Computation and IPTW with MSMs

The *g*-computational approach may be more intuitive for readers familiar with traditional techniques for modeling dynamics in a TSCS setting, so in order to illustrate the benefits of the IPTW and MSM approach, it will be helpful to demonstrate the *g*-computational approach in a special case. Here we consider estimating treatment history effects for a single outcome with two time periods, where the parameter of interest, $E[Y_i(\underline{a})]$, simplifies to $E[Y_i(a_1, a_0)]$. The specification of MSMs is sometimes not necessary for only two time periods, and we use the general presentation here in order to focus on the pertinent issues.

The *g*-computational approach for this scenario, presented below, demonstrates that the expectation of $E[Y_i(a_1, a_0)]$ can be derived using the assumptions of consistency and sequential ignorability (the *i* subscripts are removed and discrete covariates are assumed for notational simplicity). In principal, the *g*-computational formula,

$$\sum_{x_0} \sum_{x_1} E[Y|a_1, x_1, a_0, x_0] f(x_1|a_0, x_0) f(x_0),$$

allows the estimation of treatment history effects by substituting estimates for the expectation and two densities, and integrating over the covariates at time zero and time one. Note that *g*-computation deals with time-varying covariates by estimating the probability of the time one covariates (x_1) based on the time zero covariates and the time zero treatment (a_0, x_0). Hence the key component of the *g*-computational approach is the inclusion of the $f(x_1|a_0, x_0)$ density in the integral. The following proof demonstrates the validity of this approach.

$$
\begin{aligned}
&\sum_{x_0}\sum_{x_1} E[Y|a_1, x_1, a_0, x_0]f(x_1|a_0, x_0)f(x_0) \\
&\qquad = \sum_{x_0}\sum_{x_1} E[Y(a_1, a_0)|x_0, x_1, a_0, a_1]f(x_1|x_0, a_0)f(x_0) \textit{ (by consistency)} \\
&\qquad = \sum_{x_0}\sum_{x_1} E[Y(a_1, a_0)|x_1, a_0, x_0]f(x_1|a_0, x_0)f(x_0) \textit{ (by sequential ignorability)} \\
&\qquad = \sum_{x_0} E[Y(a_1, a_0)|a_0, x_0]f(x_0) \textit{ (by LTP)} \\
&\qquad = \sum_{x_0} E[Y(a_1, a_0)|x_0]f(x_0) \textit{ (by sequential ignorability)} \\
&\qquad = E[Y(a_1, a_0)]
\end{aligned}
$$

However, $f(x_1|a_0, x_0)$ requires estimating the density for combinations of x_1, x_0, and a_0 that do not exist in the data. Therefore, when x_1 has a high dimension, the estimation of $f(x_1|a_0, x_0)$ represents a difficult modeling problem. (In contrast, $f(x_0)$ need only be estimated for observed values of x_0 and hence even a high dimensional $f(x_0)$ can be estimated with the empirical distribution of the time zero covariates.)

The weighting approach discussed in this paper avoids this modeling problem. Weighting adjusts the observed data so that the treatment history is unrelated to the past covariate history. The following proof shows that the weighted conditional mean of the outcome equals the marginal mean of the potential outcomes under consistency and sequential ignorability. Here, we use $I(A_1 = a_1)$ as an indicator variable for when $A_1 = a_1$, *LTP* for the law of total probability, and *CE* to denote an equality follows from the properties of conditional expectation. We prove this result using the unstabilized weights, but it is straightforward to extend to the stabilized weights.

$$
\begin{aligned}
&E_W\left[Y|a_1,a_0\right] \\
&= E_W[I(A_1=a_1)I(A_0=a_0)Y]/E_W[I(A_1=a_1)I(A_0=a_0)] \qquad \textit{(by CE)} \\
&= E_W[I(A_1=a_1)I(A_0=a_0)Y] \qquad \textit{(by result below)} \\
&= E\big[W(A_1,X_1,A_0,X_0)I(A_1=a_1)I(A_0=a_0)Y\big] \qquad \textit{(by definition of weighted mean)} \\
&= E\big[W(A_1,X_1,A_0,X_0)I(A_1=a_1)I(A_0=a_0)Y(a_1,a_0)\big] \qquad \textit{(by consistency)} \\
&= E\left[\frac{I(A_1=a_1)I(A_0=a_0)}{f(a_1|X_1,a_0,X_0)f(a_0|X_0)}Y(a_1,a_0)\right] \qquad \textit{(by definition of weights)} \\
&= E\left[E\left[\frac{I(A_1=a_1)I(A_0=a_0)}{f(a_1|X_1,a_0,X_0)f(a_0|X_0)}Y(a_1,a_0)\middle|X_0\right]\right] \qquad \textit{(by LTP)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[\frac{I(A_1=a_1)I(A_0=a_0)}{f(a_1|X_1,a_0,X_0)}Y(a_1,a_0)\middle|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[E\left[\frac{I(A_1=a_1)I(A_0=a_0)}{f(a_1|X_1,a_0,X_0)}Y(a_1,a_0)\middle|X_1,X_0\right]\middle|X_0\right]\right] \qquad \textit{(by LTP)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[\frac{1}{f(a_1|X_1,a_0,X_0)}E\left[I(A_1=a_1)I(A_0=a_0)Y(a_1,a_0)|X_1,X_0\right]\middle|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[\frac{1}{f(a_1|X_1,a_0,X_0)}E\left[I(A_1=a_1)Y(a_1,a_0)|X_1,a_0,X_0\right]f(a_0|X_1,X_0)\middle|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[\frac{E\left[I(A_1=a_1)|X_1,a_0,X_0\right]}{f(a_1|X_1,a_0,X_0)}E\left[Y(a_1,a_0)|X_1,a_0,X_0\right]f(a_0|X_1,X_0)\middle|X_0\right]\right] \qquad \textit{(by sequential ignorability)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[E\left[Y(a_1,a_0)|X_1,a_0,X_0\right]f(a_0|X_1,X_0)\middle|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[E\left[I(A_0=a_0)Y(a_1,a_0)|X_1,X_0\right]\middle|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[I(A_0=a_0)Y(a_1,a_0)|X_0\right]\right] \qquad \textit{(by LTP)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[I(A_0=a_0)|X_0\right]E\left[Y(a_1,a_0)|X_0\right]\right] \qquad \textit{(by sequential ignorability)} \\
&= E\left[E\left[Y(a_1,a_0)|X_0\right]\right] \qquad \textit{(by CE)} \\
&= E\left[Y(a_1,a_0)\right] \qquad \textit{(by LTP)}
\end{aligned}
$$

If the treatments here are binary, it is clear that we could rely on a sample weighted mean among units that followed a particular treatment history. By the law of large numbers, this sample weighted mean will be consistent for the population conditional expectation and, thus, the mean potential outcome. The above proof relies on knowing the weights. If we replace the known weights with a consistent estimator for the weights, $\hat{W}(A_1, X_1, A_0, X_0; \hat{\alpha}) \xrightarrow{p} W(A_1, X_1, A_0, X_0)$, then it is easy to show using Slutsky's Theorem that the weighted mean will be consistent for the mean of the potential outcomes.

Above we used the fact that $E_W[I(A_1 = a_1)I(A_0 = a_0)] = 1$. Below we prove this result:

$$
\begin{aligned}
E_W[I(A_1 = a_1)I(A_0 = a_0)] &= E[W(A_1, X_1, A_0, X_0)I(A_1 = a_1)I(A_0 = a_0)] && \textit{(by weighted mean definition)} \\
&= E\left[\frac{I(A_1 = a_1)I(A_0 = a_0)}{f(a_1|X_1, a_0, X_0)f(a_0|X_0)}\right] && \textit{(by definition of weights)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[E\left[\frac{I(A_1 = a_1)I(A_0 = a_0)}{f(a_1|X_1, a_0, X_0)}\middle| X_1, X_0\right]\middle| X_0\right]\right] && \textit{(by LTP)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[\frac{1}{f(a_1|X_1, a_0, X_0)}E\left[I(A_1 = a_1)|X_1, a_0, X_0\right]f(a_0|X_1, X_0)\middle| X_0\right]\right] && \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[f(a_0|X_1, X_0)\middle| X_0\right]\right] && \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E\left[E[I(A_0 = a_0)|X_1, X_0]\middle| X_0\right]\right] && \textit{(by CE)} \\
&= E\left[\frac{1}{f(a_0|X_0)}E[I(A_0 = a_0)|X_0]\right] && \textit{(by LTP)} \\
&= 1 && \textit{(by CE)}
\end{aligned}
$$