Front-door Difference-in-Differences Estimators*

Adam Glynn† Konstantin Kashin‡

aglynn@emory.edu kkashin@fas.harvard.edu

November 9, 2015

Abstract

In this paper, we develop front-door difference-in-differences estimators that utilize information from post-treatment variables in addition to information from pre-treatment covariates. Even when the front-door criterion does not hold, these estimators allow the identification of causal effects in observational studies under an assumption of one-sided noncompliance, an exclusion restriction, and additional assumptions similar to difference-in-differences assumptions. These estimators also allow the bounding of causal effects under relaxed assumptions, and surprisingly, do not use traditional control units. We illustrate these points with an application to job training study and with an application to Florida's early in-person voting program. For the job training study, we show that these techniques can recover an experimental benchmark. For the Florida program, we find some evidence that early in-person voting had small positive effects on turnout in 2008. This provides a counterpoint to recent claims that early voting had a negative effect on turnout in 2008.

Word count: 7676

*We thank Barry Burden, Justin Grimmer, Manabu Kuroki, Kevin Quinn, and seminar participants at Emory, Harvard, Notre Dame, NYU, Ohio State, UC Davis, and UMass Amherst for comments and suggestions. Earlier versions of this paper were presented at the 2014 MPSA Conference and the 2014 Asian Political Methodology Meeting in Tokyo.

†Department of Political Science, Emory University, 327 Tarbutton Hall, 1555 Dickey Drive, Atlanta, GA 30322 (http://scholar.harvard.edu/aglynn).

‡Department of Government and Institute for Quantitative Social Science, Harvard University, 1737 Cambridge Street, Cambridge, MA 02138 (http://konstantinkashin.com).

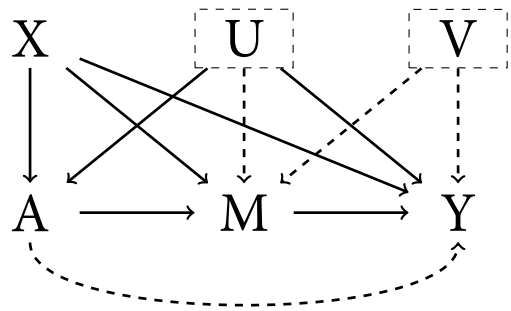
1 Introduction

One of the main tenets of observational studies is that post-treatment variables should not be included in an analysis because naively conditioning on these variables can block some of the effect of interest, leading to post-treatment bias (King, Keohane and Verba, 1994). While this is usually sound advice, it seems to contradict recommendations from the process tracing literature that information about mechanisms can be used to assess the plausibility of an effect (Collier and Brady, 2004; George and Bennett, 2005; Brady, Collier and Seawright, 2006). The front-door criterion (Pearl, 1995) and its extensions (Kuroki and Miyakawa, 1999; Tian and Pearl, 2002*a*,*b*; Shpitser and Pearl, 2006) resolve this apparent contradiction, providing a means for nonparametric identification of treatment effects using post-treatment variables. Importantly, the front-door approach can identify causal effects even when there are unmeasured common causes of the treatment and the outcome (i.e., the total effect is confounded). Figure 1 presents the directed acyclic graph associated with the front-door criterion. The formal definition of this graph can be found in Pearl (1995, 2009), but for our purposes, it will suffice to note the following: A represents the treatment/action variable, M represents a set of mediating variables (often a singleton), Y represents the outcome, X represents covariates, U and V represent sets of unobserved variables, and arrows represent the possible existence of effects from one set of variables to another.[1] Solid arrows are allowed for the front-door criterion to hold. Note the existence of solid arrows from U to both A and Y. Hence, unmeasured common causes of the treatment and outcome are allowed. As can be seen in Figure 1, the front-door approach works by identifying the effects of A on M and the effects of M on Y, and then putting them back together.

However, the front-door adjustment has been used infrequently (VanderWeele, 2009) due to concerns that the assumptions required for point identification will rarely hold (Cox and Wermuth,

[1]To simplify presentation, we have not included arrows between X, U, and V. While the graph implies that these sets of variables are independent, this is not required for the techniques below.

Figure 1: Front-door Directed Acyclic Graph (DAG). A represents the treatment/action variable, M represents a set of mediating variables, Y represents the outcome, X represents covariates, and U and V represent sets of unobserved variables. To simplify presentation, we have assumed that X, U, and V are independent (this is implied by the lack of arrows between them), but this is not required. Solid arrows are allowed for the front-door criterion to hold within this group. Dashed arrows are not allowed for the front-door criterion to hold in this group.



1995; Imbens and Rubin, 1995). These assumptions are represented by the dashed arrows in Figure 1. Hence, while common causes of A and Y are allowed for the front-door criterion to hold, common causes of M and Y (not mediated by A) are not allowed. Additionally, the front-door criterion will not hold when A has a direct effect on Y.

A number of papers have proposed weaker and more plausible sets of assumptions (Joffe, 2001; Kaufman, Kaufman and MacLehose, 2009; Glynn and Quinn, 2011) that tend to correspond to conceptions of process tracing. However, these approaches rely on binary or bounded outcomes, and even in large samples these methods only provide bounds on causal effects (i.e., partial instead of point identification). In this paper, we use bias formulas for the front-door approach and demonstrate that we can remove or ameliorate this bias via a difference-in-differences approach when there is one-sided noncompliance. We also illustrate that under one-sided noncompliance, the front-door estimator implies scaling the effect of the mediator on the outcome (so that it estimates the effect of the treatment).

We take a difference-in-differences (DD) approach to removing the bias from the front-door estimator. At the most basic level, a DD estimator tries to correct the bias coming from a standard estimator by estimating this bias from a set of observations for which there should be no effect.

In this paper, we will refer to the group of observations for which there should be no effect as the *differencing group*, and the observations on which the standard estimator operates as the *group of interest*. In many cases, an over-time DD approach is used such that the group of interest is the set of observations (both treatment and control) taken at a point in time after the treatment has been applied to the treated units. A simple estimator for that set is often the difference in mean post-treatment outcomes between the treated and control units. In an attempt to remove bias due to differences between treated and control units in the group of interest not attributable to the treatment, pre-treatment observations of the treated and control units are used as the differencing group in over-time DD.[2] The simple difference-in-means estimate for the differencing group is taken to be evidence of bias and subtracted off from the estimate for the group of interest.

Although over-time DD is the most common DD approach, the concept of a differencing group — a group of observations for which there should be no effect — is more general than pre-treament outcome observations. Non-over-time differencing groups are often found within difference-in-difference-in-differences (DDD) strategies. For example, one might use age-eligibility cutoffs to find a group of people who are not eligible for a program, and hence for whom the program should have no effect (see pages 242–243 of Angrist and Pischke (2009) for a description of this "higher order contrast" approach). Regardless of whether an over-time or a non-over-time DD is used, the standard DD approach can be conceptualized as the following: first, estimate the effect among the group of interest; second, estimate the bias for the group of interest as the estimated effect of the treatment among the differencing group; third, take the difference between the two estimates.

The front-door difference-in-differences (front-door DD) approach extends the front-door approach in a similar manner, but with two major differences. First, the differencing is done with

[2]The over-time DD is often rearranged as the difference between the over-time differences between the treated and control units. Although algorithmically distinct, the final result is numerically equivalent.

respect to the effect of the mediator, and then the estimated "mediator effect" is scaled to estimate the effect of the treatment. Second, over-time DD is sometimes not possible because mediator information may not be available in the pre-treatment period. This is often the case with repeated cross-section designs where the mediator is measured at the individual level.

With these two differences in mind, the front-door DD proceeds analogously to the DD approach. First, we identify the group of interest. Second, we identify a group of treated units distinct from our group of interest (perhaps using pre-treatment observations) for which we believe the treatment should have no effect (or a small effect). A non-zero front-door estimate for this group can then be attributed to bias. For an over-time example, we consider a job training program with the pre-program observations on individuals as the differencing group. In a non-over-time example, we estimate the effects of an early in-person (EIP) voting program on turnout for elections leveraging voters that used an absentee ballot in the previous election as a differencing group.[3]

If we further assume that the estimated bias from the differencing group is equal to the bias for our group of interest, then by subtracting the front-door estimator for this group from the front-door estimator for the group of interest, we can remove the bias from our front-door estimate for the group of interest. Note that if all effects and bias are positive, then when the estimate from the differencing group is larger than the bias for the group of interest, this differencing approach

[3]EIP was unlikely to have a large effect *on turnout* for these voters, as they had already demonstrated their ability to vote by another means. Specifically, the existence of an EIP program in 2012 might have induced some 2008 absentee ballot users to change their mode of voting in 2012 (e.g., from absentee to EIP), but it is unlikely to have caused them to vote. This is because 2008 absentee voters who voted EIP in 2012, would likely have just voted absentee in 2012 if the EIP program did not exist in 2012. Therefore, we consider non-zero front-door estimates of the turnout effect for this group to be evidence of bias. This point and evidence for it is discussed in more detail in the application.

can provide a lower bound on the effect of the program. Furthermore, if the front-door approach provides an upper bound, then the front-door and front-door DD approaches can be combined in a bracketing approach.[4] We demonstrate this bracketing within the context of the job training study with a non-over-time differencing group. Finally, we note that as with the standard difference-in-differences approach, when the estimate from the differencing group is smaller than the bias for the group of interest, one can get results that are too large.

The paper is organized as follows. Section 2 presents the bias formulas for the front-door approach to estimating average treatment effects on the treated (ATT) for nonrandomized program evaluations with one-sided noncompliance. Section 3 presents the difference-in-differences approach for front-door estimators for the simplified case and discusses the required assumptions. Section 4 presents an application of the front-door difference-in-differences estimator to the National JTPA (Job Training Partnership Act) Study. Section 5 presents an application of the front-door difference-in-differences estimator to election law: assessing the effects of early in-person voting on turnout in Florida. Section 6 concludes.

2 Bias for the Front-Door Approach for ATT

In this section, we present large-sample bias formulas for the front-door approach for estimating the average treatment effect on the treated (ATT). Throughout this paper, all references to bias will mean large-sample bias in the context of nonparametric estimation. This allows us to avoid questions of modeling.

ATT is often the parameter of interest when assessing the effects of a program or a law. For an outcome variable Y and a binary treatment/action A, we define the potential outcome under active

[4]This bracketing approach is similar in spirit to the use of fixed effects and lagged dependent variables for bracketing (see page 245 of Angrist and Pischke (2009)).

treatment as $Y(a_1)$ and the potential outcome under control as $Y(a_0)$.[5] Our parameter of interest is the ATT, defined as $\tau_{att} = E[Y(a_1)|a_1] - E[Y(a_0)|a_1] = \mu_{1|a_1} - \mu_{0|a_1}$. We assume consistency, $E[Y(a_1)|a_1] = E[Y|a_1]$, so that the mean potential outcome under active treatment for the treated units is equal to the observed outcome for the treated units such that $\tau_{att} = E[Y|a_1] - E[Y(a_0)|a_1]$. The ATT is therefore the difference between the mean outcome for the treated units and mean counterfactual outcome for these units, had they not received the treatment.

We also assume that $\mu_{0|a_1}$ is potentially identifiable by conditioning on a set of observed covariates X and unobserved covariates U. To clarify, we assume that if the unobserved covariates were actually observed, the ATT could be estimated by standard approaches (e.g., matching). For simplicity in presentation we assume that X and U are discrete, such that

$$\mu_{0|a_1} = \sum_x \sum_u E[Y|a_0, x, u] \cdot P(u|a_1, x) \cdot P(x|a_1),$$

but continuous variables can be handled analogously. However, even with only discrete variables we have assumed that the conditional expectations in this equation are well-defined, such that for all levels of X and U amongst the treated units, all units had a positive probability of receiving either treatment or control (i.e., positivity holds).

The front-door adjustment for a set of measured post-treatment variables M can be written as the following:

$$\mu^{fd}_{0|a_1} = \sum_x \sum_m P(m|a_0, x) \cdot E[Y|a_1, m, x] \cdot P(x|a_1).$$

Conditioning on a_1 is a slight adjustment from the original front-door formula (Pearl, 1995), that targets the average for the treated units instead of all units. We can thus define the large-sample

[5]Note that we must assume that these potential outcomes are well defined for each individual, and therefore we are making the stable unit treatment value assumption (SUTVA).

front-door estimator of ATT as:

$$\tau_{att}^{fd} = \mu_{1|a_1} - \mu_{0|a_1}^{fd}.$$

For the difference-in-differences estimators we consider in this paper, we use the special case of nonrandomized program evaluations with one-sided noncompliance. Following the literature in econometrics on program evaluation, we define the program impact as the ATT where the active treatment (a_1) is assignment into a program (Heckman, LaLonde and Smith, 1999), and when M indicates whether the active treatment (a_1) was actually received. We use the short-hand notation m_1 to denote that active treatment was received and m_0 if it was not.

Assumption 1 (One-sided noncompliance)
$P(m_0|a_0, x) = P(m_0|a_0, x, u) = 1$ *for all* x, u.

Assumption 1 implies that only those assigned to treatment can receive treatment.[6] The front-door

[6]We might wonder how often one-sided noncompliance is likely to hold when the treatment is not assigned randomly. Stated differently, if we control the treatment to the extent that those not assigned to treatment cannot get it, why would we not randomize the treatment. The early voting application later in the paper provides the clearest answer to this question. Often, due to logistical or ethical concerns, a treatment cannot be withheld from any individual. Additionally, we might wonder whether the effect of treatment assignment would still be of interest in this circumstance. The effect of treatment assignment (often known as the intent to treat effect) is often of interest when assignment is manipulable as a policy variable and compliance is not (Heckman et al., 1998). Again, the early voting application later in this paper provides an example of this.

large-sample estimator can be re-written in the following manner.

$$
\begin{aligned}
\tau^{fd}_{att} &= \mu_{1|a_1} - \mu^{fd}_{0|a_1} \\
&= E[Y|a_1] - \sum_x \sum_m P(m|a_0, x) \cdot E[Y|a_1, m, x] \cdot P(x|a_1) \\
&= E[Y|a_1] - \sum_x \underbrace{E[Y|a_1, m_0, x]}_{\text{treated non-compliers}} \cdot P(x|a_1) \qquad (1) \\
&= \sum_x P(x|a_1) \cdot P(m_1|x, a_1) \cdot \left\{ \underbrace{E[Y|a_1, m_1, x] - E[Y|a_1, m_0, x]}_{\text{“effect” of receiving treatment}} \right\} \qquad (2)
\end{aligned}
$$

The formulas in (1) and (2) are interesting because they do not rely upon outcomes of control units in the construction of proxies for the potential outcomes under control for treated units (see Appendix A.1 for the derivation of (2)). This is a noteworthy point that has implications for research design that we will revisit subsequently. The formula in (1) can be compared to the standard large-sample covariate adjustment for ATT:

$$
\begin{aligned}
\tau^{std}_{att} &= \mu_{1|a_1} - \mu^{std}_{0|a_1} \\
&= E[Y|a_1] - \sum_x \underbrace{E[Y|a_0, x]}_{\text{controls}} \cdot P(x|a_1). \qquad (3)
\end{aligned}
$$

Roughly speaking, standard covariate adjustment matches units that were assigned treatment to similar units that were assigned control. On the other hand, front-door estimates match units that were assigned treatment to similar units that were assigned treatment but did not receive treatment. This sort of comparison is not typical, so it is helpful to consider the informal logic of the procedure before presenting the formal statements of bias. The fundamental question is whether the treated noncompliers provide reasonable proxies for the missing counterfactuals: the outcomes that would have occurred if the treated units had not been assigned treatment. Therefore, in order for the front-door approach to be unbiased in large samples, we are effectively assuming that 1) assignment to

treatment has no effect if treatment is not received and 2) those that are assigned but don't receive treatment are comparable in some sense to those that receive treatment. This will be made more precise below.

The front-door formula in (2), with the observable proportions $P(x|a_1)$ and $P(m_1|a_1, x)$ multiplying the estimated effect of receiving the treatment, is helpful when considering the simplified front-door ATT bias, which can be written in terms of the same observable proportions (see Appendices A.2 and A.3 for proofs):

$$B^{fd}_{att} = \sum_x P(x|a_1)P(m_1|a_1, x) \sum_u \Bigg[E[Y|a_0, m_0, x, u] \cdot [P(u|a_1, m_1, x) - P(u|a_1, m_0, x)] \\ + \left\{ E[Y(a_0)|a_1, m_1, x, u] \frac{P(m_1|a_1, x, u)}{P(m_1|a_1, x)} - E[Y(a_0)|a_1, m_0, x, u] \cdot \frac{\frac{E[Y|a_1, m_0, x, u]}{E[Y(a_0)|a_1, m_0, x, u]} - P(m_0|a_1, x, u)}{P(m_1|a_1, x)} \right\} \cdot P(u|a_1, m_0, x) \Bigg]$$

The unobservable portion of this bias formula (i.e., everything after the $\sum_u$), can be difficult to interpret, but there are a number of assumptions that allow us to simplify the formula. For example, we might assume that treatment does not have an effect on the outcome for noncompliers.

Assumption 2 (Exclusion restriction)

No direct effect for noncompliers: $E[Y|a_1, m_0, x, u] = E[Y(a_0)|a_1, m_0, x, u]$.

When combined with the consistency assumption, Assumption 2 can also be written as $E[Y(a_1)|a_1, m_0, x, u] = E[Y(a_0)|a_1, m_0, x, u]$. If this exclusion restriction holds, then the bias simplifies to the following:

$$B^{fd}_{att} = \sum_x P(x|a_1)P(m_1|a_1, x) \sum_u \Bigg[E[Y|a_0, m_0, x, u] \cdot [P(u|a_1, m_1, x) - P(u|a_1, m_0, x)] \\ + \left\{ E[Y(a_0)|a_1, m_1, x, u] \frac{P(m_1|a_1, x, u)}{P(m_1|a_1, x)} - E[Y(a_0)|a_1, m_0, x, u] \cdot \frac{P(m_1|a_1, x, u)}{P(m_1|a_1, x)} \right\} \cdot P(u|a_1, m_0, x) \Bigg]$$

If instead we assume that compliance rates are constant across levels of u within levels of x,

Assumption 3 (Constant compliance rates across values of u within levels of x)

$P(m_1|a_1, x, u) = P(m_1|a_1, x)$ *for all* x *and* u,

then due to the binary measure of treatment received, we know that $P(u|a_1, m_1, x) = P(u|a_1, m_0, x)$ (see Appendix A.4), and the bias simplifies to the following:

$$B_{att}^{fd} = \sum_x P(x|a_1)P(m_1|a_1, x) \sum_u \Bigg[\Bigg\{ E[Y(a_0)|a_1, m_1, x, u] - E[Y(a_0)|a_1, m_0, x, u] \cdot \frac{\frac{E[Y|a_1,m_0,x,u]}{E[Y(a_0)|a_1,m_0,x,u]} - P(m_0|a_1, x, u)}{P(m_1|a_1, x)} \Bigg\} \cdot P(u|a_1, m_0, x) \Bigg]$$

Assumption 3 can be strengthened and the bias simplified further in some cases of clustered treatment assignment. Because the front-door estimator uses only treated units under Assumption 1, it is possible that all units within levels of x were assigned in clusters such that U is actually measured at the cluster level. We present an example of this in the early voting application, where treatment (the availability of early in-person voting) is assigned at the state level, and therefore all units within a state (e.g., Florida) have the same value of u. Formally, the assumption can be stated as the following:

Assumption 4 (u is constant among treated units within levels of x)
For any two units with a_1 and covariate values (x, u) and (x', u'), $x = x' \Rightarrow u = u'$.

When Assumption 4 holds, the u notation is redundant, and can be removed from the bias formula which simplifies as the following:

$$B_{att}^{fd} = \sum_x P(x|a_1)P(m_1|a_1, x) \Bigg\{ E[Y(a_0)|a_1, m_1, x] - E[Y(a_0)|a_1, m_0, x] \cdot \frac{\frac{E[Y|a_1,m_0,x]}{E[Y(a_0)|a_1,m_0,x]} - P(m_0|a_1, x)}{P(m_1|a_1, x)} \Bigg\} \tag{4}$$

Finally, it can be instructive to consider the formula when both Assumption 2 and Assumption 4 hold. In this scenario, the remaining bias is due to an unmeasured common cause of compliance

and the outcome.

$$B_{att}^{fd} = \sum_{x} P(x|a_1)P(m_1|a_1, x)\{E[Y(a_0)|a_1, m_1, x] - E[Y(a_0)|a_1, m_0, x]\}$$

In some applications, the bias B_{att}^{fd} may be small enough for the front-door estimator to provide a viable approach. For others, we may want to remove the bias. In the next section, we discuss a difference-in-differences approach to removing the bias.

3 Front-door Difference-in-Differences Estimators

If we define the front-door estimator within levels of a covariate x as $\tau_{att,x}^{fd}$, then the front-door estimator can be written as a weighted average of strata-specific front-door estimators where the weights are relative strata sizes for treated units:

$$\tau_{att}^{fd} = \sum_{x} P(x|a_1)\tau_{att,x}^{fd}.$$

If we further define the group of interest as the stratum g_1 and the differencing group as the stratum g_2, and we maintain Assumption 1 (one-sided noncompliance), then the front-door estimators within levels of x for these groups can be written as:

$$\tau_{att,x,g_1}^{fd} = P(m_1|x, a_1, g_1)\{E[Y|a_1, m_1, x, g_1] - E[Y|a_1, m_0, x, g_1]\},$$
$$\tau_{att,x,g_2}^{fd} = P(m_1|x, a_1, g_2)\{E[Y|a_1, m_1, x, g_2] - E[Y|a_1, m_0, x, g_2]\}.$$

Assumptions 2-4 are not needed, but can simplify interpretation (as discussed below). Using these components, the front-door difference-in-differences estimator can be written as

$$\tau^{fd-did}_{att,g_1} = \sum_x P(x|a_1,g_1)\big[\tau^{fd}_{att,x,g_1} - \frac{P(m_1|a_1,x,g_1)}{P(m_1|a_1,x,g_2)}\tau^{fd}_{att,x,g_2}\big] \tag{5}$$

$$= \sum_x P(x|a_1,g_1)P(m_1|x,a_1,g_1)\big[\{E[Y|a_1,m_1,x,g_1] - E[Y|a_1,m_0,x,g_1]\}$$

$$- \{E[Y|a_1,m_1,x,g_2] - E[Y|a_1,m_0,x,g_2]\}\big]. \tag{6}$$

Hence, (5) shows that within levels of x, the front-door difference-in-differences estimator for the group of interest is the difference between the front-door estimator from the group of interest and a scaled front-door estimator from the differencing group, where the scaling factor is the ratio of the compliance rates in the two groups. Then, the overall front-door difference-in-differences estimator is a weighted average of the estimators within levels of x, where the weights are determined by the group of interest proportions of x for treated units. Intuitively, the scaling factor is necessary because it places the front-door estimate for the differencing group on the same compliance scale as the front-door estimate for the group of interest. The necessity of this adjustment can be most easily seen in (6), where we see that the main goal is to remove the bias from the $\{E[Y|a_1,m_1,x,g_1] - E[Y|a_1,m_0,x,g_1]\}$ component of group 1 with the $\{E[Y|a_1,m_1,x,g_2] - E[Y|a_1,m_0,x,g_2]\}$ component of group 2 (i.e. remove bias from the "mediator effect").

In order for the front-door difference-in-differences estimator to remove the large sample bias from the front-door estimator of the ATT for the group of interest, we will need the following assumption to hold (where we denote bias within levels of x for the interest group g_1 as B^{fd}_{att,x,g_1}):

Assumption 5 (Bias for g_1 equal to scaled front-door formula for g_2 within levels of x)
$B^{fd}_{att,x,g_1} = \frac{P(m_1|a_1,x,g_1)}{P(m_1|a_1,x,g_2)}\tau^{fd}_{att,x,g_2}$ *for all x.*

There are two things to note about Assumption 5. First, when using an over-time approach, the compliance rates of the two groups will be equal ($P(m_1|a_1,x,g_1) = P(m_1|a_1,x,g_2)$), because time does not alter an individuals definition as a complier. Hence, Assumption 5 simplifies to $B^{fd}_{att,x,g_1} =$

τ^{fd}_{att,x,g_2} for all x in the over-time case. Second, Assumption 5 can often be weakened if only a bound is needed. For example, if the estimated effect for the differencing group is positive, and we believe the front-door bias for the group of interest is also positive but smaller than the scaled estimated effect for the differencing group, then subtracting the scaled estimated effect from the differencing group will remove too much from the estimated effect in the group of interest. Hence the front-door difference-in-differences approach will produce a lower bound.

Finally, if we believe that the front-door estimator and front-door difference-in-differences estimator have bias of different signs, then these can be used in a bracketing approach. For example, if we believe the bias in the front-door estimator is positive prior to the differencing, and we believe the bias of the front-door difference-in-differences estimator is negative, then the front-door and front-door difference-in-differences estimator can be used together to bracket the truth in large samples. This will be discussed in the context of the illustrative applications from the following sections.

If Assumptions 1 and 5 hold, then τ^{fd-did}_{att} has no large-sample bias for τ_{att} (see Appendix B.1 for a proof). However, the interpretation of Assumption 5 will often be simplified when Assumptions 2, 3, or 4 hold. This will be discussed in the context of the applications, but one special case is useful to consider for illustrative purposes. When Assumptions 1 through 4 hold, then Assumption 5 is equivalent to the following:

$$\{E[Y(a_0)|a_1, m_1, x, g_1] - E[Y(a_0)|a_1, m_0, x, g_1]\} = \{E[Y(a_0)|a_1, m_1, x, g_2] - E[Y(a_0)|a_1, m_0, x, g_2]\}$$

Note that this equality is analogous to the parallel trends assumption for standard difference-in-differences estimators.

4 Illustrative Application: National JTPA Study

We now illustrate how front-door and front-door difference-in-differences estimates for the average treatment effect on the treated (ATT) can be used to estimate and bracket the experimental truth in the context of the National JTPA Study, a job training evaluation with both experimental data and a nonexperimental comparison group. We measure program impact as the ATT on 18-month earnings in the post-randomization or post-eligibility period, where active treatment is assignment into the program (perhaps self-selected assignment).[7] We focus on the effect of sign-up on earnings for three reasons: 1) we can compare front-door estimates to the experimental benchmark, 2) this effect is the same parameter of interest as in much of the econometrics literature utilizing JTPA data (Heckman, Ichimura and Todd, 1997; Heckman and Smith, 1999), and 3) this is often the policy-relevant causal effect when considering whether or not to extend the opportunity for job training. Furthermore, (Heckman et al., 1998) showed that for the National JTPA Study, matching

[7]The Department of Labor implemented the National JTPA Study between November 1987 and September 1989 in order to gauge the efficacy of the Job Training Partnership Act (JTPA) of 1982. The Study randomized JTPA applicants into treatment and control groups at 16 study sites (referred to as service delivery areas, or SDAs) across the United States. Participants randomized into the treatment group were allowed to receive JTPA services, whereas those in the control group were prevented from receiving program services for an 18-month period following random assignment (Bloom et al., 1993; Orr et al., 1994). Crucially for our analysis, 57.3% of adult males and 61.4% of married adult men allowed to receive JTPA services actually utilized at least one of those services. Moreover, the Study also collected a nonexperimental comparison group of individuals who met JTPA eligibility criteria but chose not to apply to the program in the first place. See Appendix C for additional information regarding the ENP sample. See Smith (1994) for details of ENP screening process. Since this sample of eligible nonparticipants (ENPs) was limited to 4 service delivery areas, we restrict our entire analysis to only these 4 sites.

adjustments using the nonexperimental comparison group can come close to the experimental estimates only when one has, "detailed retrospective questions on labor force participation, job spells, earnings." In the following, we discuss the use of front-door difference-in-differences estimators to provide similar information in the absence of detailed labor force histories.

As discussed below, the simple front-door estimator is anticipated to exhibit positive bias when estimating ATT for the JTPA program for adult males. In the following subsections, we consider two front-door DD approaches to correcting this bias. First, we consider using an over-time approach to remove positive bias from the front-door estimator. Second, we consider the more conservative approach of using single adult males as a differencing group, which allows us to provide a lower bound on the effect of the program from married adult males. Because the front-door estimator provides an upper bound, these two estimators can be used in a bracketing approach.

4.1 Results: Over-Time Differencing

The most simple front-door estimator for the effects of the JTPA program takes the mean 18-month earnings of those that both signed up for the program and showed up for their training and subtracts the mean 18-month earnings of those that signed up for the program but failed showed up for their training, and then scales this estimate by the rate at which those that signed up actually showed up. Because we have not used covariates, this estimator can be written as a simplified version of (2):

$$\tau_{att}^{fd} = P(m_1|a_1) \cdot \left\{ \underbrace{E[Y|a_1, m_1] - E[Y|a_1, m_0]}_{\text{"effect" of receiving treatment}} \right\},$$

where a_1 indicates signing up for the program, m_1 indicates showing up for the program, m_0 indicates failing to show up for the program, and Y denotes 18-month earnings. Because those that show up are likely to be more diligent/disciplined than those that fail to show up, we expect this estimator to be positively biased.

In an attempt to remove the anticipated positive bias, we can use the baseline earnings of these individuals as a differencing group. The most simple version of this estimator does the following: a) takes the mean 18-month earnings of those that both signed up for the program and showed up for their training and subtracts the mean 18-month earnings of those that signed up for the program but failed showed up for their training, b) takes the mean baseline (i.e., 0-month) earnings of those that both signed up for the program and showed up for their training and subtracts the mean baseline earnings of those that signed up for the program but failed to show up for their training, c) takes the difference between these two estimates, and d) scales this difference by the proportion that showed up among those that signed up. As above, because we have not used covariates, this estimator can be written as a simplified version of (6):

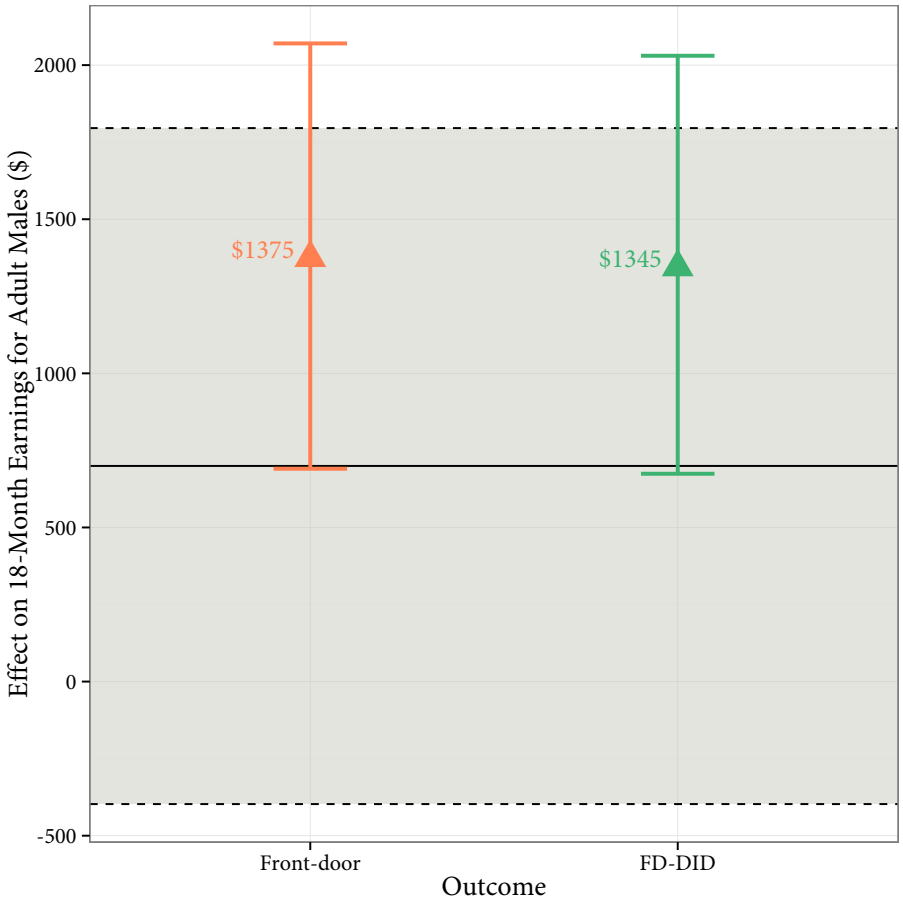
$$\begin{aligned}\tau_{att,g_1}^{fd-did} &= P(m_1|a_1,g_1)\big[\{E[Y|a_1,m_1,x,g_1] - E[Y|a_1,m_0,x,g_1]\} \\ &\quad - \{E[Y|a_1,m_1,x,g_2] - E[Y|a_1,m_0,x,g_2]\}\big],\end{aligned}$$

where a_1 indicates signing up for the program, m_1 indicates showing up for the program, m_0 indicates failing to show up for the program, g_1 indicates a post-treatment measurement (i.e., at 18 months), g_2 indicates a baseline measurement (i.e., at 0 months), and Y now can denote either 18-month or 0-month earnings, depending on whether g_1 or g_2 is in the conditioning set.

The front-door and front-door difference-in-differences estimates for the effect of the JTPA program on adult males are presented in Figure 2. The experimental benchmark (solid black line) is the only estimate that uses the experimental control units. Note that while the front-door estimator appears to exhibit some of the anticipated positive bias, the estimate lies within the 95% confidence interval from the experiment. The front-door DD estimator gets a bit closer to the experimental benchmark and its 95% interval more clearly covers the benchmark.

Although the improvement from the front-door DD estimate is minimal here, this may be due to the relatively good quality of the front-door estimate. If we didn't see the experimental results (as

Figure 2: Comparison of front-door and over-time front-door difference-in-differences estimates for the JTPA effect for adult males. The solid line is the experimental benchmark and the dashed lines represent the confidence interval. All intervals are 95% bootstrapped confidence intervals based on 10,000 replicates.



would be true for non-illustrative applications), the similarity between the front-door and front-door DD estimates would give us some confidence as to the robustness of the findings (and this confidence would not be misplaced for this example). However, if even after seeing these results we prefer a more conservative estimate of the effect of sign-up, we can define a different differencing group using the observed covariates.

4.2 Results: Single Males as a Differencing Group

If we didn't have the experimental benchmark, we might not be confident that the bias in the pre-treatment period is equal to the bias in the post-treatment period, and hence we may want to use an additional differencing group as a robustness strategy. In this subsection, we discuss the use of never married men (henceforth referred to as simply single men) as the differencing group and currently or once married adult men as the group of interest (henceforth referred to as simply married men).[8]

The use of differencing group that is a subset of the individuals (single men) adds an additional complication to the analysis. We must consider whether the effect of interest is the average effect of the program for all individuals or just the average over the individuals in the group of interest. Fortunately, conversion between the two effects is straightforward due to the assumption that the effect of the program is zero for the differencing group. Specifically, the average effect over all individuals is the average effect for the group of interest times the proportion of individuals in the group of interest. In order to simplify the presentation and because this conversion is straightforward, we continue this section focusing on the effect for the group of interest instead of for all individuals. All of the following results are substantively replicated when we convert to the analysis for all individuals.

With single men as the differencing group, we include baseline earnings as a covariate, which further complicates the analysis, and rules out the use of the simplified versions of (2) and (6) from the previous Subsection 4.1. However, the use of covariates in the analysis also allows us to compare the performance of the front-door and front-door difference-in-differences estimators to standard covariate adjustments like regression and matching.

The front-door and front-door difference-in-differences estimates for the effect of the JTPA program on married males - our group of interest - are presented in Figure 3 across a range of covari-

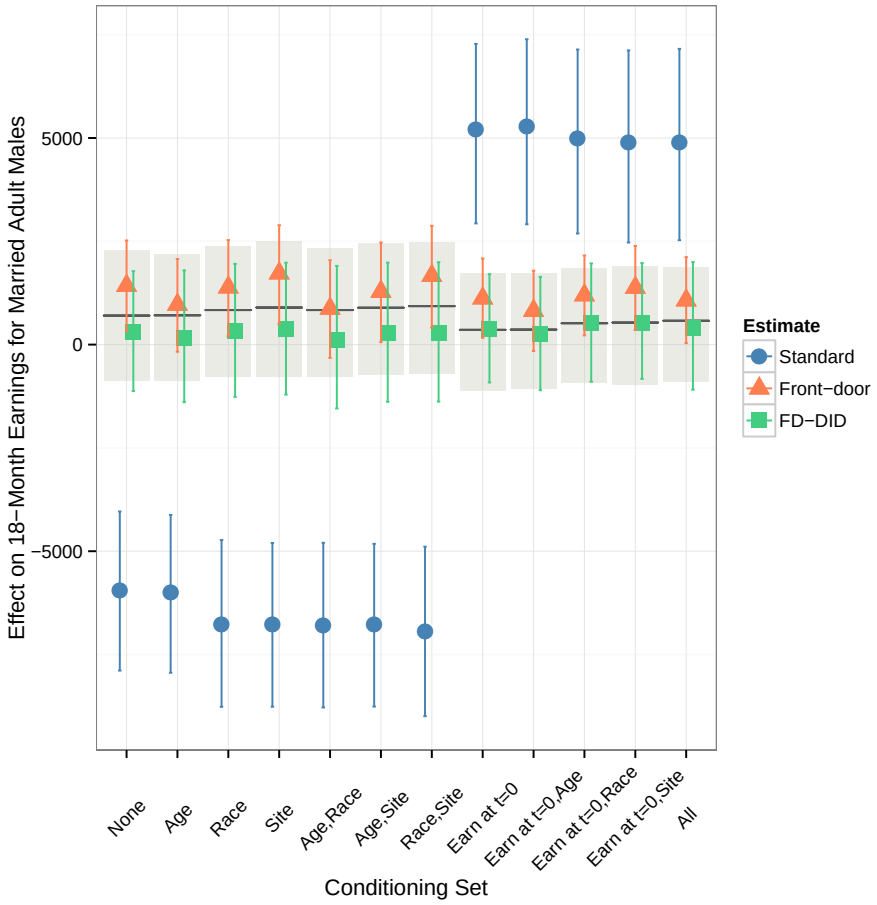
[8]Age for adult men ranges from 22 to 54 at random assignment / eligibility screening. Once married men comprises individuals who report that they are widowed, divorced, or separated.

ate sets. Additionally, we present the standard covariate adjusted estimates for comparison. We use OLS separately within experimental treated and observational control groups (the ENPs) for the standard estimates. For front-door estimates, we use OLS separately within the "experimental treated and received treatment" and "experimental treated and didn't receive treatment" groups. Therefore, these estimates assume linearity and additivity within these comparison groups when conditioning on covariates, albeit we note that we obtain similar results when using more flexible methods that relax these parametric assumptions. The experimental benchmark (dashed line), is the only estimate that uses the experimental control units.

First, note that the front-door estimates exhibit uniformly less estimation error than estimates from standard covariate adjustments across all conditioning sets in Figure 3. The error in the standard estimates for the null conditioning set and conditioning sets that are combinations of age, race, and site are negative. The error becomes positive when we include baseline earnings in the conditioning set. In sharp contrast, the stability of front-door estimates is remarkable. We thus find that front-door estimates are preferable to standard covariate adjustment when more detailed information on labor force participation and historic earnings is not available.

In spite of the superior performance of front-door estimates compared to standard covariate adjustment, the front-door estimates are slightly above the experimental benchmark across all covariate sets. As mentioned above, without seeing the experimental benchmark, we might believe these estimates are affected by positive bias because those that fail to show up to the job training program are likely to be less diligent individuals than those that show up. Given the anticipated positive bias in the front-door estimates, we use the front-door difference-in-differences estimator to either recover an unbiased point estimate or obtain a lower bound, depending on our assumptions as to the effect of the program in the differencing group. If we believe that the JTPA program had no effect for single males, and we also believe that Assumptions 1 and 5 hold, then the difference-in-differences estimator will return an unbiased estimate of the effect for the group of interest in large samples. If, on the other hand, we believe there might be a non-negative effect for single males,

Figure 3: Comparison of standard covariate adjusted estimates, front-door, and front-door difference-in-differences estimates for the JTPA effect for married adult males. The dashed line is the experimental benchmark. 95% bootstrapped confidence intervals are based on 10,000 replicates.



then we would obtain a lower bound for the effect for the group of interest. In this application, it is more likely that there was a positive effect of the JTPA program for single males, albeit one smaller than for married males. Hence, the front-door difference-in-differences estimator will likely give us a lower bound for the effect of the JTPA program for married males. In fact, in many applications we may be unable to find a differencing group with no effect, yet still be able to use front-door and front-door difference-in-differences approaches to bound the causal effect of interest given our beliefs about the sign and relative scale of effects in the group of interest and the differencing group.

When examining the empty conditioning set, the front-door estimate that we obtain for sin-

gle males is \$946.09. In order to construct the front-door difference-in-differences estimator, we have to scale this estimate by the ratio of compliance for married males to compliance for single males, which is equal to $0.614/0.524 \approx 1.172$. Subtracting the scaled front-door estimate for single males from the front-door estimate for married males as shown in (5), we obtain an estimate of \$315.41. This is slightly below the experimental benchmark and thus indeed functions as lower bound. In sharp contrast to the front-door and front-door difference-in-differences estimates that rather tightly bound the truth, the bias in the standard estimate is -\$6661.90. It is noteworthy that the front-door estimate acts as an upper bound and the front-door difference-in-differences estimate acts as a lower bound across all conditioning sets presented in Figure 3.

5 Illustrative Application: Early Voting

In this section, we present front-door difference-in-differences estimates for the average treatment effect on the treated (ATT) of an early in-person voting program in Florida. We want to evaluate the impact that the presence of early voting had on turnout for some groups in the 2008 and 2012 presidential elections in Florida. In traditional regression or matching approaches (either cross sectional or difference-in-differences), data from Florida would be compared to data from states that did not implement early in-person voting. These approaches are potentially problematic because there may be unmeasured differences between the states, and these differences may change across elections. One observable manifestation of this is that the candidates on the ballot will be different for different states in the same election year and for different election years in the same state. The front-door and front-door difference-in-differences approaches allows us to solve this problem by confining analysis to comparisons made amongst modes of voting within a single presidential election in Florida.

Additionally, by restricting our analysis to Florida, we are able to use individual-level data from the Florida Voter Registration Statewide database, maintained since January 2006 by the Florida

Department of State's Division of Elections. This allows us to avoid the use of self-reported turnout, provides a very large sample size, and makes it possible to implement all of the estimators discussed in earlier sections because we observe the mode of voting for each individual. The data contains two types of records by county: registration records of voters contained within *voter extract files* and voter history records contained in *voter history files*. The former contains demographic information - including, crucially for this paper, race - while the latter details the voting mode used by voters in a given election. The two records can be merged using a unique voter ID available in both file types. However, voter extract files are snapshots of voter registration records, meaning that a given voter extract file will not contain all individuals appearing in corresponding voter history file because individuals move in and out of the voter registration database. We therefore use voter registration files from four time periods to match our elections of interest: 2006, 2008, and 2010 book closing records, and the 2012 post-election registration record. Our total population, based on the total unique voter IDs that appear in any of the voter registration files, is 16.4 million individuals. Appendix D provides additional information regarding the pre-processing of the Florida data.

Information on mode of voting in the voter history files allows us to define compliance with the program for the front-door estimator (i.e., those that utilize EIP voting in the election for which we are calculating the effect are defined as compliers). Additionally, we use information on previous mode of voting to partition the population into a group of interest and differencing groups. In order to maximize data reliability, we define our group of interest as individuals that used EIP in a previous election (e.g., 2008 EIP voters are the group of interest when analyzing the turnout effect for the 2012 election). In other words, we are assessing what would have happened to these 2008 EIP voters in 2012 if the EIP program had not been available in 2012. To calculate the EIP effect on turnout for the 2012 election, we separately consider 2008 and 2010 EIP voters as our groups of interest. For the 2008 EIP effect on turnout, we rely upon 2006 EIP voters as our group of interest. An attempt to define the group of interest more broadly (e.g., including non-voters) or in terms of earlier elections (e.g., the 2004 election) would involve the use of less reliable data, and would therefore introduce

methodological complications that are not pertinent to the illustration presented here.[9] Therefore, the estimates presented in this application are confined only to those individuals that utilized EIP in a previous election and hence we cannot comment on the overall turnout effect.

We consider two differencing groups for each analysis: those who voted absentee and those that voted on election day in a previous election. When considering the 2012 EIP effect for 2008 EIP voters, for example, we use 2008 absentee and election day voters as our differencing groups. It is likely that the 2012 EIP program had little or no effect on 2012 turnout for 2008 absentee voters and perhaps only a minimal effect for 2008 election day voters, as these groups had already demonstrated an ability to vote by other means. For example, experimental evidence suggests that while mobilizing people to vote early increases turnout, it does not significantly alter the proportion of people that vote by mail and slightly reduces the proportion voting on election day (Mann and Mayhew, 2012). It thus seems reasonable to assume that EIP offers alternative, not additional, opportunities for voting to past absentee and election day voters. In this case, any apparent effects on turnout estimated for these groups will be primarily due to bias, and this bias can then be removed from the estimates for the group of interest. If in fact, these apparent effects represent real effects for these

[9]Following Gronke and Stewart (2013), we restrict our analysis to data starting in 2006 due to its greater reliability than data from 2004. We also might like to extend the group of interest to those that did not vote in a previous election, but we avoid assessing either 2008 or 2012 EIP effects for these voters because it is difficult to calculate the eligible electorate and consequently the population of non-voters. In their analysis of the prevalence of early voting, Gronke and Stewart (2013) use all voters registered for at least one general election between 2006 and 2012, inclusive, as the total eligible voter pool. However, using registration records as a proxy for the eligible electorate may be problematic (McDonald and Popkin, 2001). By focusing on the 2008 voting behavior of individuals who voted early in 2006, we avoid the need to define the eligible electorate and the population of non-voters.

groups, then our results will produce a lower bound. As discussed in earlier sections, the estimates from the differencing groups must be scaled according to the level of compliance for the group of interest. Finally, the existence of two differencing groups allows us to conduct a placebo test by using election day voters as the group of interest and the absentee voters as the differencing group in each case. This analysis is explored below.

Despite the limited scope of the estimates presented here, these results have some bearing on the recent debates regarding the effects of early voting on turnout. There have been a number of papers using cross state comparisons that find null results for the effects of early voting on turnout (Gronke, Galanes-Rosenbaum and Miller, 2007; Gronke et al., 2008; Fitzgerald, 2005; Primo, Jacobmeier and Milyo, 2007; Wolfinger, Highton and Mullin, 2005), and Burden et al. (2014) finds a surprising negative effect of early voting on turnout in 2008.[10] However, identification of turnout effects from observational data using traditional statistical approaches such as regression or matching rely on the absence of unobserved confounders that affect both election laws and turnout (Hanmer, 2009). If these unobserved confounders vary across elections, then traditional difference-in-differences estimators will also be biased. See Keele and Minozzi (2013) for a discussion within the context of election laws and turnout. Additionally, a reduction in Florida's early voting program between 2008 and 2012 provided evidence that early voting may encourage voter turnout (Herron and Smith, 2014).

The front-door estimators presented here provide an alternative approach to estimating turnout effects with useful properties. First, front-door adjustment can identify the effect of EIP on turnout in spite of the endogeneity of election laws that can lead to bias when using standard approaches. Second, unlike traditional regression, matching, or difference-in-differences based estimates, the front-door estimators considered here only require data from Florida within a given year. This

[10] Burden et al. (2014) examine a broader definition of early voting that includes no excuse absentee voting.

means that we can effectively include a Florida/year fixed effect in the analysis, and we do not have to worry about cross-state or cross-time differences skewing turnout numbers across elections. We also include county fixed effects in the analysis in order to control for within-Florida differences.

However, in addition to the limited scope of our analysis, it is important to note that the exclusion restriction is likely violated for this application. Since early in-person voting decreases waiting times on election day, it is possible that it actually increases turnout among those that only consider voting on election day. This would mean that front-door estimates would understate the effect if all other assumptions held because the front-door estimator would be ignoring a positive component of the effect. Alternatively, Burden et al. (2014) suggest that campaign mobilization for election day may be inhibited, such that early voting hurts election day turnout. This would mean that front-door estimates would overstate the effect because the front-door estimator would be ignoring a negative component of the effect. This can also be seen by examining the bias formula (4) (because the EIP treatment is assigned at the state level, Assumptions 1 and 4 will hold).

Taken together, the overall effect of these exclusion restrictions is unclear and would depend on the strength of the two violations. The predictions also become less clear once we consider the front-door difference-in-differences approach, where additional bias in front-door estimates might cancel with bias in the estimates for the differencing group. For the remainder of this analysis, we will assume that all such violations of the exclusion restriction cancel out in the front-door difference-in-differences estimator. This is implicit in Assumption 5.

5.1 Results

In order to construct the front-door estimate of the 2008 EIP effect for our group of interest, we calculate the turnout rate in 2008 for all individuals who voted early in 2006. We also calculate the non-complier turnout rate in 2008 by excluding all individuals who voted early in 2008 from the previous calculation. The front-door estimate of the 2008 EIP effect for 2006 early voters is thus the difference between the former and latter turnout rates. Quite intuitively, the counterfactual turnout

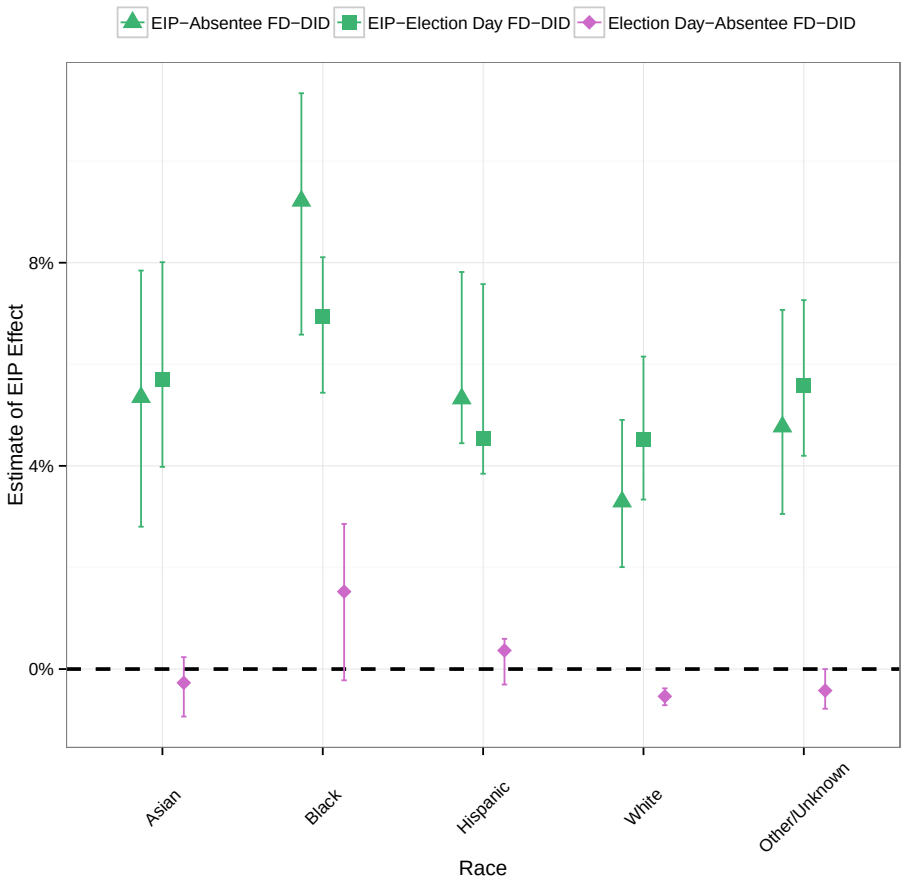
rate without EIP for the group of interest is the observed turnout rate of non-compliers in that group. We do not devote much attention to the front-door estimates seeing as they are implausibly large.[11] The positive bias stems from the fact that 2006 EIP voters would be more likely to vote in 2008, even in the absence of EIP, than the 2006 non-EIP group (this group includes individuals that did not vote in 2006). In terms of the bias formula in (4), this is equivalent to saying that $E[Y(a_0)|a_1, m_1, x] > E[Y(a_0)|a_1, m_0, x]$.

In order to address this bias, we present front-door difference-in-differences estimates for the 2008 EIP program in Figure 4. The estimates all utilize county fixed effects and are calculated separately across the racial categories.[12] The front-door difference-in-differences estimates for the group of interest (2006 EIP voters) are in green, with 2008 absentee voters (triangles) and 2008 election day voters (squares) as the differencing groups. The former, for example, is constructed as the difference between front-door estimates for 2006 early voters and the front-door estimates for 2006 absentee voters, with the front-door estimates for the differencing group scaled by the ratio of early voter compliance to absentee voter compliance as shown in (5). The purple estimates in Figure 4 represent the placebo test, with 2006 election day voters standing in as the group of interest and the absentee voters as the differencing group. In general, we note that if there exists more than one plausible differencing group, then one should conduct the analysis using each differencing group separately, as well as a placebo test to verify the plausibility of Assumption 5.

[11]Front-door estimates are available in Appendix E.

[12]We calculate the FD-DID estimates within each county and then average using the population of the group of interest as the county weight. Due to very small sample sizes in a few counties, we are occasionally unable to calculate front-door estimates. In these cases, we omit the counties from the weighted average when calculating the front-door estimates with fixed effects. We note that due to their small size, these counties are unlikely to exert any significant impact upon the estimates regardless.

Figure 4: Front-door difference-in-differences estimates for the turnout effect in 2008 for voters who voted early in 2006 (by race). All estimates include county fixed effects. 99% block bootstrapped confidence intervals are based on 10,000 replicates.



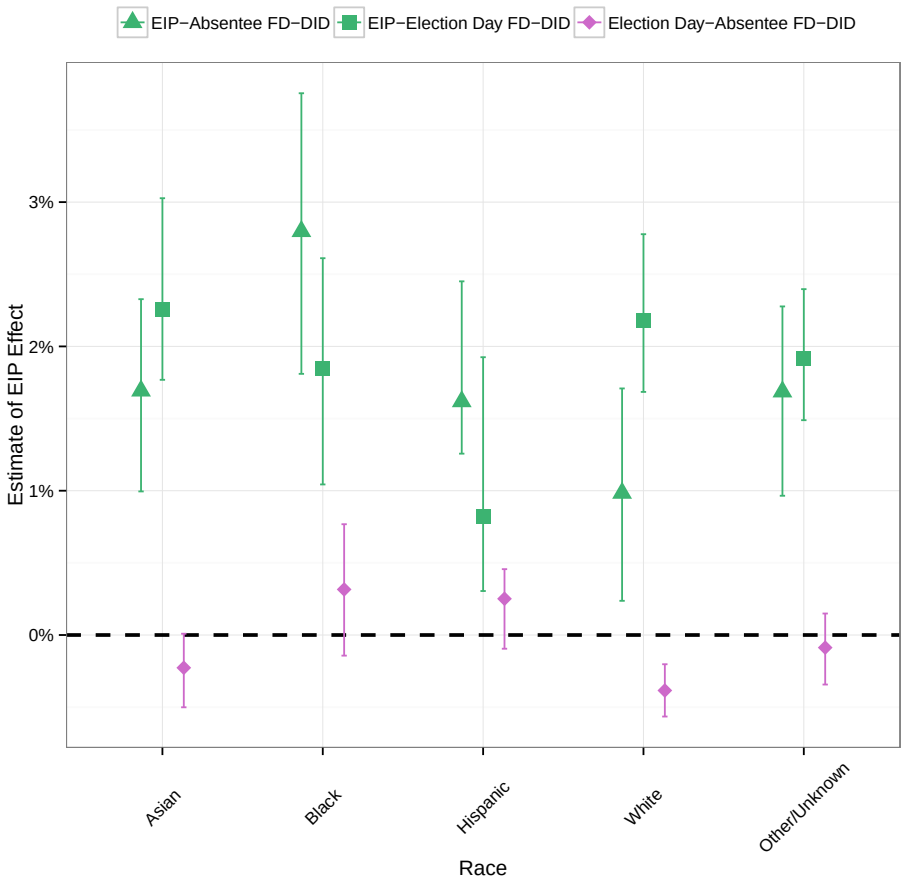
The EIP program estimates are positive and significant at the 99% level. All placebo tests, with the exception of the white estimate, are indistinguishable from zero, giving us confidence in the estimated EIP effects. Even if the slightly negative placebo estimate for whites indicates a true negative effect of the 2008 EIP program, and not bias, the weighted average of the green and the purple effects (i.e., the 2008 EIP effect for the 2006 EIP and election day voters together), again produces a slightly positive estimate. Therefore, we generally find evidence that early voting increased turnout for the subset of individuals who voted early in 2006. Moreover, comparing the point estimates across races, we find some evidence that the program had a disproportionate benefit for African-

Americans.

Our methodology uses voting behavior in 2006 only to define groups and does not compare turnout of voters across elections. Thus any differences between presidential election and midterm election voters (see e.g. Gronke and Toffey (2008)) does not pose a prima facie problem for the analysis. Moreover, using early voters in a midterm election as the group of interest for calculating the EIP effect in a presidential election does not require additional assumptions beyond what one would need if using early voters in a presidential election. Nonetheless, a potential downside of the preceeding results is that the estimated 2008 EIP is limited to those individuals who voted early in the 2006 midterm election, whereas we might want to extend the group of interest to early voters in a presidential election. Unfortunately, we cannot present the estimates with the group of interest and the differencing groups defined in terms of 2004 behavior because the data from 2004 are not reliable (as mentioned above). As a robustness check, we also estimate the effect of the early voting program in the 2012 election, for which we can define the group of interest and the differencing groups using 2008 voting behavior. However, as discussed above, Florida's early voting program was reduced between 2008 and 2012, so we should not expect the results to be equivalent.

The results of this analysis are presented in Figure 5. For the 2008 EIP voters, the 2012 EIP front-door difference-in-differences estimates (green) are positive and significant at the 99% level (based on 10,000 block bootstraps at the county level). There is some evidence of differences between the racial categories, but these differences change depending on which differencing group is used. The purple estimates are for the most part indistinguishable from zero, indicating that the placebo tests have mostly been passed. The slightly negative purple estimate for whites again indicates either bias, or perhaps a negative effect of the 2012 EIP program for white 2008 election day voters. Note that even if we believe this estimate, the weighted average of the green and the purple effects for whites (i.e., the 2012 EIP effect for the 2008 EIP and election day voters together) produces a slightly positive estimate, albeit this estimate is indistinguishable from zero. In sum, the evidence points to a slightly positive turnout effect of the 2012 EIP program on the 2008 EIP users.

Figure 5: Front-door difference-in-differences estimates for the turnout effect in 2012 for voters who voted early in 2008 (by race). All estimates include county fixed effects. 99% block bootstrapped confidence intervals are based on 10,000 replicates.



It is also notable that the size of the estimated EIP effect for 2012 is less than half the estimated EIP effect for 2008 when looking at EIP voters as the group of interest across all races. There are two potential reasons for this. First, our estimates for the 2008 EIP program are obtained using groups defined by 2006 midterm election behavior and as already mentioned, midterm election early voters are likely different than presidential election early voters. Second, the nature of the early voting program changed between the 2008 and 2012 elections, notably removing the option of voting early on the Sunday prior to the election and all-together near-halving of the early voting period from 14 days to 8 days (Gronke and Stewart, 2013; Herron and Smith, 2014). This change might possibly

reduce the effect of the EIP program in 2012 when compared to 2008 - a finding consistent with the conclusion made by Herron and Smith (2014) that individuals who voted in 2008 on the Sunday prior to the election were disproportionately less likely to vote in 2012.

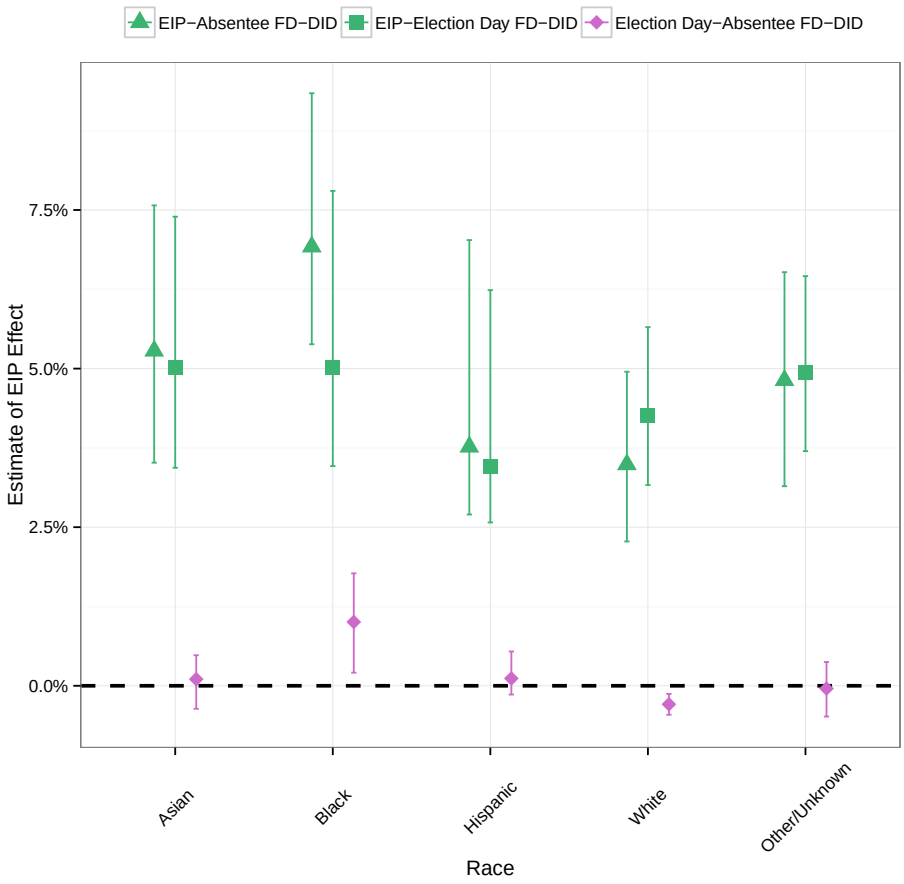
In order to isolate the consequences of the change in the early voting program from changes in the construction of the group of interest and differencing groups, we re-estimate the effects of the 2012 EIP program using 2010 EIP voters as the group of interest (green), and using 2010 absentee (triangles) and election day voters (squares) as the differencing groups. Placebo tests are reported using 2010 election day voters as the group of interest and 2010 absentee voters as the differencing group (purple). These results are presented in Figure 6, and they are quite similar to the results in Figure 4. This provides some evidence that if we were able to obtain reliable data from the 2004 election, our estimates for the 2008 EIP program would likely have produced something similar to Figure 5 when using 2004 EIP voters as the group of interest, and 2004 absentee (triangles) and election day voters (squares) as the differencing groups. However, the estimates in Figure 4 are slightly larger than estimates in Figure 6. This is consistent with the reduction in the early voting window for the 2012 election.

6 Conclusion

In this paper, we have developed front-door difference-in-differences estimators for nonrandomized program evaluations with one-sided noncompliance and an exclusion restriction. These estimators allow for asymptotically unbiased estimation via front-door techniques, even when front-door estimators have significant bias. Even when the assumptions do not hold exactly, these estimators sometimes allow for informative bounds.

We illustrated this technique with an application to the National JTPA (Job Training Partnership Act) Study and with an application to the effects of Florida's early in-person voting program on turnout. For the job training application, we showed that front-door and front-door difference-in-

Figure 6: Front-door difference-in-differences estimates for the turnout effect in 2012 for voters who voted early in 2010 (by race). All estimates include county fixed effects. 99% block bootstrapped confidence intervals are based on 10,000 replicates.



differences could be used to recover the experimental benchmark. For the application to the effects of an early in-person (EIP) voting program on turnout in Florida in 2008 and 2012, we found that for two separate differencing groups, the program had at least small but significant positive effects. While the scope of the analysis is limited, this result provides some evidence to counter previous results in the literature that early voting programs had either no effect or negative effects.

While these two applications demonstrate the efficacy of the technique, they also demonstrate the care with which it should be used. In both applications, it was possible that some of the assumptions would not hold, and we were forced to argue about the potential direction of the bias

(although we note that the assumptions may hold for the over-time DD in the JTPA example). As we showed, this did not prevent us from partially answering the research questions in either case. In particular, the bracketing approach, exemplified by the JTPA study, provides operationally similar (albeit more conservative) information to the case where the assumptions hold perfectly.

Additionally, these applications demonstrate two broad classes of applications where this approach might be helpful. The JTPA study shows an example where treatment (signing up) is more susceptible to confounding than compliance (showing up). In applications of this type, bias from the front-door approach may be minimal, and the front-door difference-in-differences approach may only be necessary as a robustness check (over-time approach), or to bracket the truth (married versus single male approach). The early voting study shows an example where both the treatment and compliance may be susceptible to large amounts of confounding so that front-door estimators may be as highly biased as standard estimators. However, the absentee voters provide a ready made differencing group, and after differencing, the front-door difference-in-differences estimator can function at least as a robustness check.

Finally, the results in this paper have implications for research design and analysis. First, the examples demonstrate the importance of collecting post-treatment variables that represent compliance with, or uptake of, the treatment. Such information allows front-door and front-door difference-in-differences analyses to be carried out as a robustness check on standard approaches. Second, the bracketing of the experimental benchmark in the JTPA application show that control units are not always necessary for credible causal inference. This is a remarkable finding that should make a number of previously infeasible studies possible (e.g., when it is unethical or impossible to withhold treatment from individuals).

References

Angrist, Joshua D. and Jörn Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. 4, 6

Bloom, Howard S., Larry L. Orr, George Cave, Stephen Bell and Fred Doolittle. 1993. "The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months." Bethesda, MD: . 15

Brady, Henry E., David Collier and Jason Seawright. 2006. "Toward a Pluralistic Vision of Methodology." *Political Analysis* 14:353–368. 2

Burden, Barry C., David T. Canon, Kenneth R. Mayer and Donald P. Moynihan. 2014. "Election Laws, Mobilization, and Turnout: The Unanticipated Consequences of Election Reform." *American Journal of Political Science* 58(1):95–109. 25, 26

Collier, David and Henry E. Brady. 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield. 2

Cox, David R. and Nanny Wermuth. 1995. "Discussion of 'Causal diagrams for empirical research.'" *Biometrika* 82:688–689. 2

Fitzgerald, Mary. 2005. "Greater Convenience but not Greater Turnout: The Impact of Alternative Voting Methods on Electoral Participation in the United States." *American Politics Research* 33:842–867. 25

George, A.L. and A. Bennett. 2005. *Case studies and theory development in the social sciences*. Mit Press. 2

Glynn, Adam and Kevin Quinn. 2011. "Why Process Matters for Causal Inference." *Political Analysis* 19(3):273–286. 3

Gronke, Paul and Charles Stewart. 2013. "Early Voting in Florida." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, IL. 24, 30, 43

Gronke, Paul and Daniel Krantz Toffey. 2008. "The Psychological and Institutional Determinants of Early Voting." *Journal of Social Issues* 64(3):503–524. 29

Gronke, Paul, Eva Galanes-Rosenbaum, Peter A. Miller and Daniel Toffey. 2008. "Convenience Voting." *Annual Review of Political Science* 11:437–455. 25

Gronke, Paul, Eva Galanes-Rosenbaum and Peter Miller. 2007. "Early Voting and Turnout." *PS: Political Science and Politics* XL. 25

Hanmer, Michael J. 2009. *Discount Voting: Voter Registration Reforms and Their Effects*. Cambridge University Press. 25

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd. 1998. "Characterizing selection bias using experimental data." *Econometrica* 66:1017–1098. 8, 15, 42

Heckman, James, Hidehiko Ichimura and Petra Todd. 1997. "Matching as an econometric evaluation estimator evidence from evaluating a job training program." *Review of Economic Studies* 64:605–654. 15

Heckman, James J. and Jeffrey A. Smith. 1999. "The pre-programme earnings dip and the determinants of participation in a social programme: implications for simple programme evaluation strategies." *Economic Journal* . 15

Heckman, James J., Robert J. LaLonde and Jeffrey A. Smith. 1999. The Economics and Econometrics of Active Labor Market Programs. In *Handbook of Labor Economics, Volume III*, ed. O. Ashenfelter and D. Card. Elsevier Science North-Holland. 8

Herron, Michael C. and Daniel A. Smith. 2014. "Race, Party, and the Consequences of Restricting Early Voting in Florida in the 2012 General Election." *Political Research Quarterly* . 25, 30, 31

Imbens, Guido and Donald Rubin. 1995. "Discussion of 'Causal diagrams for empirical research.'" *Biometrika* 82:694–695. 3

Joffe, Marshall M. 2001. "Using information on realized effects to determine prospective causal effects." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* pp. 759–774. 3

Kaufman, Sol, Jay S. Kaufman and Richard F. MacLehose. 2009. "Analytic bounds on causal risk differences in directed acyclic graphs with three observed binary variables." *Journal of Statistical Planning and Inference* 139:3473–87. 3

Keele, Luke and William Minozzi. 2013. "How Much Is Minnesota Like Wisconsin? Assumptions and Counterfactuals in Causal Inference with Observational Data." *Political Analysis* 21(2):193–216. 25

King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. 1 ed. Princeton University Press. 2

Kuroki, Manabu and Masami Miyakawa. 1999. "Identifiability Criteria for Causal Effects of Joint Interventions." *J. Japan Statist. Soc.* 29(2):105–117. 2

Mann, Christopher B. and Genevieve Mayhew. 2012. "Multiple Voting Methods, Multiple Mobilization Opportunities? Voting Behavior, Institutional Reform, and Mobilization Strategy." Paper presented at the Annual Meeting of the Southern Political Science Association, New Orleans, LA. 24

McDonald, Michael P. and Samuel L. Popkin. 2001. "The Myth of the Vanishing Voter." *American Political Science Review* 95:963–974. 24

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Winston Lin, George Cave and Fred Doolittle. 1994. "The National JTPA Study: Impacts, Benefits, And Costs of Title IIA." Bethesda, MD: . 15

Pearl, Judea. 1995. "Causal diagrams for empirical research." *Biometrika* 82:669–710. 2, 7

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2 ed. Cambridge University Press. 2

Primo, David M., Matthew L. Jacobmeier and Jeffrey Milyo. 2007. "Estimating the Impact of State Policies and Institutions with Mixed-Level Data." *State Politics & Policy Quarterly* 7:446–459. 25

Shpitser, Ilya and Judea Pearl. 2006. "Identification of Conditional Interventional Distributions." Proceedings of the Twenty Second Conference on Uncertainty in Artificial Intelligence (UAI). 2

Smith, Jeffrey A. 1994. "Sampling Frame for the Eligible Non-Participant Sample." *Mimeo* . 15

Stewart, Charles. 2012. "Declaration of Dr. Charles Stewart III." State of Florida vs. United States of America. 43

Tian, Jin and Judea Pearl. 2002*a*. A general identification condition for causal effects. In *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999 pp. 567–573. 2

Tian, Jin and Judea Pearl. 2002*b*. On the identification of causal effects. In *Proceedings of the American Association of Artificial Intelligence*. 2

VanderWeele, Tyler J. 2009. "On the relative nature of overadjustment and unnecessary adjustment." *Epidemiology* 20(4):496–499. 2

Wolfinger, Raymond E., Benjamin Highton and Megan Mullin. 2005. "How Postregistration Laws Affect the Turnout of Citizens Registered to Vote." *State Politics & Policy Quarterly* 5:1–23. 25

A ATT Proofs

A.1 Front-door Adjustment with One-Sided Noncompliance

In the special case of one-sided noncompliance, the front-door estimator can be written as the following:

$$
\begin{aligned}
\tau_{att}^{fd} &= E[Y|a_1] - \sum_x E[Y|a_1, m_0, x] \cdot P(x|a_1) \\
&= \sum_x E[Y|a_1, x] \cdot P(x|a_1) - \sum_x E[Y|a_1, m_0, x] \cdot P(x|a_1) \\
&= \sum_x P(x|a_1) \left\{E[Y|a_1, x] - E[Y|a_1, m_0, x]\right\} \\
&= \sum_x P(x|a_1) \left\{E[Y|a_1, m_1, x] \cdot P(m_1|x, a_1) + E[Y|a_1, m_0, x] \cdot P(m_0|x, a_1) - E[Y|a_1, m_0, x]\right\} \\
&= \sum_x P(x|a_1) \left\{E[Y|a_1, m_1, x] \cdot P(m_1|x, a_1) + E[Y|a_1, m_0, x] \cdot [P(m_0|x, a_1) - 1]\right\} \\
&= \sum_x P(x|a_1) \left\{E[Y|a_1, m_1, x] \cdot P(m_1|x, a_1) - E[Y|a_1, m_0, x] \cdot [1 - P(m_0|x, a_1)]\right\} \\
&= \sum_x P(x|a_1) \left\{E[Y|a_1, m_1, x] \cdot P(m_1|x, a_1) - E[Y|a_1, m_0, x] \cdot P(m_1|x, a_1)\right\} \\
&= \sum_x P(x|a_1) P(m_1|x, a_1) \left\{E[Y|a_1, m_1, x] - E[Y|a_1, m_0, x]\right\}
\end{aligned}
$$

A.2 Large-Sample Bias Under One-Sided Noncompliance

The front-door and standard covariate adjustment ATT bias can be written as the following, utilizing the fact that $P(m_0|a_0) = 1$ and $P(m_0|a_1) = 0$:

$$
\begin{aligned}
B_{att}^{fd} &= \mu_1 - \mu_{0|a_1}^{fd} - (\mu_1 - \mu_{0|a_1}) \\
&= \mu_{0|a_1} - \mu_{0|a_1}^{fd} \\
&= -B_{a_1}^{fd} \\
&= \sum_x P(x|a_1) \sum_u E[Y|u, a_0, m_0, x] P(u|a_1, x) \\
&- \sum_x P(x|a_1) \sum_u E[Y|u, a_1, m_0, x] P(u|a_1, x, m_0)
\end{aligned}
$$

Adding and subtracting $\sum_x P(x) \sum_u E[Y|a_0, m_0, u] \cdot P(u|a_1, m_0)$:

$$
\begin{aligned}
&= \sum_x P(x|a_1) \sum_u E[Y|u, a_0, m_0, x] \cdot [P(u|a_1, x) - P(u|a_1, x, m_0)] \\
&- \sum_x P(x|a_1) \sum_u \{E[Y|u, a_1, m_0, x] - E[Y|u, a_0, m_0, x]\} \cdot P(u|a_1, m_0, x)
\end{aligned}
$$

$$
\begin{aligned}
B_{att}^{std} &= \mu_1 - \mu_{0|a_1}^{std} - (\mu_1 - \mu_{0|a_1}) \\
&= \mu_{0|a_1} - \mu_{0|a_1}^{std} \\
&= -B_{a_1}^{std} \\
&= \sum_x P(x|a_1) \sum_u E[Y|u, a_0, m_0, x] \cdot [P(u|a_1, x) - P(u|a_0, x)]
\end{aligned}
$$

A.3 Front-door Bias Simplification

The front-door bias under one-sided noncompliance can be written as:

$$
B_{att}^{fd} = \sum_x P(x|a_1) \sum_u E[Y|a_0, m_0, x, u] \underbrace{[P(u|a_1, x) - P(u|a_1, m_0, x)]}_{\varepsilon} \tag{7}
$$

$$
+ \sum_x P(x|a_1) \sum_u \{\underbrace{E[Y|a_0, m_0, x, u] - E[Y|a_1, m_0, x, u]}_{\eta}\} P(u|a_1, m_0, x). \tag{8}
$$

ε can be rewritten as:

$$
\begin{aligned}
\varepsilon &= P(u|a_1, x) - P(u|a_1, m_0, x) \\
&= P(u|a_1, m_1, x)P(m_1|a_1, x) + P(u|a_1, m_0, x)P(m_0|a_1, x) - P(u|a_1, m_0, x) \\
&= P(u|a_1, m_1, x)P(m_1|a_1, x) + P(u|a_1, m_0, x)[P(m_0|a_1, x) - 1] \\
&= P(u|a_1, m_1, x)P(m_1|a_1, x) - P(u|a_1, m_0, x)P(m_1|a_1, x) \\
&= P(m_1|a_1, x)[P(u|a_1, m_1, x) - P(u|a_1, m_0, x)].
\end{aligned}
$$

We can also expand η as:

$$
\begin{aligned}
\eta &= E[Y|a_0, m_0, x, u] - E[Y|a_1, m_0, x, u] \\
&= E[Y|a_0, x, u] - E[Y|a_1, m_0, x, u] \\
&= E[Y(a_0)|a_0, x, u] - E[Y|a_1, m_0, x, u] \\
&= E[Y(a_0)|a_1, x, u] - E[Y|a_1, m_0, x, u] \\
&= E[Y(a_0)|a_1, m_1, x, u]P(m_1|a_1, x, u) + E[Y(a_0)|a_1, m_0, x, u]P(m_0|a_1, x, u) - E[Y|a_1, m_0, x, u] \\
&= E[Y(a_0)|a_1, m_1, x, u]P(m_1|a_1, x, u) - E[Y(a_0)|a_1, m_0, x, u] \cdot \{\frac{E[Y|a_1, m_0, x, u]}{E[Y(a_0)|a_1, m_0, x, u]} - P(m_0|a_1, x, u)\} \\
&= P(m_1|a_1, x) \left\{ E[Y(a_0)|a_1, m_1, x, u]\frac{P(m_1|a_1, x, u)}{P(m_1|a_1, x)} - E[Y(a_0)|a_1, m_0, x, u] \cdot \frac{\frac{E[Y|a_1, m_0, x, u]}{E[Y(a_0)|a_1, m_0, x, u]} - P(m_0|a_1, x, u)}{P(m_1|a_1, x)} \right\}.
\end{aligned}
$$

We note that the bias can be written as scaled by the compliance proportion within levels of x ($P(m_1|a_1, x)$).

We can thus rewrite front-door bias under one-sided noncompliance as:

$$
\begin{aligned}
B^{fd}_{att} = \sum_x P(x|a_1)P(m_1|a_1, x) \sum_u \Bigg[& E[Y|a_0, m_0, x, u] \cdot [P(u|a_1, m_1, x) - P(u|a_1, m_0, x)] \\
+ \left\{ E[Y(a_0)|a_1, m_1, x, u]\frac{P(m_1|a_1, x, u)}{P(m_1|a_1, x)} - E[Y(a_0)|a_1, m_0, x, u] \cdot \frac{\frac{E[Y|a_1, m_0, x, u]}{E[Y(a_0)|a_1, m_0, x, u]} - P(m_0|a_1, x, u)}{P(m_1|a_1, x)} \right\} & P(u|a_1, m_0, x) \Bigg].
\end{aligned}
$$

A.4 Front-door Bias Under Assumption 3

Assumption 3 and binary M implies that $\varepsilon = 0$:

$$
\begin{aligned}
P(m_1|a_1, x, u) &= \frac{P(u|a_1, m_1, x) \cdot P(m_1|a_1, x)}{P(u|a_1, x)} \\
1 &= \frac{P(u|a_1, m_1, x)}{P(u|a_1, x)} \\
P(u|a_1, x) &= P(u|a_1, m_1, x)
\end{aligned}
$$

Since M is binary, by similar logic as above we know that $P(u|a_1, x) = P(u|a_1, m_0, x)$.

Therefore:

$$\begin{aligned}\varepsilon &= P(m_1|a_1,x)[P(u|a_1,m_1,x) - P(u|a_1,m_0,x)] \\ &= P(m_1|a_1,x)[P(u|a_1,x) - P(u|a_1,x)] \\ &= 0\end{aligned}$$

Under Assumption 3, we can simplify front-door bias to:

$$\begin{aligned}B^{fd}_{att} = \sum_x P(x|a_1)P(m_1|a_1,x)\sum_u \Bigg[& E[Y(a_0)|a_1,m_1,x,u]\frac{P(m_1|a_1,x,u)}{P(m_1|a_1,x)} \\ & - E[Y(a_0)|a_1,m_0,x,u]\cdot\frac{\frac{E[Y|a_1,m_0,x,u]}{E[Y(a_0)|a_1,m_0,x,u]} - P(m_0|a_1,x,u)}{P(m_1|a_1,x)}\Bigg]\cdot P(u|a_1,x).\end{aligned}$$

B Front-Door Difference-in-Differences Proofs

B.1 No Large-Sample Bias in the Front-door Difference-in-Differences Estimator

First define $\tau_{att,x,g_1} = E[Y(a_1)|a_1,x,g_1] - E[Y(a_0)|a_1,x,g_1]$. It is well known that $\tau_{att,g_1} = \sum_x \tau_{att,x,g_1}P(x|a_1,g_1)$. Therefore in order to show that τ^{fd-did}_{att,g_1} has no bias, we need only show a lack of bias for τ_{att,x,g_1} within levels of x. If Assumptions 1 and 5 hold, then the front-door difference-in-differences estimator has no large-sample bias:

$$\begin{aligned}\tau^{fd-did}_{att,x,} &= \tau^{fd}_{att,x,g_1} - \frac{P(m_1|a_1,x,g_1)}{P(m_1|a_1,x,g_2)}\tau^{fd}_{att,x,g_2} \\ &= \tau_{att,x} + B^{fd}_{att,x,g1} - \frac{P(m_1|a_1,x,g_1)}{P(m_1|a_1,x,g_2)}\tau^{fd}_{att,x,g_2} \\ &= \tau_{att,x} + B^{fd}_{att,x,g1} - B^{fd}_{att,x,g1} \\ &= \tau_{att,x}\end{aligned}$$

C National JTPA Study Data

Our paper uses the following samples from the National JTPA Study: experimental active treatment group, experimental control group, and the nonexperimental / eligible nonparticipant (ENP) group. For our purposes, the active treatment group means receving any JTPA service, although the type of

services actually received varied across individuals.[13] In this analysis, we only examine adult males and follow the sample restrictions in Appendix B1 of Heckman et al. (1998). We also restrict our attention to the 4 *service delivery areas* at which the ENP sample was collected: Fort Wayne, IN; Corpus Christi, TX; Jackson, MS, and Providence, RI. The final sample sizes (by marital status) are presented in Table 1.

Table 1: Sample sizes for adult males by marital status. The treated units are broken up into compliers (C) and noncompliers (NC). Control denotes experimental control and ENP denotes the eligible nonparticipants.

	Treated		Control	ENP
	C	*NC*		
Non-single	484	304	274	292
Single	350	318	266	92

We use the same dataset as Section 11 of Heckman et al. (1998). The data contains all three samples in our analysis, as well as compliance information for the experimental treated group sample. We obtained this dataset in communication with Jeffrey Smith and Petra Todd. We cross-checked the covariates we utilize in our analysis against the raw data, available as part of the National JTPA Study Public Use Data from the Upjohn Institute. We established that all covariates in our conditioning sets are identical. However, the marital status variable that denotes whether individuals are currently, or were once, married was imputed as described in Appendix B3 of Heckman et al. (1998) and thus does not exactly match the raw data. We treat all individuals with values of the marital status variable that fall between 0 and 1 (non-inclusive) as married. We note that any given coding scheme is highly unlikely to alter results since only 3% of observations fall in this range.

The outcome variable we use in the analysis is total 18-month earnings in the period following random assignment (for experimental units) or eligiblity screening (for ENPs). We have verified the earnings data from our data against the earnings variables in the public use data (`totearn` variable), and they match exactly except for several individuals where Heckman et al. (1998) have imputed missing monthly data. The imputation applies to around 1% of observations and thus is unlikely to substantively alter results. A unit-by-unit comparison of earnings across the raw data and the data we are using can be obtained from us upon request. Note also that some individuals had missing earnings data for some months. In the construction of the 18-month total earnings variable, we mean impute the missing months using the average of the individual's available monthly earnings. Details on the extent of missingness are available from authors upon request.

[13]The National JTPA Study classified services received into the following 6 categories: classroom training in occupational skills, on-the-job training, job search assistance, basic education, work experience, and miscellaneous.

D Florida Voting Data

To construct our population of eligible voters, we examine individuals that have appeared in one of four voter registration snapshots: book closing records from 10/10/2006, 10/20/2008, and 10/18/2010, as well as a 2012 election recap record from 1/4/2013. This yields a total population of 16,371,725 individuals that we are able to subset by race (Asian / Pacific Islander, Black (not Hispanic), Hispanic, White (Not Hispanic), and Other). Note that the Other category contains individuals who self-identify as American Indian / Alaskan Native, Multiracial, or Other, as well as individuals for whom race is unknown. In cases where race changes across voter registration snapshots for the same voter, we use the latest available self-reported race. Such changes affect only 1.1% of observations. The breakdown of the the population by race is presented in the rightmost column of Table 2.

We use voter history files from 08/03/2013 to subset the population by voting mode in each election. The voter history files required pre-processing before we could use them for estimation. As mentioned in Gronke and Stewart (2013) and Stewart (2012), there is an issue of duplication of voter identification numbers within the same election. In some cases, this duplication is rather innocuous because the voting mode is identical across records. In these cases, we simply remove duplicate records and include the voter in our analysis. In other cases, voters are recorded as both having voted in a given election and not having voted (code "N"). In these cases, we assume that the voter did indeed cast a ballot and use that code. Finally, there are a few instances in which a voter is recorded to have voted in multiple ways. For example, a voter history file may indicate that a voter voted both absentee and early at a given election. While Gronke and Stewart (2013) indicates that voters may legitimately appear multiple times in the voter history file, this makes the task of stratifying by voting mode difficult. As a result, we choose to exclude individuals who are recorded to have voted using more than one mode. When analyzing the 2008 election subsetting by 2006 voting modes, we exclude 385 individuals. The corresponding numbers for analysis of the 2012 election subsetting using 2008 and 2010 voting groups are 1951 and 2998, respectively. These figures are dwarfed by the sample sizes and thereby highly unlikely to exert any serious effect upon our estimates.

We also made several choices regarding the definition of voting modes. Specifically, we classified anyone who voted absentee (code "A") and whose absentee ballot was not counted (code "B") as having voted absentee. We classified anyone who voted early (code "E") and anyone cast a provisional ballot early (code "F") as having voted early. Finally, we classify anyone who voted at the polls (code "Y") and cast a provisional ballot at the polls (code "Z") as having voted on election day. We do not use the code "P", which indicates that an individual cast a provisional ballot that was not counted since we cannot ascertain whether it was cast on election day, early, or as an absentee voter.

Another difficulty with the data is defining the eligible electorate and thus individuals who did not vote. While the voter history files have a code "N" for did not vote, most individuals who do not vote are not present in the voter history files at all. For example, for the 2008 election there were no "N" codes at all in the voter history files. Therefore, we count an individual as not having voted in a given election if they appeared in the voter registration files at one point but are either not present in the voter history file for that election or are coded as "N".

Table 2: Voting modes as percent of population in 2006, 2008, and 2010 elections. Population is defined as anyone who has appeared in voter registration records from 2006-2012. Note that percentages of individuals who did not vote, whose provisional ballots were not counted, or who are dropped due to conflicting voting modes are not shown.

	2006			2008			2010			
Race	Early	Absentee	Election Day	Early	Absentee	Election Day	Early	Absentee	Election Day	Total
Asian	2.87	2.61	12.21	15.15	10.45	20.43	4.75	5.69	13.76	233664
Black	2.81	1.85	16.86	27.67	7.14	16.97	6.41	4.24	18.57	2159473
Hispanic	2.46	2.83	12.46	15.05	8.60	22.72	4.11	6.01	13.37	2049683
White	5.89	5.60	23.16	14.48	13.57	25.32	7.39	9.05	20.63	11179293
Other	2.60	2.43	11.97	13.67	8.12	19.31	3.87	4.19	12.26	749612

Table 3: Rate of compliance (percent) with early voting program by prior voting mode and race in 2006-2008, 2008-2012, and 2010-2012 transitions (e.g., for the 2006-2008 transition, prior voting mode is based on voting behavior in 2006, while compliance rate is proportion voting early in 2008). Note that percentages of individuals who did not vote, whose provisional ballots were not counted, or who are dropped due to conflicting voting modes are not shown.

	2006-2008			2008-2012			2010-2012		
Race	Early	Absentee	Election Day	Early	Absentee	Election Day	Early	Absentee	Election Day
Asian	55.43	12.14	27.59	37.48	8.50	12.68	57.16	9.12	24.65
Black	71.53	18.66	53.70	49.12	12.35	19.48	67.91	15.05	43.65
Hispanic	56.70	8.98	29.50	33.18	6.81	11.02	53.03	7.31	21.63
White	53.87	9.19	22.98	40.54	7.69	12.33	56.69	7.11	20.41
Other/Unknown	55.42	10.43	28.10	37.15	7.33	11.06	57.43	8.28	24.50

E Early Voting Results

Table 4: Front-door estimates for EIP effect by race for 2006-2008, 2008-2012, and 2010-2012 transitions (e.g., for the 2006-2008 transition, prior voting mode is based on voting behavior in 2006 and EIP estimate is for 2008). All estimates use county fixed effects. 99% block boostrapped confidence intervals are reported in brackets.

Race	2006-2008	2008-2012	2010-2012
Asian	0.1548	0.1587	0.1476
	[0.1421, 0.1857]	[0.1608, 0.1778]	[0.14, 0.1823]
Black	0.2239	0.2101	0.1793
	[0.1923, 0.2455]	[0.1899, 0.2325]	[0.1582, 0.2191]
Hispanic	0.1442	0.1305	0.128
	[0.132, 0.2037]	[0.1131, 0.1664]	[0.1096, 0.1821]
White	0.1292	0.1539	0.1325
	[0.1064, 0.1564	[0.136, 0.173]]	[0.1131, 0.1563]
Other/Unknown	0.1699	0.1756	0.1623
	[0.1421, 0.2021]	[0.1608, 0.1897]	[0.14, 0.1869]

Table 5: Front-door difference-in-differences estimates for EIP effect by race for 2006-2008, 2008-2012, and 2010-2012 transitions (e.g., for the 2006-2008 transition, prior voting mode is based on voting behavior in 2006 and EIP estimate is for 2008). Estimates are reported across different group of interest and differencing groups. All estimates use county fixed effects. 99% block boostrapped confidence intervals are reported in brackets.

Race	Group of interest - Differencing group	2006-2008	2008-2012	2010-2012
Asian	EIP-Absentee	0.0535	0.0169	0.0528
		[0.028, 0.0785]	[0.01, 0.0233]	[0.0352, 0.0757]
	EIP-Election Day	0.0569	0.0225	0.0501
		[0.0398, 0.0801]	[0.0177, 0.0303]	[0.0344, 0.0739]
	Election Day-Absentee	-0.0027	-0.0023	0.001
		[-0.0093, 0.0023]	[-0.005, 1e-04]	[-0.0036, 0.0048]
Black	EIP-Absentee	0.0922	0.028	0.0692
		[0.0658, 0.1134]	[0.0181, 0.0375]	[0.0538, 0.0934]
	EIP-Election Day	0.0694	0.0185	0.0502
		[0.0544, 0.0811]	[0.0104, 0.0261]	[0.0346, 0.078]
	Election Day-Absentee	0.0152	0.0032	0.0101
		[-0.0022, 0.0286]	[-0.0014, 0.0077]	[0.0021, 0.0177]
Hispanic	EIP-Absentee	0.0532	0.0162	0.0377
		[0.0445, 0.0782]	[0.0126, 0.0245]	[0.027, 0.0703]
	EIP-Election Day	0.0454	0.0082	0.0346
		[0.0385, 0.0758]	[0.003, 0.0193]	[0.0257, 0.0624]
	Election Day-Absentee	0.0037	0.0025	0.0012
		[-0.003, 0.0059]	[-9e-04, 0.0046]	[-0.0014, 0.0054]
White	EIP-Absentee	0.033	0.0098	0.0349
		[0.0201, 0.049]	[0.0024, 0.0171]	[0.0228, 0.0495]
	EIP-Election Day	0.0452	0.0218	0.0426
		[0.0334, 0.0615]	[0.0168, 0.0278]	[0.0316, 0.0565]
	Election Day-Absentee	-0.0054	-0.0038	-0.0029
		[-0.0071, -0.0038]	[-0.0056, -0.002]	[-0.0046, -0.0013]
Other/Unknown	EIP-Absentee	0.0477	0.0169	0.0482
		[0.0305, 0.0707]	[0.0096, 0.0228]	[0.0315, 0.0652]
	EIP-Election Day	0.0558	0.0191	0.0493
		[0.042, 0.0726]	[0.0149, 0.024]	[0.037, 0.0646]
	Election Day-Absentee	-0.0042	-9e-04	-4e-04
		[-0.0078, 0]	[-0.0034, 0.0015]	[-0.0048, 0.0037]