Blueprint Labs

Discussion Paper #2026.01

One Instrument, Many Treatments: Instrumental Variables Identification of Multiple Causal Effects

Joshua Angrist
Andres Santos
Otávio Tecchio

January 2026

MIT Department of Economics
77 Massachusetts Avenue, Bldg. E53-390
Cambridge, MA 02139

National Bureau of Economic Research
1050 Massachusetts Avenue, 3rd Floor
Cambridge, MA 02138

One Instrument, Many Treatments: Instrumental Variables Identification of Multiple Causal Effects*

Joshua Angrist[†] Andres Santos[‡] Otávio Tecchio[§]

January 2026

Abstract

Many instrumental variables applications specify a single Bernoulli treatment. But instruments may change outcomes through multiple pathways or by varying treatment intensity. Lottery instruments that boost charter school enrollment, for instance, may affect outcomes by lengthening time enrolled in a charter school and by moving students between charter schools of different types. We analyze the identification problem such scenarios present in a framework that generalizes the always-taker/never-taker/complier partition of treatment response types to cover a wide range of multinomial and ordered treatments with heterogenous potential outcomes. This framework yields novel estimators in which a single randomly assigned instrument identifies (i) causal effects averaged over complier types and (ii) a causal conditional expectation function that captures effects for each element in a set of response types. Three empirical applications demonstrate the utility of these results. The first extends an earlier analysis of the Head Start Impact Study allowing for multiple fallbacks. The second examines two causal channels for the impact of post-secondary financial aid on degree completion. The third estimates effects of additional births (an ordered treatment) on mothers' earnings.

*We thank participants at the Harvard-MIT Econometrics Seminar for helpful comments, Pat Kline and Chris Walters for help with the Head Start data and for helpful comments, and Fatima Djalalova and Abigail Orbe for extraordinary research assistance.

[†]MIT Department of Economics and NBER. Email: angrist@mit.edu

[‡]UCLA. Department of Economics. Email: andres@econ.ucla.edu

[§]MIT Department of Economics. Email: otecchio@mit.edu

Instrumental variables (IV) research designs exploit natural experiments or intentional random assignment as a source of variation in treatment variables of all kinds. Many of the most compelling instruments leverage random assignment. For instance, many IV studies of charter schools (surveyed in Chabrier et al. (2016)) use conditionally-random admissions lotteries as instruments for charter-school enrollment. Under plausible assumptions, IV estimates using randomized charter offers to instrument a charter enrollment dummy identify causal effects of charter attendance in the population induced to enroll by random assignment. Other lottery designs derive instruments for Medicaid (analyzed in Finkelstein et al. (2012)) and the draft lottery that determined eligibility for Vietnam-Era conscription (analyzed in Angrist (1990)). In clinical trials with partial compliance, random assignment to treatment, also called intention-to-treat, can be used to instrument treatment received, identifying effects on trial participants who comply with the trial protocol (Angrist and Hull, 2023).

Random assignment alone does not guarantee instrument validity. Angrist and Imbens (1995) and Andresen and Huber (2021) note that quotidian implementation details such as recoding an ordered treatment as Bernoulli may cause an exclusion restriction failure, thereby muddling IV estimates. Rose and Shem-Tov (2024) likewise highlight this point, arguing that the Medicaid treatment in Finkelstein et al. (2012) can be seen as years of exposure rather than as a dummy treatment indicating any Medicaid enrollment. Relying on an additive linear model for identification, Abdulkadiroglu et al. (2017) and other charter lottery studies similarly code charter treatment exposure as years enrolled. Angrist and Evans (1998), by contrast, uses dummies indicating multiple second births and samesex sibships to instrument a dummy indicating the presence of a third child. Exclusion fails here if these instruments shift childbearing at parities higher than three.

Exclusion failures can arise when a single instrument affects outcomes through multiple channels. For example, Angrist and Chen (2011) uses randomly assigned Vietnam-Era draft eligibility as an instrument for schooling. This identification strategy is motivated by the GI Bill, a major source of post-secondary aid for Vietnam veterans and a possible explanation of earnings gaps by veteran status. But draft-eligibility also increases the likelihood of military service, arguably the dominant channel through which draft eligibility affects earnings and other long-run outcomes. Angrist and Chen (2011) allows for this using additive constant-effects models with two treatments, identified by multiple instruments. In the charter school research domain, Abdulkadiroglu et al. (2017) uses parametric models to distinguish causal effects of different sorts of charter schools. However, two-stage least squares (2SLS) estimators with two additive treatments may fail to identify average causal effects in a world of heterogeneous potential outcomes (a point made in Behagel et al. (2013) and Kirkeboen et al. (2016)).

The interpretation of IV estimates is also complicated by heterogeneous counterfactuals. Abdulkadiroglu et al. (2017) notes, for instance, that a given charter school offer may pull applicants from other charters, possibly differing in type or quality, as well as increasing the odds of any-charter enrollment. Similarly, in a re-analysis of data from a randomized evaluation of the Head Start preschool program, Kline and Walters (2016) uses the random assignment of Head Start seats and a structural model to distinguish causal effects of Head Start for children drawn into the program from home care and from competing preschools. Potential outcomes in these scenarios are determined by multinomial treatments any one of which may be changed by the instrument at hand.

Some multinomial treatment variables are meant to *explain* causal effects. For example, Angrist et al. (2022) analyzes data from a randomized trial that offered generous financial aid to college-bound high school seniors in the treatment group. This intervention boosted BA completion rates sharply. Angrist et al. (2022) argues that the resulting degree gains are caused exclusively by early four-year engagement as measured by college credits earned in freshman year. But other channels such as a reduction in the burden of growing college debt may matter as much or more than early engagement. Revisiting this application, we show how a single experiment can be used to distinguish the early engagement and loan reduction channels generating causal effects.

This paper develops identification and estimation strategies for a broad class of such problems. Specifically, we state assumptions and establish identification results for potential outcomes models that use a single instrument to identify a class of average causal effects in settings with ordered and multinomial treatments. We also extend the weighting strategy introduced in Abadie (2003) to identify a causal conditional expectation function that characterizes mean potential outcomes conditional on potential Bernoulli treatments. Our extensions identify average causal effects in settings beyond those with Bernoulli treatments, while delineating the assumptions needed to deliver identification in these contexts.

Our analysis contributes to a large literature on IV identification focused on beyond-Bernoulli treatments. We share with Heckman et al. (2006), Lee and Salanié (2018), Pinto (2021), Dutz et al. (2021), Goff (2024), and Nibbering and Oosterveen (Forthcoming) (among others) use of restrictions on choice behavior to identify causal parameters. Mogstad and Torgovitsky (2024) surveys alternative IV paradigms for analysis of models with heterogeneous potential outcomes. Our analysis also complements Torgovitsky (2015), D'Haultfœuille and Février (2015), and Chernozhukov et al. (2024) which likewise study identification using discrete instruments, though using assumptions unrelated to ours. In extensions of Imai et al. (2010), Frölich and Huber (2017), Kwon and Roth (2025), and Athey et al. (2025) explore the identifying power of IV-inspired restrictions on potential treatments like monotonicity in causal mediation analysis.

The identification strategies developed here are closely related to the principal stratification approach to causal inference pioneered in Frangakis and Rubin (2002). Principal strata are conditioning sets defined by latent response types; conditional on these (and possibly on covariates), treatment is independent of potential outcomes. The next section starts with a restatement of this foundational result. Heckman and Pinto (2018) extends the principal strata idea to IV models with unordered treatments and discusses identification under various restrictions on latent response types. Unlike Heckman and Pinto (2018), we introduce limited homogeneity restrictions on potential outcomes to facilitate identification of average causal effects for individual response types. This leads to looser, easily-verified alternatives to the restrictions on response types considered in earlier work.

We also build on Navjeevan et al. (2023), which gives high-level necessary and sufficient conditions for identification of causal parameters in a general class of potential outcomes models. We apply and extend lessons from Navjeevan et al. (2023) by introducing new causal parameters and novel, widely-applicable identifying restrictions. Our results leverage unique features of models powered by a single Bernoulli instrument to obtain transparent, verifiable (in the context of a particular model) necessary and sufficient conditions for identification. Finally, we derive simple weighted least squares estimands for parameters identified by our extensions of Abadie (2003), as well as straightforward IV estimands. Our results are unique in their derivation of a minimum mean-squared error approximation to underlying, possibly nonlinear, causal models for beyond-Bernoulli treatments. An appendix collects asymptotic results for estimators constructed as the sample analog of the estimands developed here. The utility of our theoretical results is demonstrated in empirical applications revisiting questions addressed in Kline and Walters (2016), Angrist et al. (2022), and Angrist and Evans (1998).

1 Econometric Framework

We begin by describing the causal models of interest. The data consist of an outcome $Y \in \mathbf{R}$, a discrete treatment $T \in \mathbf{T} \equiv \{t_1, \ldots, t_d\}$, a Bernoulli instrument $Z \in \mathbf{Z} \equiv \{0, 1\}$, and covariates $X \in \mathbf{X}$. Since d may exceed two, treatments can be ordered or multinomial. Observed outcomes

$$Y = \sum_{t \in \mathbf{T}} Y^*(t) 1\{T = t\}$$

are determined by a vector of potential outcomes $Y^* \equiv (Y^*(t_1), \ldots, Y^*(t_d)) \in \mathbf{R}^d$. The realized treatment determines which potential outcome we see. Observed treatments $T = T^*(0)(1 - Z) + T^*(1)Z$ are determined by a pair of potential treatments, $T^* \equiv (T^*(0), T^*(1)) \in \mathbf{T} \times \mathbf{T}$. Response type for each individual is defined by the pair of potential treatments.

The following assumption encapsulates this setup.

Assumption 1.1. *(i)* $Y^* \equiv (Y^*(t_1), \ldots, Y^*(t_d)) \in \mathbf{R}^d$, $Z \in \mathbf{Z} \equiv \{0, 1\}$, $X \in \mathbf{X}$, *and* $T^* = (T^*(0), T^*(1))$ *with* $T^*(z) \in \mathbf{T} \equiv \{t_1, \ldots, t_d\}$ *for* $z \in \{0, 1\}$; *(ii)* $T = T^*(0)(1 - Z) + T^*(1)Z$ *and* $Y = \sum_{t \in \mathbf{T}} Y^*(t) 1\{T = t\}$; *(iii)* (Y^*, T^*) *is independent of* Z *conditional on* X; *(iv)* $E[|Y^*(t)|] < \infty$ *for all* $t \in \mathbf{T}$ *and* $P(Z = 1|X) \in (0, 1)$ *with probability one.*

Assumption 1.1(iii) says that instrumental variable Z is as good as randomly assigned in the sense of being independent of response types and potential outcomes (conditional on covariates). This assumption implies an exclusion restriction since, for a given treatment value, potential outcomes are presumed to be unrelated to the value of Z. Assumption 1.1(iv) also ensures that Z varies conditional on covariates and that average potential outcomes are well-defined.

Given Assumption 1.1, average causal effects are identified by restrictions on the manner in which an instrument affects treatment. Imbens and Angrist (1994) *monotonicity*, for example, says that Z can move a Bernoulli treatment in one direction only. Monotonicity is extended here to a wide class of restrictions suitable for ordered and multinomial treatments. These restrictions limit the set of possible response types. Specifically, T^* is presumed to belong to a known set $\mathbf{M}$, of which a subset are induced by an instrument to change treatment status. In formalizing this, we use the notation $t^* = (t^*(0), t^*(1))$ to denote possible values of $T^* \in \mathbf{M}$.

Assumption 1.2. *(i)* $P(T^* \in \mathbf{M}) = 1$ *for a set of response types,* $\mathbf{M} \subseteq \mathbf{T} \times \mathbf{T}$; *(ii) Subset* $\mathbf{S} \equiv \{t^* \in \mathbf{M} : t^*(1) \neq t^*(0)\}$ *of* $\mathbf{M}$, *is non-empty and* $P(T^* \in \mathbf{S}) > 0$.

Assumption 1.2(ii) says that an assignment model characterized by $\mathbf{M}$ includes a non-empty subset of potential response types, denoted $\mathbf{S}$, induced by the instrument to *switch* from one treatment to another. Our results identify average causal effects for switchers, that is, for response types in $\mathbf{S}$.

The Imbens and Angrist (1994) setup without covariates illustrates this framework. In this case, $T \in \{0,1\}$ and monotonicity restricts T^* to satisfy $T^*(1) \geq T^*(0)$ with probability one. In our notation, monotonicity is equivalent to setting $\mathbf{M} = \{t^* : t^*(1) \geq t^*(0)\}$. The subset of switchers, $\mathbf{S}$, contains a single type (compliers), so $\mathbf{S} = \{(0,1)\}$. Under monotonicity and Assumptions 1.1 and 1.2, a Wald-IV estimand identifies a local average treatment effect (LATE):

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[T|Z=1] - E[T|Z=0]} = E[Y^*(1) - Y^*(0)|T^*(1) > T^*(0)] \equiv \tau_w.$$

The causal effect here is for a single switcher type, the set of compliers, while the set of allowable response types includes compliers plus always-takers (with $T^*(1) = T^*(0) = 1$) and never-takers (with $T^*(1) = T^*(0) = 0$). The fact that causal effects are identified only for a subset of potential response types, those induced to change treatment status by changing Z, is key to the development that follows.

As a stepping stone to more general results, note that τ_w is the slope in a *causal conditional expectation function* (CCEF) that describes the expectation of potential outcomes conditional on potential assignments. This is a consequence of Assumption 1.1(iii) (independence of Z and (Y^*, T^*) conditional on X), Assumption 1.1(iv) ($P(Z=1|X) \in (0,1)$) implying compliers are treated with probability strictly between zero and one, and Assumption 1.2(ii) ($P(T^* \in \mathbf{S}) > 0$) requiring that the probability of compliance be positive. The CCEF for basic LATE is:

$$\begin{aligned} E[Y|T, T^*(1) > T^*(0)] &= E[Y^*(T)|T, T^*(1) > T^*(0)] \\ &= E[Y^*(0)|T^*(1) > T^*(0)] + \tau_w \times T. \end{aligned}$$

Because T disappears from the conditional expectation function for potential outcomes, the CCEF is generated by a notional randomized trial in which T is randomly assigned in the population of compliers. The CCEF is causal since it generates comparisons of mean potential outcomes for a fixed reference population, in this case, the population of compliers.

Heckman and Pinto (2018) note that this insight extends to a general potential-outcomes model with more elaborate treatment variables. Specifically, for an ordered or multinomial treatment, the conditional expectation of outcomes given treatment, response types, and covariates identifies average causal effects of treatment on outcome Y for treatment values and response types compatible with Assumption 1.2. In the terminology introduced in Frangakis and Rubin (2002), response types define principal strata. The following proposition expresses this result in our notation.[1]

Proposition 1.1 (CCEF Properties)**.** *If Assumptions 1.1, 1.2 hold, then for any $t^* \in \mathbf{M}$ with $P(T^* = t^*|X) > 0$ and $t \in \{T^*(0), T^*(1)\}$ the* CCEF, $E[Y|T, T^* = t^*, X]$, *satisfies:*

$$E[Y|T = t, T^* = t^*, X] = E[Y^*(t)|T^* = t^*, X].$$

[1] Appendix A.1 proves the theoretical results stated in Sections 1 and 2.

Therefore, for any $t^* \in \mathbf{M}$ such that $P(T^* = t^*|X) > 0$, the CCEF *generates response-type-specific average treatment effects as follows:*

$$E[Y^*(T^*(1)) - Y^*(T^*(0))|T^* = t^*, X] \\ = E[Y|T = T^*(1), T^* = t^*, X] - E[Y|T = T^*(0), T^* = t^*, X]. \tag{1}$$

The causal contrasts in (1) are useful only for response types that *switch* from one treatment to another in response to the instrument—otherwise $T^*(0)$ equals $T^*(1)$ and (1) is zero. Proposition 1.1 implies that, for switchers, conditional-on-T^* comparisons of mean outcomes between groups with $T^*(1) = t$ and $T^*(0) = t'$ equal average causal effects of changing treatment from t' to t. We therefore seek to identify the CCEF and type-specific average causal effects:

$$\tau(T^*) \equiv E(E[Y^*(T^*(1)) - Y^*(T^*(0))|T^*, X]|T^*), \tag{2}$$

for $T^* \in \mathbf{S}$. For instance, with an ordered treatment in $\{0, 1, 2\}$ that satisfies monotonicity (a scenario considered in Angrist and Imbens (1995)), $\tau(T^*)$ takes up to three values according to whether T^* equals $(0, 1), (0, 2)$, or $(1, 2)$.

The basic LATE scenario illustrates a key feature of the CCEF that we aim to exploit. Suppose again that $T \in \{0, 1\}$, we can ignore covariates, and potential response types satisfy monotonicity. In this case, $\tau(T^*)$ equals τ_w, an average causal effect for the single response type (compliers). At the same time, τ_w is also the slope coefficient in a regression of Y on T for compliers. That is, τ_w equals β_1 in the regression

$$(\beta_0, \beta_1) \equiv \arg\min_{b_0,b_1} E[(E[Y|T, T^*(1) > T^*(0)] - b_0 - b_1 T)^2 | T^*(1) > T^*(0)] \\ = \arg\min_{b_0,b_1} E[(Y - b_0 - b_1 T)^2 1\{T^*(1) > T^*(0)\}]. \tag{3}$$

Moreover, while *individual* compliers are not identified (because one of $T^*(1), T^*(0)$ is counterfactual), Abadie (2003) shows that (3) is equivalent to a feasible weighted-least squares minimand. Specifically, the objective in (3) satisfies

$$E[(Y - b_0 - b_1 T)^2 1\{T^*(1) > T^*(0)\}] = E[(Y - b_0 - b_1 T)^2 \kappa(T, Z)],$$

where

$$\kappa(T, Z) = 1 - \frac{T(1 - Z)}{P(Z = 0)} - \frac{(1 - T)Z}{P(Z = 1)}. \tag{4}$$

Without covariates, the regression function obtained by κ-weighting is the basic LATE CCEF. With covariates, the Abadie (2003) κ-weighted minimand generates a best linear approximation to the CCEF.

The next section extends IV and κ-weighting identification strategies to accommodate ordered and multinomial treatments of the sort allowed by Assumption 1.1. Importantly, in this more general setting, the set of switchers may contain more than one type. The generalization of the LATE Theorem established here therefore identifies causal contrasts $\tau(T^*)$ averaged over possible types T^* in $\mathbf{S}$. This parameter, which we call switcher LATE (SLATE), is defined by:

$$\tau(\mathbf{S}) \equiv E[\tau(T^*)|T^* \in \mathbf{S}].$$

SLATE generalizes an estimand defined in Kline and Walters (2016) for a trinomial treatment variable. We revisit this connection in Sections 2.2 and 3.

For basic (Bernoulli treatment) LATE with no covariates, $\tau(\mathbf{S})$ coincides with the effect on compliers obtained using (3). Beyond this scenario, however, κ-weighted construction of the CCEF potentially delivers more than SLATE, distinguishing average causal effects for distinct switcher types. This leads us to derive conditions under which the CCEF is identified. The solution here is characterized by a model, denoted $F(T, X; \beta)$, that depends on parameter vector, β. These parameters are defined by:

$$\begin{aligned}\beta &\equiv \arg\min_b E[(E[Y|T, T^*, X] - F(T, X; b))^2 | T^* \in \mathbf{S}] \\ &= \arg\min_b E[(Y - F(T, X; b))^2 1\{T^* \in \mathbf{S}\}],\end{aligned}$$

which we show are identified by a feasible weighted least squares problem. This minimization problem is shown to provide a foundation for identification of the response-type-specific average causal effects defined in (2).

2 Identification of Causal Parameters

This section begins by establishing necessary and sufficient conditions for identification of SLATE. These conditions are then extended to derive a weighted least squares minimand that identifies the CCEF. The latter identification problem is more demanding and therefore requires stronger assumptions than the former.

2.1 SLATE

SLATE averages type-specific causal contrasts, $\tau(T^*)$, over the subset of switchers $\mathbf{S} \subseteq \mathbf{M}$ defined in Assumption 1.2. The following lemma shows that identification of the probability of being a switcher is key to identification of SLATE.

Lemma 2.1. *Suppose Assumptions 1.1 and 1.2 are satisfied. Then,*

$$E[\tau(T^*) 1\{T^* \in \mathbf{S}\}] = E(E[Y|Z = 1, X] - E[Y|Z = 0, X])$$

and therefore

$$\tau(\mathbf{S}) = \frac{E(E[Y|Z = 1, X] - E[Y|Z = 0, X])}{P(T^* \in \mathbf{S})}.$$

This result raises the question of when $P(T^* \in \mathbf{S})$ is identified so that the lemma can be used to construct $\tau(\mathbf{S})$. The answer is surprisingly simple and determined by whether a particular set of linear equations is consistent, that is, has a solution, or inconsistent, meaning no solution exists. This can be shown by supplementing Corollary 4.3 in Navjeevan et al. (2023) and Assumptions 1.1 and 1.2 with two regularity conditions.[2] In stating these conditions, P_{AB} and $P_{A|B}$ denote the joint and conditional distributions of any two random variables (A, B), while $P_A \ll P_B$ means that P_A is absolutely continuous with respect to P_B.[3]

[2] Heckman and Pinto (2018) gives sufficient conditions for identification of $P(T^* \in \mathbf{S})$.

[3] A distribution P_A is absolutely continuous with respect to a distribution P_B if any set C that has zero probability under P_B also has zero probability under P_A – i.e. $P_B(C) = 0$ implies $P_A(C) = 0$.

Theorem 2.1 (Identification of $P(T^* \in \mathbf{S})$). *Suppose Assumptions 1.1, 1.2 hold. Assume also that: (i) For some $\varepsilon > 0$, with probability one, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ and $P(T^* = t^*|X) \geq \varepsilon$ for all $t^* \in \mathbf{M}$; (ii) $P_{Y^*|T^*X} \ll \{P_{Y|X}\}^d \ll P_{Y^*|T^*X}$ and $\tilde{P} \equiv P_{Y^*|X}P_{T^*ZX}$ satisfies $\tilde{P} \ll P_{Y^*T^*ZX}$ with $d\tilde{P}/dP_{Y^*T^*ZX}$ bounded. Then $P(T^* \in \mathbf{S})$ is identified if and only if there are functions ν_0 and ν_1 of T that satisfy:*

$$\nu_0(T^*(0)) + \nu_1(T^*(1)) = 1\{T^* \in \mathbf{S}\}, \tag{5}$$

with probability one.

The regularity conditions imposed by Theorem 2.1 require the instrument-assignment propensity score, $P(Z = 1|X)$, to be bounded away from zero and one, and the probability of response types allowed for by $\mathbf{M}$ to be bounded away from zero. The theorem also imposes absolute continuity requirements that restrict the joint support of the vector of potential outcomes in a manner that seems uncontroversial.[4]

Theorem 2.1 gives conditions under which restrictions on $\mathbf{M}$ are necessary and sufficient to identify $P(T^* \in \mathbf{S})$. The most important of these is equation (5). Importantly, because $T^* \in \mathbf{M}$ under Assumption 1.2, solutions to (5) depend on the restrictions embodied in $\mathbf{M}$ but not on the distribution of the data. We call (5) ISP as a reminder that this is equivalent to an *identified switcher probability*. ISP is formalized as:

Assumption ISP *There are functions ν_0 and ν_1 of T satisfying*

$$\nu_0(T^*(0)) + \nu_1(T^*(1)) = 1\{T^* \in \mathbf{S}\}, \tag{6}$$

with probability one over $T^* \in \mathbf{M}$.

Unsurprisingly, ISP is satisfied in the basic LATE scenario featured in Imbens and Angrist (1994). Recall that $\mathbf{M}$ in this case allows for three response types, $T^* = (T^*(0), T^*(1))$ equal to $(0,0), (1,1)$, or $(0,1)$. Setting $\nu_0(t) = -t$ and $\nu_1(t) = t$ for $t \in \{0,1\}$, we have:

$$\nu_0(T^*(0)) + \nu_1(T^*(1)) = T^*(1) - T^*(0),$$

which equals $1\{T^*(1) > T^*(0)\} = 1\{T^* \in \mathbf{S}\}$, as required by ISP. More generally, ISP generates a system of $|\mathbf{M}|$ linear equations that may or may not have a solution.

Given Assumptions 1.1 and 1.2, ISP is sufficient for a Wald-type estimand to identify $\tau(\mathbf{S})$, as shown in the following result.

Theorem 2.2 (ISP Identifies SLATE). *Suppose Assumptions 1.1, 1.2, and* ISP *hold, and, for some $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ with probability one. Let*

$$\tilde{Z} \equiv \frac{Z - P(Z = 1|X)}{P(Z = 0|X)P(Z = 1|X)} \tag{7}$$

[4] These require, for instance, the support of Y^* conditional on T^* to be unrelated to T^* and are used to show the necessity part of the statement (i.e., that identification of $P(T^* \in \mathbf{S})$ implies that a solution to (5) exists). Theorem 2.2 below shows that the existence of a solution to (5) is sufficient for identification of $P(T^* \in \mathbf{S})$ without absolute continuity conditions.

and, for functions ν_0 *and* ν_1 *satisfying* ISP*, let* $V \equiv \nu_1(T)Z - \nu_0(T)(1-Z)$*. Then:*

$$P(T^* \in \mathbf{S}) = E(E[V|Z=1,X] - E[V|Z=0,X]) = E[\tilde{Z}V]$$

$$\tau(\mathbf{S}) = \frac{E(E[Y|Z=1,X] - E[Y|Z=0,X])}{E(E[V|Z=1,X] - E[V|Z=0,X])} = \frac{E[\tilde{Z}Y]}{E[\tilde{Z}V]}. \tag{8}$$

This theorem identifies SLATE as an IV estimand in which treatment variable V is determined by functions ν_0 and ν_1 satisfying ISP. The first representation in (8) marginalizes over covariates in the manner discussed by Frölich (2007) for basic LATE. The second is derived using inverse propensity-score weighting (IPW), as suggested by Tan (2006).[5]

When does ISP fail? Consider a simplified ordered treatment model with $T \in \{0,1,2\}$, no covariates, and monotonicity. The set of response types $T^* = (T^*(0), T^*(1))$ allowed by $\mathbf{M}$ then consists of three types of stayers, $\{(0,0),(1,1),(2,2)\}$, and three types of switchers, $\{(0,1),(0,2),(1,2)\} = \mathbf{S}$. In this case, ISP does not hold because no functions ν_0 and ν_1 satisfy restriction (6). To see why, note first that evaluation of restriction (6) at staying response type $(T^*(0), T^*(1)) = (1,1)$ yields

$$\nu_0(1) + \nu_1(1) = 0. \tag{9}$$

Similarly, evaluating (6) at switching response types $T^* \in \{(0,1),(0,2),(1,2)\}$ gives

$$\nu_0(0) + \nu_1(1) = 1 \text{ for } T^* = (0,1) \tag{10}$$

$$\nu_0(0) + \nu_1(2) = 1 \text{ for } T^* = (0,2) \tag{11}$$

$$\nu_0(1) + \nu_1(2) = 1 \text{ for } T^* = (1,2). \tag{12}$$

Equations (9)-(12) have no solution (in ν_0, ν_1) because (10) and (11) imply $\nu_1(1) = \nu_1(2)$, which together with (12) yields $\nu_0(1) + \nu_1(1) = 1$. This in turn contradicts (9). Since ISP does not hold, SLATE is unidentified without further restrictions on $\mathbf{M}$.[6]

The identifying power of ISP is highlighted by a model adding Rose and Shem-Tov (2024) *extensive margin compliance only* (EMCO) assumption to our simple ordered-treatment story. EMCO says that an instrument induces shifts only from no treatment to positive treatment. With three treatment levels, EMCO allows three types of stayers, $\{(0,0),(1,1),(2,2)\}$, but only two types of switchers, $\{(0,1),(0,2)\} = \mathbf{S}$. Because EMCO rules out switcher type $T^* = (1,2)$, we can drop (12). Evaluating (6) for stayers, ISP requires ν_0 and ν_1 satisfy

$$\nu_0(t) + \nu_1(t) = 0 \text{ for all } t \in \{0,1,2\}. \tag{13}$$

The system of equations consisting of (10), (11), and (13) is solved by $\nu_1(t) = 1\{t > 0\}$ and $\nu_0(t) = -\nu_1(t)$. Hence, ISP holds and SLATE is identified. Specifically, since $V \equiv \nu_1(T)Z -$

[5] An IPW estimand for the conditional reduced form in the second line of (8) is

$$E\left[\frac{YZ}{P(Z=1|X)} - \frac{Y(1-Z)}{1-P(Z=1|X)}\right] = E[\tilde{Z}Y].$$

The equality above is obtained by bringing the elements of this expression over a common denominator. A parallel operation yields the denominator in the characterization of $\tau(\mathbf{S})$.

[6] Weighted average LATEs computed over *overlapping* groups of switchers may be identified without ISP, as shown in Angrist and Imbens (1995). Angrist et al. (2024) extends that approach to cover dynamic treatment effects in randomized trials with time-varying compliance.

$\nu_0(T)(1-Z)$ equals $V = 1\{T > 0\}$, Theorem 2.2 implies:

$$\tau(\mathbf{S}) = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[1\{T>0\}|Z=1] - E[1\{T>0\}|Z=0]}, \tag{14}$$

a result obtained (following a different path) in Rose and Shem-Tov (2024).[7]

ISP is also satisfied by a latent-index ordered choice model like ordered probit. To see this, suppose ε is continuously distributed on the real line, and that treatment status is governed by:

$$T^*(z) = \begin{cases} 0 & \text{if } z\pi + \varepsilon \le c_1 \\ 1 & \text{if } c_1 < z\pi + \varepsilon \le c_2 \\ 2 & \text{if } c_2 < z\pi + \varepsilon, \end{cases} \tag{15}$$

where ε is independent of Z and $\pi > 0$. This model imposes restrictions on T^* that go beyond monotonicity. Suppose initially that $\pi \ge c_2 - c_1$. In this scenario, $T^*(0) = 1$ (i.e., $c_1 < \varepsilon \le c_2$) implies $T^*(1) = 2$, because then $\pi + \varepsilon > c_2$. Hence, while there are up to three types of switchers, $\{(0,1),(0,2),(1,2)\}$, this restriction leaves only two types of stayers, $\{(0,0),(2,2)\}$.

Evaluating the ISP restrictions in (6) at these five possible response types yields:

$$\begin{aligned} \nu_0(0) + \nu_1(1) &= 1 \text{ for } T^* = (0,1) \\ \nu_0(0) + \nu_1(2) &= 1 \text{ for } T^* = (0,2) \\ \nu_0(1) + \nu_1(2) &= 1 \text{ for } T^* = (1,2) \\ \nu_0(0) + \nu_1(0) &= 0 \text{ for } T^* = (0,0) \\ \nu_0(2) + \nu_1(2) &= 0 \text{ for } T^* = (2,2). \end{aligned}$$

These equations are solved by $\nu_0(t) = -1\{t = 2\}$ and $\nu_1(t) = 1\{t > 0\}$. Therefore, ISP holds, and Theorem 2.2 implies SLATE equals:

$$\tau(\mathbf{S}) = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[1\{T>0\}|Z=1] - E[1\{T=2\}|Z=0]}. \tag{16}$$

What happens if $\pi < c_2 - c_1$? ISP is satisfied even so. To see this, note first that, in this case, $T^*(0) = 0$ (i.e., $\varepsilon \le c_1$) implies $T^*(1) < 2$, because then $\pi + \varepsilon < c_2$. So, when $\pi < c_2 - c_1$, there are no switchers with $T^* = (0,2)$. That leaves three stayer types, $\{(0,0),(1,1),(2,2)\}$, and two switcher types, $\{(0,1),(1,2)\}$. Evaluating the ISP restrictions at these five response types yields:

$$\begin{aligned} \nu_0(0) + \nu_1(1) &= 1 \text{ for } T^* = (0,1) \\ \nu_0(1) + \nu_1(2) &= 1 \text{ for } T^* = (1,2) \\ \nu_0(0) + \nu_1(0) &= 0 \text{ for } T^* = (0,0) \\ \nu_0(1) + \nu_1(1) &= 0 \text{ for } T^* = (1,1) \\ \nu_0(2) + \nu_1(2) &= 0 \text{ for } T^* = (2,2). \end{aligned}$$

These equations are solved by $\nu_1(t) = t$ and $\nu_0(t) = -t$, again implying ISP holds and SLATE

[7] Rose and Shem-Tov (2024) also derives κ-weights that identify expectations of functions of observables for switchers under EMCO. As we discuss in the next section, these weights are important, but not sufficient, for identifying the CCEF.

is identified. Applying Theorem 2.2 in this scenario yields:

$$\tau(\mathbf{S}) = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[T|Z=1] - E[T|Z=0]}, \tag{17}$$

and hence SLATE equals the Wald-IV estimand in which we instrument for T. Thus, the construction and interpretation of SLATE turns on first-stage parameter values. Yet, the ordered assignment model delivers an average causal effect either way. Appendix A.2 shows that this conclusion holds for latent-index ordered-assignment models with more than three treatments, gives general formulas for ν_0 and ν_1, characterizes the set of switchers in these models, and gives restrictions under which SLATE can be constructed by instrumenting for T.

SLATE by κ-Weighting

A restriction on allowable assignment types beyond ISP leads to new κ-weighted least squares estimands for both SLATE and the CCEF. This *in/out/stay* (IOS) assumption uses the following taxonomy of response types. For each possible treatment, t, define:

$$\begin{aligned}
\texttt{in}(\mathrm{t}) &\equiv \{t^* \in \mathbf{M} : t^*(1) \neq t^*(0) \text{ and } t^*(1) = t\} \\
\texttt{out}(\mathrm{t}) &\equiv \{t^* \in \mathbf{M} : t^*(1) \neq t^*(0) \text{ and } t^*(0) = t\} \\
\texttt{stay}(\mathrm{t}) &\equiv \{t^* \in \mathbf{M} : t^*(0) = t^*(1) = t\}.
\end{aligned}$$

This taxonomy divides response types that may select treatment level t into: (i) Switchers who select t when $Z = 1$, that is, they switch *in* to t when the instrument is turned on; (ii) Switchers who select t when $Z = 0$, that is, they switch *out* of t when the instrument is turned on; and (iii) Stayers who select t regardless of Z. In the basic LATE setup, for instance, $\texttt{stay}(0)$ contains never-takers, $\texttt{stay}(1)$ contains always-takers, $\texttt{out}(0)$ and $\texttt{in}(1)$ contain compliers, while $\texttt{in}(0)$ and $\texttt{out}(1)$ are empty.

Using this notation, IOS is stated as:

Assumption IOS. *For all* $t \in \mathbf{T}$*, at least one of* $\texttt{in}(\mathrm{t})$*,* $\texttt{out}(\mathrm{t})$*, and* $\texttt{stay}(\mathrm{t})$ *is empty.*

IOS says that there is no treatment level that is selected by some switchers when $Z = 1$, selected by other switchers when $Z = 0$, and at which some stayers stay regardless of Z. In the context of a particular model, IOS is more easily checked than ISP since, for the former, no system of equations need be solved. Rather, for each treatment level, t, IOS is verified by examining response types in $\mathbf{M}$ that can choose t.

Under IOS, both a Wald-type estimand and κ-weighting identify SLATE. The probability of T^* belonging to any set of types $\texttt{in}(\mathrm{t})$ or $\texttt{out}(\mathrm{t})$ is also identified under IOS. These results build on a partition of the set of possible treatment levels into four groups:

$$\begin{aligned}
\mathbf{T}_{\texttt{i}} &\equiv \{t \in \mathbf{T} : \texttt{in}(t) \neq \emptyset \text{ and } \texttt{out}(t) = \emptyset\} \\
\mathbf{T}_{\texttt{o}} &\equiv \{t \in \mathbf{T} : \texttt{in}(t) = \emptyset \text{ and } \texttt{out}(t) \neq \emptyset\} \\
\mathbf{T}_{\texttt{io}} &\equiv \{t \in \mathbf{T} : \texttt{in}(t) \neq \emptyset \text{ and } \texttt{out}(t) \neq \emptyset\} \\
\mathbf{T}_{\texttt{s}} &\equiv \{t \in \mathbf{T} : \texttt{in}(t) = \emptyset \text{ and } \texttt{out}(t) = \emptyset\}.
\end{aligned}$$

The set $\mathbf{T}_{\texttt{i}}$ consists of treatments that response types may switch into, but not out of, when

$Z = 1$. Set $\mathbf{T}_{\mathtt{o}}$ consists of treatments that response types may switch out of, but not into, when $Z = 0$. Switchers can switch in or out of treatments in $\mathbf{T}_{\mathtt{io}}$, while treatments in $\mathbf{T}_{\mathtt{s}}$ are never selected by switchers. In the basic LATE scenario, for example, $\mathbf{T} = \{0, 1\}$ and monotonicity implies $\mathbf{T}_{\mathtt{i}} = \{1\}$, $\mathbf{T}_{\mathtt{o}} = \{0\}$, and both $\mathbf{T}_{\mathtt{io}}$ and $\mathbf{T}_{\mathtt{s}}$ are empty.

We first use the partition of $\mathbf{T}$ into these four subsets to derive generalized κ-type weights under IOS, denoted $\kappa^\star$. Specifically, for $\tilde{Z}$ as defined in (7), define

$$\kappa^\star(T, Z, X) \equiv \tilde{Z}(1\{T \in \mathbf{T}_{\mathtt{i}}\}P(Z = 1|X) - 1\{T \in \mathbf{T}_{\mathtt{o}}\}P(Z = 0|X)) + 1\{T \in \mathbf{T}_{\mathtt{io}}\}. \quad (18)$$

IOS enables SLATE identification via $\kappa^\star$-weighting. This is formalized by the following theorem:

Theorem 2.3 ($\kappa^\star$ for SLATE)**.** *Suppose Assumptions 1.1, 1.2, and* IOS *hold.*

(i) Then, ISP *holds with* $\nu_0(T) = -1\{T \in \mathbf{T}_{\mathtt{i}}\}$ *and* $\nu_1(T) = 1\{T \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}\}$*. Therefore,* $\tau(\mathbf{S})$ *equals* IV *estimand* (8) *with* $V = 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\}Z$*. Moreover,*

$$P(T^* \in \mathbf{S}) = E[\kappa^\star(T, Z, X)].$$

(ii) Let $h_0(X) \equiv 1/P(Z = 0|X)$ *and* $h_1(X) \equiv 1/[P(Z = 0|X)P(Z = 1|X)]$*. Also, suppose that* $E[Y^2] < \infty$ *and, for some* $\varepsilon > 0$, $\varepsilon \le P(Z = 1|X) \le 1 - \varepsilon$ *with probability one. Then,* $\tau(\mathbf{S}) = \beta_1$ *defined by*

$$(\beta_0, \beta_1) \equiv \arg\min_{(b_0, b_1)} E\left[\left(Y - b_0\frac{h_0(X)}{E[h_0(X)|T^* \in \mathbf{S}]} - b_1\frac{Vh_1(X)}{E[h_1(X)|T^* \in \mathbf{S}]}\right)^2 \kappa^\star(T, Z, X)\right].$$

The first part of Theorem 2.3 says that IOS implies ISP. As we illustrate below, however, some models satisfy ISP only. Theorem 2.3(i) also says that the Wald-IV representation for SLATE obtained in Theorem 2.2 simplifies to IV with Bernoulli treatment $V = 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\}Z$. The second part of the theorem says that SLATE equals a $\kappa^\star$-weighted least squares estimand. In particular, without covariates, Theorem 2.3(ii) implies SLATE is the slope of a $\kappa^\star$-weighted regression of Y on a constant and V. With covariates, SLATE can be obtained using a weighting scheme involving assignment propensity scores and the conditional variance of assignment. This result appears to be new even in the Bernoulli treatment case with monotonicity and covariates studied by Abadie (2003).[8]

To illustrate Theorem 2.3, we return to $T \in \{0, 1, 2\}$ without covariates and the ordered latent-index model for T^* given by (15). This setup satisfies ISP. Here, $\mathtt{in}(1)$ and $\mathtt{out}(1)$ are non-empty. But when $\pi < c_2 - c_1$, $\mathtt{stay}(1)$ is also non-empty since $c_1 < \varepsilon < c_2 - \pi$ generates response type $T^* = (1, 1)$. Therefore, IOS fails even though ISP holds. If $\pi \ge c_2 - c_1$, however, $\mathtt{stay}(1)$ is empty. Since $\mathtt{in}(0)$ and $\mathtt{out}(2)$ are also empty, IOS holds. Moreover, $\mathbf{T}_{\mathtt{i}} = \{2\}$ and $\mathbf{T}_{\mathtt{io}} = \{1\}$, so $V = 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\}Z$ equals $1\{T = 2\} + 1\{T = 1, Z = 1\}$. Theorem 2.3(i) then implies

$$\tau(\mathbf{S}) = \frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[1\{T > 0\}|Z = 1] - E[1\{T = 2\}|Z = 0]},$$

[8] In a basic LATE scenario with covariates required for identification, Abadie (2003) minimizes $E[(Y - b_0 - b_1T - X'b_2)^2\kappa(T, Z, X)]$ or some other parametric model for the CCEF allowing for covariate effects on potential outcomes. The minimand in Theorem 2.3(ii) requires no such specification, requiring only a model for the assignment propensity score.

the result obtained in (16) using ISP and Theorem 2.2.

Now, replace the latent-index model for T^* with EMCO. As we have noted, EMCO allows five response types, three stayers ($\{(0,0),(1,1),(2,2)\}$) and two switchers ($\{(0,1),(0,2)\}$). IOS holds here because $\texttt{in}(0)$, $\texttt{out}(1)$, and $\texttt{out}(2)$ are empty. Moreover, since $\mathbf{T}_{\texttt{i}} = \{1,2\}$ while $\mathbf{T}_{\texttt{io}}$ is empty, $V = 1\{T \in \mathbf{T}_{\texttt{i}}\} + 1\{T \in \mathbf{T}_{\texttt{io}}\}Z$ equals $1\{T > 0\}$. Theorem 2.1(i) therefore implies:

$$\tau(\mathbf{S}) = \frac{E[Y|Z=1] - E[Y|Z=1]}{E[1\{T>0\}|Z=1] - E[1\{T>0\}|Z=0]},$$

the result obtained in (14) using ISP and Theorem 2.2.

SLATE is an easily computed summary causal effect that is likely to be useful in many applications. For example, Kline and Walters (2016) interprets a conventional Wald estimand as SLATE in a scenario with a trinomial unordered treatment. The treatment variable in this case is defined by preschool sector, either Head Start, competing (non-Head Start) preschools, or home care. The Kline and Walters (2016) version of SLATE averages effects of movements into Head Start from the other two sectors. We reconsider this application in Section 2.2 and Section 3. As it turns out, the key identifying assumption here implies IOS.

In applications with ordered treatments, SLATE is likely to be most useful under monotonicity. To see this, consider a modified basic LATE scenario with no always-takers while allowing defiance. IOS fails here because $\texttt{stay}(0), \texttt{in}(0), \texttt{out}(0)$ are non-empty. ISP holds, however, with $\nu_0(t) = \nu_1(t) = t$. SLATE therefore equals:

$$\begin{aligned}\tau(\mathbf{S}) &= \frac{E[Y|Z=1] - E[Y|Z=1]}{E[T|Z=1] + E[T|Z=0]} \\ &= p_c E[Y^*(1) - Y^*(0)|T^*(1) > T^*(0)] + (1-p_c)E[Y^*(0) - Y^*(1)|T^*(1) < T^*(0)],\end{aligned}$$

where $p_c \equiv P(T^*(1) > T^*(0)|T^* \in \mathbf{S})$. Suppose now that the causal effect of treatment is a constant, τ. The usual Wald-IV estimand identifies τ. However, $\tau(\mathbf{S})$ differs from τ because SLATE averages $Y(T^*(1)) - Y(T^*(0))$ for compliers, for whom $T^*(1) = 1$ and $T^*(0) = 0$, with the corresponding effects for defiers, for whom $T^*(1) = 0$ and $T^*(0) = 1$. Under constant effects, the first effect is τ while the second is $-\tau$; SLATE weights these by p_c and $1 - p_c$, respectively.[9]

SLATE does not distinguish causal effects for different types of switchers, but IOS identifies features of the distribution of switcher types. Specifically, under IOS, the probability T^* belongs to $\texttt{in}(\mathrm{t})$ and the probability T^* belongs to $\texttt{out}(\mathrm{t})$ are identified for all $t \in \mathbf{T}$ consistent with $\mathbf{M}$. The following lemma formalizes this result.

Lemma 2.2 (Probabilities of $\texttt{in}(\mathrm{t})$ and $\texttt{out}(\mathrm{t})$)**.** *If Assumptions 1.1, 1.2, and* IOS *hold, then*

$$\begin{aligned} P(T^* \in \texttt{in}(\mathrm{t})\,|X) &= E[1\{T=t\}|Z=1,X] - E[1\{T=t\}|Z=0,X] \qquad \text{for all } t \in \mathbf{T}_{\texttt{i}}, \\ P(T^* \in \texttt{out}(\mathrm{t})\,|X) &= E[1\{T=t\}|Z=0,X] - E[1\{T=t\}|Z=1,X] \qquad \text{for all } t \in \mathbf{T}_{\texttt{o}}, \end{aligned}$$

[9]The de Chaisemartin (2017) compliers-defiers condition, requiring that there be a subpopulation of compliers, C_F, of the same size and with the same average treatment effect as defiers, yields $\tau(\mathbf{S}) = P(C_V|T^* \in \mathbf{S})E[Y^*(1) - Y^*(0)|C_V]$, where C_V is the complement of C_F in the set of compliers. This follows from the reduced form here equaling $P(C_V)E[Y^*(1) - Y^*(0)|C_V]$. This result also holds when there are more compliers than defiers conditional on the difference in potential outcomes.

and, for all $t \in \mathbf{T}_{\texttt{io}}$,

$$P(T^* \in \texttt{in}(\mathrm{t})\,|X) = E[1\{T=t\}|Z=1,X]$$
$$\textit{and} \;\; P(T^* \in \texttt{out}(\mathrm{t})\,|X) = E[1\{T=t\}|Z=0,X].$$

Moreover, defining $\tilde{Z} \equiv (Z - P(Z=1|X))/(P(Z=1|X)P(Z=0|X))$, *we obtain*

$$P(T^* \in \texttt{in}(\mathrm{t})) = E[1\{T=t\}\tilde{Z}] \qquad \text{for all } t \in \mathbf{T}_{\texttt{i}},$$
$$P(T^* \in \texttt{out}(\mathrm{t})) = -E[1\{T=t\}\tilde{Z}] \qquad \text{for all } t \in \mathbf{T}_{\texttt{o}},$$

and, for all $t \in \mathbf{T}_{\texttt{io}}$,

$$P(T^* \in \texttt{in}(\mathrm{t})) = E[1\{T=t\}Z\tilde{Z}]$$
$$\textit{and} \;\; P(T^* \in \texttt{out}(\mathrm{t})) = -E[1\{T=t\}(1-Z)\tilde{Z}].$$

This lemma links the distribution of switcher types to first-stage relationships for the set of treatment indicators of the form $1\{T=t\}$. For some models, this result also identifies the probability distribution of switcher types, T^*. This is illustrated in the context of Head Start at the end of the next section.

2.2 CCEF Identification

When are the underlying average causal effects defined by (2) identified? This question is answered by studying identification of the CCEF. Beyond basic LATE as in Imbens and Angrist (1994), resolution of this more challenging identification problem typically requires stronger assumptions than does identification of SLATE.

The CCEF is assumed to be modeled or approximated by a function $F(T, X; \beta)$ of parameter vector, β. The parameters we seek are defined by fitting F to the CCEF for switchers:

$$\beta \equiv \arg\min_b E[(E[Y|T,T^*,X] - F(T,X;b))^2 | T^* \in \mathbf{S}]$$
$$= \arg\min_b E[(Y - F(T,X;b))^2 1\{T^* \in \mathbf{S}\}].$$

Proposition 1.1 implies that if F is correctly specified (i.e., matches the CCEF), then covariate-specific average causal effects for a given switcher type T^* are given by:

$$E[Y^*(T^*(1)) - Y^*(T^*(0))|T^*, X] = F(T^*(1), X; \beta) - F(T^*(0), X; \beta). \tag{19}$$

When F differs from the CCEF, $F(T, X; \beta)$ is a minimum mean-squared error (MMSE) approximation to it.

We start by assuming that the CCEF is a function of T and X only, omitting T^*. This is formalized by the following *representation in observables* (RIO) assumption:

Assumption RIO. *There is a function* $C(T, X)$ *satisfying*

$$P(E[Y|T,T^*,X] = C(T,X)|T^* \in \mathbf{S}) = 1$$

In a basic LATE scenario without covariates, RIO is ensured since monotonicity allows only one type of switcher. In particular, the LATE CCEF can be written:

$$E[Y|T, T^* = (0, 1)] = C(T) = \beta_0 + \beta_1 T,$$

where β_1 equals LATE and $\beta_0 = E[Y^*(0)|T^* = (0, 1)]$.

Applications with non-Bernoulli treatments often generate multiple switcher types, in which case, RIO is more demanding. The following corollary to Proposition 1.1 characterizes the restrictions RIO imposes on the CCEF.

Corollary 2.1 (Interpreting RIO). *Suppose Assumptions 1.1, 1.2 hold, and that* $P(T^* = t^*|X) > 0$ *for all* $t^* \in \mathbf{S}$ *with probability one. Then,* RIO *is satisfied if and only if*

$$E[Y^*(t)|T^* = t^*, X] = E[Y^*(t)|T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t}), X], \tag{20}$$

for all $t^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})$.

RIO limits heterogeneity by requiring mean potential outcomes at treatment level t to be the same for all switcher types that can choose t. This restriction facilitates CCEF identification since it implies that, conditional on treatment level, the CCEF does not depend on unobserved switcher type T^*. For treatment values chosen by a single switcher type, (20) is irrelevant. Returning to $T \in \{0, 1, 2\}$ under EMCO, for instance, the instrument moves some switchers from 0 to 1 and some from 0 to 2. Corollary 2.1 requires the mean of $Y^*(0)$ be the same for these two types, while leaving the means of $Y^*(1)$ and $Y^*(2)$ for switchers unrestricted. Although strong, RIO has antecedents in the literature on IV identification (for example, in Mountjoy, 2022). RIO-imposed homogeneity of potential outcome *levels* contrasts with homogeneity restrictions imposed on causal effects in Hull (2018), Caetano et al. (2023), and Angrist et al. (2024), among others.

The following result shows that IOS is central to identification of a RIO-compliant CCEF.

Theorem 2.4 (The CCEF and IOS). *Suppose Assumption 1.1, 1.2, and* RIO *hold, and: (i) For some* $\varepsilon > 0$, *with probability one,* $\varepsilon \leq P(Z = 1|X) \leq 1-\varepsilon$, $P(T^* = t^*|X) \geq \varepsilon$ *for all* $t^* \in \mathbf{M}$, *and* $\mathrm{Var}\{Y^*(t)|X\} \geq \varepsilon$*; (ii)* $Y^*(t)$ *is bounded for all* $t \in \mathbf{T}$*; (iii)* $P_{Y^*|T^*X} \ll \{P_{Y|X}\}^d \ll P_{Y^*|T^*X}$ *and* $\tilde{P} \equiv P_{Y^*(t_1)|X} \ldots P_{Y^*(t_d)|X} P_{T^*ZX}$ *satisfies* $\tilde{P} \ll P_{Y^*T^*ZX}$ *with* $d\tilde{P}/dP_{Y^*T^*ZX}$ *bounded. Then the* CCEF *is identified if and only if* IOS *holds.*

Remarkably, IOS is both necessary and sufficient for CCEF identification under the conditions of Theorem 2.4.

IOS is also sufficient for identification of β as the solution to a feasible $\kappa^\star$-weighted least squares minimand. This is formalized by:

Theorem 2.5 ($\kappa^\star$ for the CCEF). *Suppose Assumptions 1.1, 1.2, and* IOS *hold, and that, for some* $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ *with probability one. Then:*

(i) For any g *satisfying* $E[|g(Y, T, X)|] < \infty$,

$$E[g(Y, T, X)\kappa^\star(T, Z, X)] = E[g(Y, T, X)1\{T^* \in \mathbf{S}\}]. \tag{21}$$

(ii) Suppose that in addition to the conditions for part (i), RIO *holds and* $g(Y,T,X) = (Y - F(T,X;b))^2$ *satisfies* $E[|g(Y,T,X)|] < \infty$ *for all* b*. Then, the CCEF, denoted* $C(T,X)$*, equals or is approximated by* $F(T,X;\beta)$*, where*

$$\beta \equiv \arg\min_b E[(C(T,X) - F(T,X;b))^2 | T^* \in \mathbf{S}] \tag{22}$$
$$= \arg\min_b E[(Y - F(T,X;b))^2 \kappa^\star(T,Z,X)]$$

and the solution to (22) *is assumed to be unique.*

(iii) If $\mathbf{T_s} = \emptyset$*, then:*

$$\kappa^\star(T,Z,X) = 1 - \frac{V(1-Z)}{P(Z=0|X)} - \frac{(1-V)Z}{P(Z=1|X)}$$

for $V = 1\{T \in \mathbf{T_i}\} + 1\{T \in \mathbf{T_{io}}\}Z$*.*

Part (i) of Theorem 2.5 says that $\kappa^\star$-weighting delivers identification of the expectation of any function of (Y,T,X) for switchers. This result holds without RIO and enables identification of, for instance, the weights used to construct SLATE in Theorem 2.3(ii).[10] Theorem 2.5(ii) uses $\kappa^\star$-weights and RIO to identify the CCEF, or to obtain an MMSE approximation to it. For a saturated linear CCEF specifications without covariates, the first order conditions for the minimization problem in the theorem imply uniqueness of β as long as the probabilities in Lemma 2.2 are nonzero.

The third part of Theorem 2.5 says that in models in which every treatment value $t \in \mathbf{T}$ is selected by some switcher, so $\mathbf{T_s} = \emptyset$, the $\kappa^\star$-weights used in (22) are of the form derived by Abadie (2003) for basic LATE. Here, however, a conventional treatment dummy is replaced by auxiliary treatment variable V that depends on (non-Bernoulli) T and, possibly, instrument Z.

From SLATE to the CCEF with an Unordered Treatment

Head Start is a large and long-running federally-funded program providing free preschool to disadvantaged children. An investigation of Head Start enrollment effects on test scores inspired by Kline and Walters (2016) illustrates identification of causal effects using Theorems 2.3 and 2.5. The instrument here is a randomized offer of Head Start placement in a sample of three- and four-year-old Head Start applicants participating in the Head Start Impact Study (HSIS). The outcome is a composite test score constructed from achievement tests taken after random assignment.

Randomized offers of Head Start seats increase the probability of Head Start enrollment in the year after random assignment by around 68 percentage points. This is also an estimate of the overall probability of switching. Importantly, Head Start switchers (defined as children placed in Head Start as a consequence of these offers) are drawn from two groups: applicants enrolled in competing non-Head Start preschools and applicants in home care (meaning no preschool). This leads to a scenario with a trinomial unordered treatment variable that we represent by $T \in \{h, c, n\}$, where h denotes Head Start attendance, c denotes competing preschool attendance, and n denotes home care. All HSIS children fall into one of these three groups.

[10] To do this, set $g(Y,T,X)$ equal to $h_0(X)$ and $h_1(X)$.

Kline and Walters (2016) and Feller et al. (2016) assume Head Start offers leave the choice between c and n unchanged since both options are available to everyone regardless of random assignment. In our notation, this can be written as:

$$T^*(1) \neq T^*(0) \Rightarrow T^*(1) = h. \tag{23}$$

This EMCO-like restriction leaves two types of switchers, $T^* = (n, h)$ and $T^* = (c, h)$. These switchers generate two corresponding Head Start effects, $E[Y^*(h) - Y^*(c)|T^* = (c, h)]$ for switchers out of competing preschools and $E[Y^*(h) - Y^*(n)|T^* = (n, h)]$ for switchers out of home care. Response types also include stayers at all three levels (a few applicants not offered Head Start manage to find their way to a Head Start program anyway). Restriction (23) is sufficient for IOS since $\texttt{out}(h)$, $\texttt{in}(c)$, and $\texttt{in}(n)$ are empty.

Ignoring covariates and applying Theorem 2.3(i), SLATE equals

$$\tau(\mathbf{S}) = \frac{E[Y|Z=1] - E[Y|Z=0]}{E[V|Z=1] - E[V|Z=0]},$$

where

$$\begin{aligned} V &= 1\{T \in \mathbf{T}_{\texttt{i}}\} + 1\{T \in \mathbf{T}_{\texttt{io}}\}Z \\ &= 1\{T = h\}, \end{aligned}$$

since $\mathbf{T}_{\texttt{io}}$ is empty. In this case, therefore, $\tau(\mathbf{S})$ is a conventional Wald-IV estimand instrumenting $1\{T = h\}$. As noted in Kline and Walters (2016) and Feller et al. (2016), this IV estimand is a weighted average of $E[Y^*(h) - Y^*(n)|T^* = (n, h)]$ and $E[Y^*(h) - Y^*(c)|T^* = (c, h)]$. The weights are given by the share of each switcher type among switchers.

Our framework disentangles switcher-specific average causal effects by adding RIO to (23). Since $t = h$ is the only treatment level selected by both types of switchers, Corollary 2.1 implies that RIO is equivalent to the restriction

$$E[Y^*(h)|T^* = (c, h)] = E[Y^*(h)|T^* = (n, h)]. \tag{24}$$

This says that average outcomes in the Head Start treatment are the same for both switcher types, while allowing unrestricted counterfactuals. In the same spirit, Kline and Walters (2016) restricts the relationship between selection bias and covariates.

Given (24), RIO holds and the CCEF is therefore a function of T only:

$$F(T; \beta) = 1\{T = h\}\beta_h + 1\{T = c\}\beta_c + 1\{T = n\}\beta_n.$$

Since IOS is satisfied, Theorem 2.5(ii) implies $\beta \equiv (\beta_h, \beta_c, \beta_n)$ can be obtained as

$$\beta = \arg\min_b E[(Y - F(T; b))^2 \kappa^\star(T, Z)]. \tag{25}$$

Moreover, because $V = 1\{T = h\}$, Theorem 2.5(iii) implies that $\kappa^\star$-weights simplify to

$$\kappa^\star(T, Z) = 1 - \frac{1\{T = h\}(1 - Z)}{P(Z = 0)} - \frac{1\{T \neq h\}Z}{P(Z = 1)}.$$

These match Abadie (2003) κ-weights for Bernoulli treatment $1\{T = h\}$.

The CCEF identified by (25) generates causal effects for the two switcher types:

$$E[Y^*(h) - Y^*(c)|T^* = (c,h)] = \beta_h - \beta_c$$
$$E[Y^*(h) - Y^*(n)|T^* = (n,h)] = \beta_h - \beta_n.$$

RIO notwithstanding, this parameterization allows causal effects of Head Start to differ for switchers into Head Start from home and switchers into Head Start from competing preschools. This feature of our setup holds more generally because RIO leaves potential outcomes seen for only one switcher type unrestricted. In the Head Start study, switcher types are distinguished by the (unique) nature of their no-Head Start counterfactuals.

An IV interpretation of $\kappa^\star$-weighting illuminates the identifying information that drives these results. Setting $p = P(Z = 1)$, the first order conditions for the $\kappa^\star$-weighted least squares problem in (25) can be written:

$$\begin{aligned} 0 &= E[(Z-p)1\{T = t\}(Y - \beta_t)] \\ &= E[(Z-p)1\{T = t\}Y] - \beta_t E[(Z-p)1\{T = t\}], \end{aligned}$$

implying

$$\beta_t = \frac{E[(Z-p)1\{T = t\}Y]}{E[(Z-p)1\{T = t\}]}, \tag{26}$$

for each $t \in \{h, c, n\}$. Equation (26) casts each β_t as an instrumental variable estimand with outcome $1\{T = t\}Y$ and treatment $1\{T = t\}$, instrumented by Z.

Given IOS, Lemma 2.2 ensures that the first stages in this context (the denominators of (26) at each t, divided by the variance of Z) identify the probability switchers select treatment level t for any $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}$. Applying the lemma to Head Start,

$$E[1\{T = h\}|Z = 1] - E[1\{T = h\}|Z = 0] = P(T^* \in \mathtt{in}(h)) = P(T^* \in \mathbf{S}),$$

where the second equality follows from the fact that all switchers switch in to h. Similarly,

$$E[1\{T = c\}|Z = 1] - E[1\{T = c\}|Z = 0] = -P(T^* \in \mathtt{out}(c)) = -P(T^* = (c,h)),$$

and likewise for $T^* = (n, h)$. In this application, each of the relevant identified switcher sets includes only one type; the share of each type is therefore identified. Note also that both switcher types must occur with positive probability for the full set of β_t to be identified. Among the 68% of children in the Head Start study who are switchers, we show below that a little over one-third of these come from competing preschools.

2.3 Relaxing RIO

In requiring (20), RIO rules out some types of selection bias. RIO can be partly relaxed, however, by allowing the CCEF to be a function of the instrument as well as of T and X. We dub this more flexible CCEF restriction ARIO, a mnemonic for "augmented" RIO. This condition is formalized by:

Assumption ARIO *There is a function $C_a(T, Z, X)$ satisfying*

$$P(E[Y|T, T^*, X] = C_a(T, Z, X)|T^* \in \mathbf{S}) = 1.$$

Although the CCEF defined in Proposition 1.1 does not condition on Z, we show below that interactions between T and Z contain information regarding T^* for specifications of $\mathbf{M}$ such that $\mathbf{T}_{\texttt{io}}$ is non-empty. This information is useful because it allows CCEF dependence on T^*.

The following corollary to Proposition 1.1 characterizes the more limited homogeneity imposed by ARIO.

Corollary 2.2 (Interpreting ARIO)**.** *Suppose Assumptions 1.1 and 1.2 hold, and that $P(T^* = t^*|X) > 0$ for all $t^* \in \mathbf{S}$ with probability one. Then:*

(i) ARIO *is satisfied if and only if:*

$$
\begin{aligned}
E[Y^*(t)|T^* = t^*, X] &= E[Y^*(t)|T^* \in \texttt{in}(\mathrm{t}), X] \qquad \text{for all } t^* \in \texttt{in}(\mathrm{t}) \\
E[Y^*(t)|T^* = t^*, X] &= E[Y^*(t)|T^* \in \texttt{out}(\mathrm{t}), X] \qquad \text{for all } t^* \in \texttt{out}(\mathrm{t})\,.
\end{aligned}
\tag{27}
$$

(ii) If ARIO *holds, covariate-specific causal effects for any $t^* = (t^*(0), t^*(1)) \in \mathbf{S}$ equal*

$$
E[Y^*(T^*(1)) - Y^*(T^*(0))|T^* = t^*, X] = C_a(t^*(1), 1, X) - C_a(t^*(0), 0, X). \tag{28}
$$

Moreover, ARIO *implies there are functions $C_1(T, X)$ and $C_2(T, Z, X)$ such that*

$$
P(C_a(T, Z, X) = C_1(T, X)1\{T \in \mathbf{T}_{\texttt{i}} \cup \mathbf{T}_{\texttt{o}}\} + C_2(T, Z, X)1\{T \in \mathbf{T}_{\texttt{io}}\}|T^* \in \mathbf{S}) = 1.
$$

The first part of the corollary shows that ARIO differs from RIO when there are treatment levels t such that some switcher types are observed at $T = t$ when the instrument is switched on and some switcher types are observed at $T = t$ when the instrument is switched off. RIO restricts mean potential outcomes at this level to be the same across these switcher types while ARIO does not. In other words, ARIO drops restrictions across switcher types for treatment values in $\mathbf{T}_{\texttt{io}}$. Suppose, for instance, $T \in \{0, 1, 2\}$ under monotonicity but not EMCO, ignoring covariates. Then $\texttt{in}(1) = \{(0, 1)\}$, $\texttt{out}(1) = \{(1, 2)\}$, and $\mathbf{T}_{\texttt{io}} = \{1\}$. By Corollary 2.1, RIO requires

$$
E[Y^*(1)|T^* = (0, 1)] = E[Y^*(1)|T^* = (1, 2)], \tag{29}
$$

because the union of $\texttt{in}(1)$ and $\texttt{out}(1)$ equals $\{(0, 1), (1, 2)\}$. ARIO, by contrast, relaxes this restriction. Because $\texttt{out}(0) = \{(0, 1), (0, 2)\}$ and $\texttt{in}(2) = \{(0, 2), (1, 2)\}$ while $\texttt{in}(0)$ and $\texttt{out}(2)$ are empty, Corollary 2.2 implies ARIO here requires only

$$
\begin{aligned}
E[Y^*(0)|T^* = (0, 2)] &= E[Y^*(0)|T^* = (0, 1)] \\
E[Y^*(2)|T^* = (0, 2)] &= E[Y^*(2)|T^* = (1, 2)],
\end{aligned}
$$

as does RIO.

The second part of Corollary 2.2 explains how the CCEF can be used to derive switcher-type specific average causal effects under ARIO. This part of the corollary also shows that under ARIO, the CCEF for switchers depends on the instrument only at values of the treatment in $\mathbf{T}_{\texttt{io}}$. We therefore focus on CCEF specifications satisfying

$$
F_a(T, Z, X; \beta) = F_1(T, X; \beta)1\{T \in \mathbf{T}_{\texttt{i}} \cup \mathbf{T}_{\texttt{o}}\} + F_2(T, Z, X; \beta)1\{T \in \mathbf{T}_{\texttt{io}}\}, \tag{30}
$$

for functions F_1 and F_2.

Instrument-treatment interactions distinguish between switcher types that select treatment levels $t \in \mathbf{T}_{\mathtt{io}}$. Returning to the scenario where $T \in \{0, 1, 2\}$ under monotonicity without EMCO, treatment level $T = 1$ is chosen by two types of switchers, $T^* = (0, 1)$ and $T^* = (1, 2)$. Conditional on T^* being a switcher, the interaction term $1\{T = 1\}(1 - Z)$ indicates type $T^* = (1, 2)$ since this type of switcher has treatment level 1 when the instrument is switched off. Likewise, the interaction term $1\{T = 1\}Z$ indicates a switcher of type $T^* = (0, 1)$ since this type of switcher has treatment level 1 when the instrument is switched on.

CCEF specifications satisfying (30) are identified as follows.

Theorem 2.6 (*Z*-augmented Identification)**.** *Assume 1.1, 1.2, and* IOS *hold, that F_a satisfies (30), and that for some $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ with probability one. Then,*

$$E[(Y - F_a(T, Z, X; b))^2 \kappa^\star(T, Z, X)] = E[(Y - F_a(T, Z, X; b))^2 1\{T^* \in \mathbf{S}\}] \tag{31}$$

provided $E[(Y - F_a(T, Z, X; b))^2] < \infty$ for all b. Moreover, given ARIO*, the CCEF equals or is approximated by $F_a(T, Z, X; \beta)$, where β is the (assumed to be unique) solution to*

$$\begin{aligned} \beta &\equiv \arg\min_b E[(C_a(T, Z, X) - F_a(T, Z, X; b))^2 | T^* \in \mathbf{S}] \\ &= \arg\min_b E[(Y - F_a(T, Z, X; b))^2 \kappa^\star(T, Z, X)]. \end{aligned} \tag{32}$$

Absent covariates, ARIO and Corollary 2.2(ii) ensure that the CCEF takes the form of a linear model saturated in T with a full set of T, Z interactions for treatment levels in $\mathbf{T}_{\mathtt{io}}$. Theorem 2.6 delivers identification of the coefficients in such a specification, and hence of the CCEF, by $\kappa^\star$-weighted least squares. With covariates, F_a is not guaranteed to match the CCEF unless covariates are discrete and F_a is fully saturated in X, including X interactions with T and TZ. When F_a does not match the CCEF, Theorem 2.6 provides an MMSE approximation in the class of functions to which $F_a(T, Z, X; b)$ belongs.

2.4 IV and $\kappa^\star$ Estimands for the CCEF

Under ARIO, the CCEF can be constructed from a set of IV estimands. This result, which uses Theorem 2.6 and the first-order conditions for $\kappa^\star$-weighted minimization to generalize (26), is formalized below.

Theorem 2.7 (CCEF IV)**.** *Suppose Assumptions 1.1, 1.2, and* IOS *hold, and that, for some $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ with probability one. Assume also that $P(T^* = t^*|X) > 0$ for all $t^* \in \mathbf{S}$ with probability one and let $\tilde{Z}$ be as defined in (7).*

(i) Suppose ARIO *holds. Then, for any $t^* \in \mathbf{S}$ and $t \in \{t^*(0), t^*(1)\}$,*

$$E[Y^*(t)|T^* = t^*, X] = \frac{E[\tilde{Z} \times 1\{T = t\}Y|X]}{E[\tilde{Z} \times 1\{T = t\}|X]} \qquad \text{if } t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}$$

$$E[Y^*(t)|T^* = t^*, X] = \frac{E[\tilde{Z} \times 1\{T = t\}ZY|X]}{E[\tilde{Z} \times 1\{T = t\}Z|X]} \qquad \text{if } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \mathtt{in}(\mathrm{t})$$

$$E[Y^*(t)|T^* = t^*, X] = \frac{E[\tilde{Z} \times 1\{T = t\}(1 - Z)Y|X]}{E[\tilde{Z} \times 1\{T = t\}(1 - Z)|X]} \qquad \text{if } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \mathtt{out}(\mathrm{t})$$

(ii) Replace ARIO *with the following unconditional analog:*

$$E[Y^*(t)|T^* = t^*] = E[Y^*(t)|T^* \in \texttt{in}(\text{t})] \quad \text{for all } t^* \in \texttt{in}(\text{t}) \tag{33}$$
$$E[Y^*(t)|T^* = t^*] = E[Y^*(t)|T^* \in \texttt{out}(\text{t})] \quad \text{for all } t^* \in \texttt{out}(\text{t})\,.$$

Then, for any $t^* \in \mathbf{S}$ *and* $t \in \{t^*(0), t^*(1)\}$,

$$E[Y^*(t)|T^* = t^*] = \frac{E[\tilde{Z} \times 1\{T = t\}Y]}{E[\tilde{Z} \times 1\{T = t\}]} \qquad \textit{if } t \in \mathbf{T}_{\texttt{i}} \cup \mathbf{T}_{\texttt{o}}$$
$$E[Y^*(t)|T^* = t^*] = \frac{E[\tilde{Z} \times 1\{T = t\}ZY]}{E[\tilde{Z} \times 1\{T = t\}Z]} \qquad \textit{if } t \in \mathbf{T}_{\texttt{io}},\ t^* \in \texttt{in}(\text{t})$$
$$E[Y^*(t)|T^* = t^*] = \frac{E[\tilde{Z} \times 1\{T = t\}(1-Z)Y]}{E[\tilde{Z} \times 1\{T = t\}(1-Z)]} \qquad \textit{if } t \in \mathbf{T}_{\texttt{io}},\ t^* \in \texttt{out}(\text{t})$$

Part (i) of Theorem 2.7 shows how the CCEF can be constructed from IV estimands that condition on X. The first of these estimands, identifying the CCEF at levels $t \in \mathbf{T}_{\texttt{i}} \cup \mathbf{T}_{\texttt{o}}$ for all switchers that select t, uses $\tilde{Z}$ to instrument $1\{T = t\}$ in an estimating equation with dependent variable $1\{T = t\}Y$. The second, identifying the CCEF at levels $t \in \mathbf{T}_{\texttt{io}}$ for response types $t^* \in \texttt{in}(\text{t})$, instruments $1\{T = t\}Z$ in an estimating equation with dependent variable $1\{T = t\}ZY$. The third, identifying the CCEF at levels $t \in \mathbf{T}_{\texttt{io}}$ for response types $t^* \in \texttt{out}(\text{t})$, instruments $1\{T = t\}(1-Z)$ in an estimating equation with dependent variable $1\{T = t\}(1-Z)Y$.

The second part of the theorem lays the foundation for IV identification of unconditional (on covariates) average treatment effects of the form

$$\tau(t^*) = E[Y^*(T^*(1)) - Y^*(T^*(0))|T^* = t^*].$$

These effects (introduced in equation (2)) depend on the marginalized (over X) CCEF, evaluated at a particular value $T^* = t^*$:

$$E[Y^*(t)|T^* = t^*] = E(\underbrace{E[Y^*(t)|T^* = t^*, X]}_{\text{CCEF}}|T^* = t^*).$$

This marginalization weights the CCEF at each value of X by the distribution of X for a *particular* switcher type, t^*. However, Theorems 2.5 and 2.6 identify the distribution of X for the set of all switchers rather than for specific switcher types. This problem is solved by replacing ARIO with (33) in Theorem 2.7(ii).

Back to $\kappa^\star$

Many causal investigations feature an exploration of variations in impact with covariates. IV estimators based on Theorem 2.7(ii), however, leave the role of covariates in the CCEF unspecified (instrument $\tilde{Z}$ is uncorrelated with X by construction). Moreover, neither conditional nor unconditional IV estimands in Theorem 2.7 lend themselves to imposition of parametric restrictions on the CCEF or to nonlinear models for the CCEF like probit or logit.

The $\kappa^\star$-weighted least squares estimand in Theorem 2.6 facilitates identification and estimation of CCEF models and approximations that incorporate covariates. In addition to generating interactions that may be of substantive interest, covariates in F_a can boost the precision of

CCEF estimates, much as covariates in an OLS estimator boost precision by reducing residual variance. Moreover, as noted in Abadie (2003), estimands of this form readily accommodate nonlinear models for the CCEF.

In practice, $\kappa^\star$-weighted least squares is complicated by the fact that $\kappa^\star$ can be negative, making the minimand non-convex. Theorem 2.7(i) motivates a computationally attractive solution to this problem. Define

$$\begin{aligned} \tilde{Y}_t &= \frac{\tilde{Z} \times 1\{T = t\}Y}{E[\tilde{Z} \times 1\{T = t\}|X]} \\ \tilde{Y}_{t,0} &= \frac{\tilde{Z} \times 1\{T = t\}(1 - Z)Y}{E[\tilde{Z} \times 1\{T = t\}(1 - Z)|X]} \\ \tilde{Y}_{t,1} &= \frac{\tilde{Z} \times 1\{T = t\}ZY}{E[\tilde{Z} \times 1\{T = t\}Z|X]} \end{aligned} \tag{34}$$

and positive weights

$$\begin{aligned} \omega_t(X) &= P(Z = 1|X)(P(T = t|Z = 1, X) - P(T = t|Z = 0, X))_+ && \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ \omega_t(X) &= P(Z = 0|X)(P(T = t|Z = 0, X) - P(T = t|Z = 1, X))_+ && \text{if } t \in \mathbf{T}_{\mathtt{o}} \\ \omega_{t,z}(X) &= P(T = t, Z = z|X) && \text{if } t \in \mathbf{T}_{\mathtt{io}}, \end{aligned} \tag{35}$$

where $a_+ \equiv \max(0, a)$ for any a. The following theorem shows that β solves an ω-weighted least squares problem whenever F_a satisfies $F_a(T, Z, X; b) = F_1(T, X; b)1\{T \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}\} + F_2(T, Z, X; b)1\{T \in \mathbf{T}_{\mathtt{io}}\}$ for some F_1 and F_2.

Theorem 2.8. *Suppose the conditions of Theorems 2.6 and 2.7(i) are satisfied. Then,*

$$\begin{aligned} \beta = \arg\min_b \; E\Bigg[&\sum_{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}} (\tilde{Y}_t - F_1(t, X; b))^2 \omega_t(X) \\ &+ \sum_{t \in \mathbf{T}_{\mathtt{io}}, z \in \{0,1\}} (\tilde{Y}_{t,z} - F_2(t, z, X; b))^2 \omega_{t,z}(X) \Bigg] \end{aligned} \tag{36}$$

for parameter vector β defined in Theorem 2.6.

This result, which combines key relationships in our framework, is most easily seen when $\mathbf{T}_{\mathtt{io}}$ and $\mathbf{T}_{\mathtt{s}}$ are empty, so $\mathbf{T} = \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}$. Note first that Corollary 2.2 and Theorem 2.6 then imply

$$\begin{aligned} \beta &\equiv \arg\min_b E\left[(C_1(T, X) - F_1(T, X; b))^2 | T^* \in \mathbf{S}\right] \\ &= \arg\min_b E\left[\sum_{t \in \mathbf{T}} (C_1(t, X) - F_1(t, X; b))^2 E[1\{T = t\}\kappa^\star(t, Z, X)|X]\right]. \end{aligned} \tag{37}$$

Using the $\kappa^\star$ representation from Theorem 2.5(iii) to evaluate $E[1\{T = t\}\kappa^\star(T, Z, X)|X]$, we have that:

$$\begin{aligned} &E[1\{T = t\}\kappa^\star(T, Z, X)|X] \\ &\qquad = \begin{cases} P(Z = 1|X)[P(T = t|Z = 1, X) - P(T = t|Z = 0, X)] & \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ P(Z = 0|X)[P(T = t|Z = 0, X) - P(T = t|Z = 1, X)] & \text{if } t \in \mathbf{T}_{\mathtt{o}} \end{cases} \end{aligned} \tag{38}$$

Since the two quantities on the right-hand side of (38) are positive by Lemma 2.2, this establishes $E[1\{T = t\}\kappa^\star(T, Z, X)|X] = \omega_t(X)$. Using this to substitute for $E[1\{T = t\}\kappa^\star(T, Z, X)|X]$ in (37) gives

$$\beta = \arg\min_b E\left[\sum_{t\in\mathbf{T}} (C_1(t, X) - F_1(t, X; b))^2 \omega_t(X)\right] \tag{39}$$

$$= \arg\min_b E\left[\sum_{t\in\mathbf{T}} (\tilde{Y}_t - F_1(t, X; b))^2 \omega_t(X)\right], \tag{40}$$

where the final equality follows from the fact that Theorem 2.7(i) implies $E[\tilde{Y}_t|X] = C_1(t, X)$. The argument for the case where $\mathbf{T}_{\mathtt{io}}$ is not empty is similar, using $\omega_{t,z}(X) = E[1\{T = t, Z = z\}\kappa^\star(T, Z, X)|X]$ for any $t \in \mathbf{T}_{\mathtt{io}}$ to simplify the term involving $F_2(T, Z, X; b)1\{T \in \mathbf{T}_{\mathtt{io}}\}$.

Theorem 2.8 offers a novel solution to the problem of negative weights in an estimand for the full CCEF. This strategy can be compared with that in Abadie et al. (2002), which replaces κ-weights with theoretically non-negative $E[\kappa(T, Z, X)|T, X, Y]$ in a basic LATE setting. Nonparametric estimates of $E[\kappa(T, Z, X)|T, X, Y]$ can be sensitive to the choice of smoothing parameters when Y is continuous. Theorem 2.8, by contrast, requires only estimates of the distribution of (Z, T) given X. Since Z and T are discrete, this estimation is relatively straightforward when X is also discrete.

CCEF Identification Under Ordered Selection

The added flexibility afforded by ARIO is evident in a scenario with $T \in \{0, 1, 2\}$, no covariates, and ordered selection model (15). This model satisfies ISP for all π, so SLATE is identified. Moreover, when $\pi \geq c_1 - c_0$ IOS also holds.

In combination, ARIO and IOS identify the CCEF. Assuming $\pi \geq c_1 - c_0$ so IOS holds, and that ε has large support, this model has two stayer types, $T^* \in \{(0, 0), (2, 2)\}$, and up to three switcher types, $T^* \in \{(0, 1), (0, 2), (1, 2)\}$. By Corollary 2.2(i), ARIO then requires

$$E[Y^*(0)|T^* = (0, 1)] = E[Y^*(0)|T^* = (0, 2)] \quad \text{(from (27) at } t = 0)$$
$$E[Y^*(2)|T^* = (0, 2)] = E[Y^*(2)|T^* = (1, 2)] \quad \text{(from (27) at } t = 2).$$

Since treatment level $t = 1$ is in $\mathbf{T}_{\mathtt{io}}$, RIO adds an additional restriction to these two:

$$E[Y^*(1)|T^* = (0, 1)] = E[Y^*(1)|T^* = (1, 2)] \quad \text{(from (20) at } t = 1).$$

Under ARIO, Corollary 2.2(ii) implies that the CCEF can be written

$$F_a(T, Z; \beta) = \underbrace{1\{T = 0\}\beta_0 + 1\{T = 2\}\beta_2}_{F_1(T;\beta)1\{T\in\mathbf{T}_{\mathtt{i}}\cup\mathbf{T}_{\mathtt{o}}\}} + \underbrace{1\{T = 1\}(\beta_1 + Z\beta_{11})}_{F_2(T,Z;\beta)1\{T\in\mathbf{T}_{\mathtt{io}}\}}, \tag{41}$$

since $\mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} = \{0, 2\}$ and $\mathbf{T}_{\mathtt{io}} = \{1\}$. This specification generates three causal effects:

$$E[Y^*(2) - Y^*(0)|T^* = (0, 2)] = F_a(2, 1; \beta) - F_a(0, 0; \beta) = \beta_2 - \beta_0 \tag{42}$$
$$E[Y^*(2) - Y^*(1)|T^* = (1, 2)] = F_a(2, 1; \beta) - F_a(1, 0; \beta) = \beta_2 - \beta_1$$
$$E[Y^*(1) - Y^*(0)|T^* = (0, 1)] = F_a(1, 1; \beta) - F_a(0, 0; \beta) = (\beta_1 + \beta_{11}) - \beta_0.$$

Note that effects for switchers from 0 to 1 and from 1 to 2 need not sum to the effect for switchers from 0 to 2.

Theorem 2.7 shows how to compute β using IV. Since $\mathbf{T}_{\mathtt{o}} = \{0\}$ and switcher types $T^* = (0,1)$ and $T^* = (0,2)$ can both select treatment level 0, we have:

$$E[Y^*(0)|T^* = (0,1)] = E[Y^*(0)|T^* = (0,2)] = \frac{E[(Z-p)1\{T=0\}Y]}{E[(Z-p)1\{T=0\}]} = \beta_0,$$

where $p = P(Z = 1)$. Similarly, since $\mathbf{T}_{\mathtt{i}} = \{2\}$ and switcher types $T^* = (0,2)$ and $T^* = (1,2)$ can both select treatment level 2, we also have:

$$E[Y^*(2)|T^* = (0,2)] = E[Y^*(2)|T^* = (1,2)] = \frac{E[(Z-p)1\{T=2\}Y]}{E[(Z-p)1\{T=2\}]} = \beta_2.$$

Finally, $\mathbf{T}_{\mathtt{io}} = \{1\}$ with switcher type $T^* = (0,1)$ belonging to $\mathtt{in}(1)$, and switcher type $T^* = (1,2)$ belonging to $\mathtt{out}(1)$. Theorem 2.7 therefore implies

$$E[Y^*(1)|T^* = (1,2)] = \frac{E[(Z-p)1\{T=1\}(1-Z)Y]}{E[(Z-p)1\{T=1\}(1-Z)]} = \beta_1$$
$$E[Y^*(1)|T^* = (0,1)] = \frac{E[(Z-p)1\{T=1\}ZY]}{E[(Z-p)1\{T=1\}Z]} = \beta_1 + \beta_{11}.$$

These IV estimands are easily rearranged to generate the causal effects of interest.

Theorems 2.6 and 2.8 allow us to identify and estimate the CCEF and associated causal effects by $\kappa^\star$-weighting. Since $\mathbf{T}_{\mathtt{i}} = \{2\}$ and $\mathbf{T}_{\mathtt{io}} = \{1\}$, V as written in Theorem 2.5(iii) equals $V = 1\{T = 2\} + 1\{T = 1\}Z$. The relevant $\kappa^\star$-weights therefore simplify to

$$\begin{aligned} \kappa^\star(T,Z) &= 1 - \frac{V(1-Z)}{P(Z=0)} - \frac{(1-V)Z}{P(Z=1)} \qquad (43) \\ &= 1 - \frac{1\{T=2\}(1-Z)}{P(Z=0)} - \frac{1\{T=0\}Z}{P(Z=1)}. \end{aligned}$$

Using these weights, Theorem 2.6 identifies β as the solution to

$$\beta = \arg\min_b E[(Y - F_a(T,Z;b))^2 \kappa^\star(T,Z)],$$

where F_a is as defined in (41). Theorem 2.8 provides a computationally attractive implementation of this minimand.

3 Applications

Three empirical applications illustrate our theoretical results. The first revisits the Head Start study discussed in Section 2.2. We compute both IV and $\kappa^\star$-weighted estimates of Head Start effects for two switcher types. The second analyzes data from the Angrist et al. (2022) post-secondary aid randomized trial. The effect of financial aid on degree completion is modeled as mediated by a pair of underlying treatments. The third application examines effects of child-bearing, an ordered treatment, on mothers' weeks worked. As in Angrist and Evans (1998), childbearing effects are identified using dummies for multiple second births and samesex sibships to instrument the number of children in households with at least two children. The family-size application features treatment levels in $\mathbf{T_{io}}$, so CCEF specifications that depend on Z are fruitful. The applications discussion in the text skips most estimation issues; Appendix A.3 details estimation procedures and establishes consistency and asymptotic normality of estimators based on the identification results in Section 2.

3.1 Multinomial Head Start Effects

Section 2.2 shows that SLATE for Head Start can be estimated using a dummy indicating randomly assigned offers of a place in Head Start (Z) to instrument a dummy for Head Start enrollment ($1\{T = h\}$). This IV estimand is a weighted average of $E[Y^*(h) - Y^*(n)|T^* = (n, h)]$ and $E[Y^*(h) - Y^*(c)|T^* = (c, h)]$. These are Head Start achievement effects for two switcher types: children drawn into Head Start from home and children drawn into Head Start from competing preschools.

Our HSIS estimation sample includes 3,571 Head Start applicants aged 3 and 4 at the time of application.[11] The outcome variable is a composite test score derived from tests taken in the first year following random assignment. Test scores are standardized to be mean zero and standard deviation one in the random assignment control group, separately by application cohort and age. The HSIS design was such that interested families applied to one of 383 Head Start centers and were either offered a seat there or turned away. The probability applicants were offered a place in Head Start varies by center, but estimates that control for center dummies are similar to those reported here.[12]

Conventional 2SLS estimates of the effects of Head Start enrollment using random assignment to instrument a dummy for Head Start enrollment appear in the first two columns of Table 1. These show that randomly assigned offers of a Head Start seat boosted Head Start enrollment by 68 percentage points. The 2SLS estimate generated by this first stage suggests Head Start enrollment yielded an achievement gain of 0.25 standard deviations; this replicates results from a comparable specification reported in Kline and Walters (2016) Table VIII. As in Kline and Walters (2016), these estimates control for dummies indicating applicant sex, race, Spanish language use, language of testing, age cohort, teen mother, mother's marital status, mother's level of education, presence of both parents in the home, family size, special education status, income quartile dummies, urban status, a cubic function of baseline scores, and indicators for assigned center transportation and quality (quality measured by center characteristics and practices).

In the context of our HSIS analysis, SLATE and 2SLS *estimators* differ in two ways: (i) SLATE instrument $\tilde{Z}$ is centered using a nonlinear model for $E[Z|X]$ (our applications specifies probit), while 2SLS uses a linear model to control for covariates; (ii) SLATE marginalizes covariate-specific reduced-form and first-stage estimates in the numerator and denominator of (8) using inverse propensity score weighting, while 2SLS implicitly averages covariate-specific

[11] These data are distributed by the Inter-university Consortium for Political and Social Research (ICPSR). See United States Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation (2018) for documentation.

[12] Applicants are coded as enrolling in Head Start if they enrolled any time between assignment and test-taking. Competing preschool enrollment is determined from parent and provider interviews. Kline and Walters (2016) details the analysis sample further.

LATEs, weighting these by the conditional variance of first-stage fitted values. In other words, the 2SLS weighting function estimates (or approximates) the variance of $E[D|Z,X]$ given X (Angrist and Imbens, 1995).

Estimates of the overall switching probability and of SLATE, reported in columns 3 and 4 of Panel A in Table 1, are virtually indistinguishable from the 2SLS estimates reported in the first two columns. The similarity here reflects the fact that $E[D|Z,X]$ and the assignment propensity score, $E[Z|X]$, vary little with X. This scenario makes the 2SLS conditional-variance weighting scheme over covariate-specific causal effects inconsequential.[13] κ^*-based estimates of SLATE, computed using Theorem 2.3 and reported in Panel B of Table 1, also equal the corresponding 2SLS estimates up to reported precision.

The restrictions on Head Start response types described in Section 2.2 satisfy IOS. Given RIO, these restrictions imply a CCEF with three values, generating distinct causal effects for two types of switchers conditional on X. In particular, assuming average *Head Start* outcomes conditional on X are the same for the two switcher types, the CCEF is a function of T and X only. This CCEF is identified by the IV estimands in Theorem 2.7(i). Provided the unconditional version of RIO given in (24) holds, the marginalized-over-covariates CCEF is identified by the IV estimands in Theorem 2.7(ii).

Marginalized CCEF estimates computed from the sample analog of the IV estimands in Theorem 2.7(ii), reported in Panel A of Table 1, indicate children in Head Start have the highest achievement levels, with average test scores of 0.24 standard deviations. This is followed by the average achievement of children in competing preschools at 0.094, with children in home care scoring well below the scores of children enrolled in either type of preschool. These estimates appear (along with standard errors) in column 5 of Table 1. As can be seen in column 7 of the table, roughly 24% of Head Start applicants are switchers drawn from competing preschool programs, while 44% are switchers drawn from home care. These quantities amount to 35% and 65% of all switchers, respectively.

Estimated causal effects of Head Start enrollment differ markedly across switcher types. The estimates in column 8 of Panel A in Table 1 indicate Head Start boosts achievement by roughly 0.14 for children drawn from competing programs, while gains for children switching in to Head Start from home care are estimated to be over twice as large at 0.31. Estimated Head Start effects for both switcher types are significantly different from zero; these estimates are also significantly different from each other. While qualitatively similar to the estimates of switcher-specific effects reported in Kline and Walters (2016), our estimates of these parameters are much more precise. The Kline and Walters (2016) control function estimates of the effect of Head Start for children drawn from competing programs aren't significantly different from zero, while we obtain an estimate of roughly 0.14 with a standard error around 0.054 for this group.[14]

IV estimates of SLATE and the CCEF are compatible in the sense that SLATE is a type-

[13] 2SLS weights can be written $V(E[D|Z,X]|X)$, where $V(\cdot|X)$ denotes conditional variance. When the first stage is additive in X and Z, this is proportional to $V(Z|X)$. Blandhol et al. (2022) highlights the fact that convex 2SLS weighting requires either saturated covariate control or a correct model for the assignment propensity score.

[14] Standard errors for estimates of $E[Y^*(h)-Y^*(c)|T^*=(c,h)]$ in Kline and Walters (2016) Table VIII, labeled "subLATEs", range from $0.15-0.47$. Feller et al. (2016) distinguishes switcher-specific effects using a parametric Bayesian framework, focusing on an outcome different from that examined here and in Kline and Walters (2016).

share weighted average of CCEF-implied causal effects for all response types. Appendix Lemma A.5 establishes this equivalence for SLATE and CCEF parameters, for the corresponding IV estimands, and for IV estimates of SLATE and the CCEF (assuming these are constructed using the same estimated propensity score). Consistent with this, the SLATE estimate of 0.25 in column 4 of Table 1 equals $.238 \times .136 + .443 \times .312$ divided by an overall probability of switching estimated at 0.68.

Estimates generated by minimizing the sample analog of the $\kappa^{\star}$-weighted least squares minimand in Theorem 2.8 are similar to the corresponding IV estimates (using a CCEF specification with additive control for covariates). In particular, $\kappa^{\star}$-based estimates of the CCEF and the corresponding causal effects by switcher type, reported in Panel B of Table 1, show the same order of achievement levels and much larger Head Start gains for children drawn from home care than for children drawn from competing preschool programs.[15] The $\kappa^{\star}$-based treatment effect estimated for switchers from competing preschools exceeds the corresponding IV estimate, though only by about one standard error. Estimates of the CCEF in Panel B are more precise than those in Panel A, though standard errors for the two pairs of estimated treatment effects are close.

The $\kappa^{\star}$-weighted estimand in Theorem 2.8 facilitates estimation of CCEF models allowing for interactions between switcher type and covariates. These estimates can then be transformed into causal effects using formula (19) under RIO or (28) under ARIO. In the context of Head Start, for instance, we'd like to know whether differences in impact across covariate groups are similar for both switcher types and whether differences in impact by type are similar across covariates. Such interactions are easily explored through restrictions on a CCEF approximation, $F(T, X; \beta)$, identified by $\kappa^{\star}$-weighting. This can be more straightforward than an IV strategy that estimates points in the support of $C(T, X)$ one at a time. In our application, $\kappa^{\star}$-weighting also yields a precision gain for estimates of the CCEF. This likely reflects the fact that the weighting estimator models CCEF dependence on covariates while the IV estimator does not.

Many evaluations of education interventions report differences in impact by sex.[16] CCEF estimates allowing differences in Head Start impact by sex suggest Head Start produces larger achievement gains for girls than for boys. These estimates appear in Panel A of Figure 1, which graphs estimated treatment effects separately by sex within switcher types. Sex differences in impact are around .16 for both types, though only the difference in impact for switchers from home care clears 10% significance. Differences in impact by whether English is the main language at home, graphed in Panel B of the figure, are more mixed. Gains for switchers from home care are largest for English-speakers (a marginally significant gap of around 0.21). But estimates for switchers from competing programs are similar for the two language groups. Finally, differences in impact by switcher type are similar in three out of four demographic groups. Statistical tests of whether differences in impact by switcher type differ by sex and

[15]To compute the weights and dependent variables used in Theorem 2.8, propensity score $E[Z|X]$ is estimated by probit and $P(T|Z, X)$ is estimated using multinomial probit. Probabilities $P(T, Z|X)$ and $P(T|X)$ used to construct the terms in (34) are obtained by manipulating these estimates. Both probit specifications are additive in explanatory variables. Standard errors and the p-value in Panel B come from a cluster bootstrap procedure (clustering on center), as discussed in Appendix A.3.

[16]For instance, Angrist and Lavy (2009) finds high school girls respond more than boys to financial incentives, while Gray-Lobe et al. (2023) reports larger long-term preschool effects for boys.

language fail to reject the null.

3.2 Multiple Mediators of Financial Aid Effects

Does post-secondary financial aid boost degree completion or simply transfer resources to students who would earn degrees anyway? This question motivated a large randomized trial that allocated full grant funding to public college and university students in Nebraska. In the trial, aid awards came from the Susan Thompson Buffett Foundation (STBF) Buffett Scholars program. Buffett Scholarships fund roughly 11% of Nebraska public university and college enrollment. A key outcome in this research, results of which are reported in Angrist et al. (2022), is BA completion.

Among STBF aid applicants who indicated a desire to attend four-year programs, STBF awards boosted BA completion in the treatment group by around eight percentage points. Since awards were randomly assigned (with groups or strata), this overall impact is easily estimated. It's harder to identify the mechanisms explaining why grant aid matters. On one hand, STBF awards might help aid recipients who've enrolled in a four-year degree programs stay on track, perhaps by reducing the need to work while in school. Grant aid also reduces student loans, thereby reducing worries related to increasing indebtedness while in school. Alternatively, or in addition, grant aid might encourage or allow Buffett scholars who would otherwise attend two-year schools or remain unenrolled to embark on a four-year program. Angrist et al. (2022) refers to this channel for aid effects as *early engagement*.

The early-engagement hypothesis is assessed here using a model that allows for two treatment variables that might mediate BA completion. The first treatment mediator indicates students who borrowed little in their first-year financial aid package (STBF awards are roughly constant across years for students in good academic standing). The second indicates Buffett applicants strongly engaged with a four-year degree program in the first year after random assignment. STBF awards affect both channels, reducing loans and boosting early engagement. We'd like to know whether aid awards boost BA completion when loans are reduced without changing early four-year engagement, or whether increased early engagement is *necessary* for loan reduction to have an impact.[17]

For the purposes of this investigation, loan reduction is coded as a dummy variable, denoted by D, indicating students who borrowed 1000$ or less as part of their first-year financial aid package. Roughly 46% of controls borrowed less than 1000$ in the first year after random assignment. STBF awards boosted this low-loan rate by roughly 25 percentage points. Early four-year engagement is coded as a dummy variable, denoted by S, indicating students who completed at least 24 credits at a four-year school (the STBF definition of a full course load) in the year after random assignment. STBF awards boosted this by roughly 10 percentage points, with a control-group base of 60%.[18]

This setting can be described by treatment vector $T = (D, S) \in \{0, 1\} \times \{0, 1\}$. Response

[17] Angrist et al. (2022) explores an alternative version of the early engagement hypothesis, without allowing for direct loan effects.

[18] Our analysis is limited to applicants who listed a four-year school as their first choice on the STBF aid application. In this sample, mean first-year loan amounts are $1366 and $4282 for treatment and control applicants, respectively.

types $T^*(z) = (D^*(z), S^*(z))$ are indexed against Bernoulli instrument Z indicating randomly assigned STBF awards. We distinguish D and S effects by restricting T^*. Specifically, we consider the identifying power of D-dominance, a restriction discussed in Navjeevan et al. (2023). In this context, D-dominance says that:

1. Awards weakly reduce student loans;
2. Awards weakly increase early engagement;
3. Awards increase early engagement only if they reduce student loans.

D-dominance therefore consists of monotonicity restrictions for each underlying treatment while also requiring awards to reduce loans if they are to boost early engagement. In other words, awards without financial consequences do not change behavior. This restriction is supported by the timing of STBF awards, which are announced before students commit to program type or course load.[19] In the notation of Assumption 1.2, D-dominance means:

$$\mathbf{M} \equiv \{T^* : D^*(1) \geq D^*(0),\ S^*(1) \geq S^*(0),\ \text{and}\ D^*(1) = D^*(0) \Rightarrow S^*(1) = S^*(0)\}.$$

Under D-dominance, IOS is satisfied. This can be seen by recoding the multinomial (D, S) treatment into a scalar treatment variable, T, as follows:

$$\begin{aligned} T &= 0 \textbf{ if } (D,S) = (0,0) \\ T &= 1 \textbf{ if } (D,S) = (1,0) \\ T &= 2 \textbf{ if } (D,S) = (0,1) \\ T &= 3 \textbf{ if } (D,S) = (1,1). \end{aligned}$$

D-dominance allows 3 types of switchers, $T^* \in \{(0,1),(2,3),(0,3)\}$ and four types of stayers, one for each value of T. IOS is satisfied because the sets $\texttt{in}(0), \texttt{out}(1), \texttt{in}(2), \texttt{out}(3)$ are empty.

Because IOS holds, Theorem 2.3(i) identifies SLATE by instrumenting $V = 1\{T \in \mathbf{T}_{\texttt{i}}\} + 1\{T \in \mathbf{T}_{\texttt{io}}\}Z$ with $\tilde{Z}$. Moreover, since $\mathbf{T}_{\texttt{io}}$ is empty, auxiliary treatment variable V simplifies to

$$V = 1\{T \in \mathbf{T}_{\texttt{i}}\} = D.$$

The second equality here is a consequence of the fact that $\mathbf{T}_{\texttt{i}}$ contains treatment values $T = 1$ and $T = 3$, and these are observed if and only if $D = 1$. In this application, SLATE is a weighted average of causal effects for the three switcher types observed under D-dominance. Specifically, SLATE averages effects of loan reduction without early engagement, effects of loan reduction for students who engage early regardless of award status, and an effect of combined loan reduction and early engagement. With covariates, SLATE averages over the covariate distribution as well as over these three switcher types.

As can be seen in Panel A of Table 2, the estimated (linear) first stage effect of Buffett awards on treatment variable D, computed with saturated controls for randomization strata, is around

[19] Nibbering and Oosterveen (Forthcoming) uses a version of this assumption to bound the effect of going from $(0,0)$ to $(1,1)$ for all compliers.

0.25. The corresponding 2SLS estimate, reported in column 3 of Table 2, is roughly 0.33.[20] As noted in our discussion of 2SLS for Head Start, with saturated controls for discrete strata (or when the assignment propensity score is additive), a conventional 2SLS estimate averages strata-specific conditional IV estimands using a conditional-variance weighting scheme.

Estimates of switching probabilities and SLATE, reported in columns 4-5 of Panel A in Table 2 are similar to the corresponding linear first stage and 2SLS estimates. These estimates are computed as described in Appendix A.3. Briefly, estimated SLATE is the sample analog of semiparametric IV estimand (8), with $E[Z|X]$ estimated by probit. Estimated $P(T^* \in \mathbf{S})$ is the denominator of this IV estimate. Standard errors for these estimates are computed using a generalized method of moments (GMM) procedure that accounts for sampling error in the estimated $E[Z|X]$; see Appendix A.3 for details. In principle, 2SLS estimates differ from the sample analog of (8) by virtue of the weights used to marginalize over X. In practice, the award propensity score is close to one-half across strata. This and the fact that $V = D$ likely explains the similarity of 2SLS and SLATE estimates in Table 2.

Results with baseline covariates added to the set of randomization strata controls appear in Panel B of Table 2 (the additional controls are dummies for the baseline characteristics listed in the table note). Covariates here are additive, meaning conditioning vector X in this case does not amount to a saturated parameterization. The addition of numerous baseline controls leaves both 2SLS and SLATE estimates virtually unchanged. This likely reflects the fact that, conditional on randomization strata, the instrument is independent of baseline covariates.

SLATE for effects of STBF financial aid is arguably harder to interpret than SLATE for Head Start. This reflects the fact that SLATE in the aid application averages the effects of two distinct treatments (D and S), while SLATE for Head Start averages effects of a common Head Start treatment relative to two possible counterfactuals. But because IOS identifies the CCEF, we can potentially distinguish switcher-specific effects. This distinction turns in part on RIO, which in this case says that:

$$E[Y^*(0)|T^* = (0,1)] = E[Y^*(0)|T^* = (0,3)]$$
$$E[Y^*(3)|T^* = (2,3)] = E[Y^*(3)|T^* = (0,3)].$$

As with Head Start, these restrictions equate potential outcome means at shared levels for switcher types that access these levels. For instance, RIO requires average $Y^*(0)$ to be the same for switcher types $T^* = (0,1)$ and $T^* = (0,3)$ since both switch out of 0. At the same time, RIO allows causal effects to differ for all three switcher types.[21]

The distribution of T, reported in column 2 of Table 3, shows that most STBF applicants either had reduced loans or engaged early with a four-year degree program. For many, however, only one or the other of these mediators is switched on, and roughly 12% neither reduced loans nor engaged early. CCEF estimates for each of the four treatments, computed using

[20]STBF awards were offered at different rates conditional on application year and the school applicants identified as the one they aspire to attend. These variables define randomization strata.

[21]The two RIO restrictions can be benchmarked against the range of CCEF specifications possible in this context. Three switcher types with two potential outcome levels identified for each yields six distinct potential outcome means when unrestricted. The most restrictive homogeneity assumption equates all of these for a total of five restrictions and treatment effects constant at zero. Non-zero constant effects imposes two restrictions on a particular linear combination of CCEF levels.

the IV estimands defined in Theorem 2.7, appear in column 3 of the table. Early engagement is associated with a high probability of four-year degree completion regardless of borrowing behavior. But students who borrow heavily without engaging (for whom $(D, S) = (0, 0)$) are substantially more likely to earn a BA than are students who borrow little without engaging ($(D, S) = (1, 0)$). Without covariates, the estimated rate of degree completion in the latter group is negative, though small and imprecise. The addition of baseline covariates moves this parameter estimate to a positive 0.023 while also increasing precision.

The distribution of switcher types indicates that roughly 10% of STBF applicants are induced by awards to both reduce loans and engage early. Estimated switching probabilities appear in column 5 of Table 3. Switchers induced to borrow less while engaging in a BA program whether or not they receive STBF awards are more prevalent, amounting to roughly 17% of the applicant population. But loan reduction without engagement is rare: the estimated prevalence of switcher types reducing loans without engagement is −0.025. It should be noted, however, that the fact that the estimate here is negative may signal problems with either D-dominance or with the independence restriction invoked in Assumption 1.1(iii).

Differences in CCEF levels for each switcher type give switcher-specific causal effects. Causal effects of STBF awards, reported in column 6 of Table 3, are broadly consistent with the early-engagement hypothesis. Estimates for switcher type $T^* = (0, 3)$ imply that STBF applicants induced by awards both to borrow less and to engage early saw their odds of degree completion increase by a little over 60 percentage points—relative to 0.22 when both channels are switched off. Induced loan reduction alone, however, did little for switcher type $T^* = (2, 3)$, the group of applicants who engage early with or without the benefit of an STBF award. These estimates are reasonably tight zeros, with standard errors of 0.036. As in Table 2, the addition of baseline covariates matters little for the estimates CCEF.

While the estimates for switcher types $T^* = (0, 3)$ and $T^* = (2, 3)$ are precisely estimated and offer a sharp contrast, results for switcher type $T^* = (0, 1)$ are muddier. Estimates for this group, which reduces borrowing without engaging early, are negative though not significantly different from zero. The imprecision of these estimates reflects the paucity of switcher switcher type $T^* = (0, 1)$. It seems noteworthy that control for covariates moderately boosts the precision of the IV estimate of the $T^* = (0, 1)$ effect, though the estimate remains imprecise.

3.3 Ordered Models for Childbearing Effects

Do falling fertility rates explain rising female labor force participation in the twentieth century? Births are strongly negatively correlated with mothers' work and earnings, but these correlations might reflect selection bias. Specifically, mothers in larger families may be predisposed to work less regardless of family size. Angrist and Evans (1998) tackles selection bias with two instrumental variables. The twins instrument indicates mothers who experienced multiple second births, while a samesex instrument indicates same-sex sibships at births one and two. Twin births increase family size sharply for the 1% of mothers who have twins, while samesex sibship pairs, which occur in roughly half of the population having two children, motivate further childbearing among parents who seek a diversified sibling-sex portfolio. Assuming twins and samesex instruments satisfy Assumption 1.1, they arguably identify causal effects of childbearing.

Angrist and Evans (1998) codes treatment as a dummy variable indicating mothers with more than two children in a sample of mothers having at least two. Among women with more than two children, however, fertility instruments can shift the distribution of family size at parities above two. Evidence on this point includes Angrist et al. (2010), which shows that the event of a multiple second birth moves some Israeli mothers from three to more than three children as well as from two to more than two. Similarly, samesex sibships at births one and two prompt some mothers to have more than three children in pursuit of a mixed-sex sibship. As first shown in Angrist and Imbens (1995), instrument-induced changes in the distribution of treatment can make IV estimates based on a Bernoulli treatment variable misleading.

These considerations lead us to consider identification strategies for effects of an ordered treatment, namely the number of children rather than a dummy for more than two. In this case, treatment $T \in \mathbf{T} \equiv \{2, \ldots, 5\}$, where $T = 5$ for women with five or more children (a rare event in the Angrist and Evans (1998) samples). For both twins and samesex instruments, potential treatment is assumed to satisfy monotonicity, that is,

$$T^*(1) \geq T^*(0).$$

This restriction allows six possible switcher types:

$$\mathbf{S} = \{(2,3), (2,4), (2,5), (3,4), (3,5), (4,5)\}.$$

An ordered probit model of treatment assignment reduces the number of latent response types, thereby identifying SLATE and, possibly, the CCEF. To this end, treatment status is assumed here to follow (15) with $\pi > 0$ and a normal error, extended to five levels of treatment.[22] Ordered probit first-stage estimates for twins and samesex instruments appear in Table 4. Estimates in even-numbered columns are from models that include as covariates mother's age, age at first birth, ages of first two children, and indicators for male first and second births, Black, Hispanic, and other race. With or without covariates, both instruments are estimated to induce large, precisely-estimated increases in family size. The first-stage coefficient for twins exceeds that for samesex by almost an order of magnitude.

The first-stage estimates in Table 4 rule out twins-induced switchers from two to five. This follows from the fact that type $T^* = (2,5)$ requires cutoffs such that $\varepsilon \leq c_3$ and $\pi + \varepsilon > c_5$. This event has positive probability in the ordered probit model if and only if $\pi > c_5 - c_3$, which doesn't hold for the estimates in columns 1-2. This leaves five switcher types for twins. On the other hand, since the estimated samesex π is less than than any consecutive difference in cutoffs, samesex sibships induce at most one additional child. This leaves up to three samesex switcher types, $(2,3), (3,4)$, and $(4,5)$. Lemma A.4 generalizes these observations, characterizing the set of switcher types for any latent-index ordered choice model where the latent variable has sufficiently large support.

As noted in Section 2.1, an ordered probit assignment scheme satisfies ISP. SLATE for both instruments is therefore identified and given by the IV estimand in Theorem 2.2. Formulas for auxiliary treatment variable V for each instrument follow from the first-stage estimates in

[22] This specification is in the spirit of Kline and Walters (2016), which uses a multinomial probit first stage in a model identified by a control function.

Table 4 and from Lemma A.4 in Appendix A.2. The estimated π for twins exceeds the largest difference in consecutive ordered probit cutoffs. Lemma therefore A.4(ii) implies:

$$
\begin{aligned}
V &= 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\}Z \\
&= 1\{T = 5\} + 1\{2 < T < 5\}Z
\end{aligned}
$$

for twins. At the same time, since the estimated samesex π is smaller than the minimum difference in consecutive ordered probit cutoffs, Lemma A.4(iii) says that samesex SLATE is identified by instrumenting $V = T$.[23]

Conventional 2SLS estimates of the effect of number of children on mothers' weeks worked, reported in the first two columns of Table 5, range from around -3.2 when estimated using twins and to roughly -4.4 when estimated using samesex. This difference in magnitudes mirrors that reported in Angrist and Evans (1998) and discussed in Angrist and Fernandez-Val (2013). Outside of a linear constant-effects world, however, it's unclear whether 2SLS can be seen as estimating weighted average marginal effects of the effects of additional childbearing. SLATE, by contrast, averages causal effects over switcher types. For twins, these types consist of a mix of switchers induced to have 1-2 additional children while samesex switchers increment childbearing by one child only.

As it turns out, the estimated SLATE for samesex switchers, reported in column 6 of Table 5, is indistinguishable from the corresponding 2SLS estimate. Since $V = T$ for samesex, the only difference between samesex SLATE and 2SLS is in the weighting scheme used to marginalize over covariates. But samesex is essentially orthogonal to covariates, so the covariate weighting scheme here is inconsequential. In contrast with samesex, twins auxiliary treatment variable V differs from T. Estimated twins SLATE is smaller in magnitude than the corresponding 2SLS estimate. This is explained by the fact that the overall probability of switching for twins, reported in the first row of column 5 in Table 5, exceeds the corresponding twins first stage.[24] Even so, the difference here is only about one standard error using the standard error associated with either estimate.

The first stage estimates in Table 4 imply that the set of possible switcher types for twins satisfies IOS. This is a consequence of the fact that π exceeds the largest consecutive difference in cutoffs, so there are no stayers at treatment levels other than endpoints 2 and 5. At the same time, since π is positive, no one is switched into 2 or out of 5. The upshot is that the CCEF for twins switchers is identified under ARIO by either IV as in Theorem 2.7 or using $\kappa^\star$ as in Theorem 2.8. ARIO supersedes RIO in this context because $\mathbf{T}_{\mathtt{io}}$ contains treatment levels 3 and 4. For these treatment levels, we obtain a pair of CCEF values according to whether switchers are of a type that has $T^* \in \mathtt{in}(\mathrm{t})$ or of a type that has $T^* \in \mathtt{out}(\mathrm{t})$.

[23] Since π is smaller than any consecutive difference in cutoffs, there are stayers at all treatment levels and samesex sibships induce one additional child among switchers. Therefore, $T^*(1) - T^*(0) = 0$ for stayers and $T^*(1) - T^*(0) = 1$ for switchers, which implies $\nu_1(T) = -\nu_0(T) = T$ satisfy the condition for ISP given in (6). This in turn means $V = \nu_1(T)Z - \nu_0(T)(1 - Z) = T$ in the IV estimand for SLATE.

[24] Conventional first-stage estimates can be written as a weighted average of the effect of the instrument on treatment for each switcher type, weighted by the probability of this type. Since switchers types are mutually exclusive and all types are induced to increase childbearing by at least 1 child, this average should exceed the estimated overall probability of switching. In this case, however, ordered probit (which we use to determine the set of switchers $\mathbf{S}$ and therefore the formula for V) generates something higher, a sign the ordered probit model may be misspecified.

Since two twins switcher types taking values in $\mathbf{T}_{\mathtt{io}}$, ARIO is the relevant homogeneity restriction for identification of the twins CCEF. With five switcher types accessing two levels each, the twins CCEF has 10 point of support. ARIO imposes four pairwise restrictions on these:

$$
\begin{aligned}
E[Y^*(2)|T^* = (2,3)] &= E[Y^*(2)|T^* = (2,4)] \\
E[Y^*(3)|T^* = (3,4)] &= E[Y^*(3)|T^* = (3,5)] \\
E[Y^*(4)|T^* = (2,4)] &= E[Y^*(4)|T^* = (3,4)] \\
E[Y^*(5)|T^* = (3,5)] &= E[Y^*(5)|T^* = (4,5)].
\end{aligned}
$$

Although switcher types $T^* = (2,3)$ and $T^* = (3,4)$ both access $T = 3$, one type belongs to $\mathtt{in}(3)$ and one type belongs to $\mathtt{out}(3)$. ARIO therefore leaves the difference in CCEF values for this pair unrestricted. Likewise, ARIO imposes no equality restriction on the pair $T^* = (2,3)$ and $T^* = (3,5)$ and the pair $T^* = (2,4)$ and $T^* = (4,5)$.

Estimates of the CCEF for twins instruments, reported in columns 3-4 of Table 6, show monotonically declining weeks worked as a function of family size for both switcher types. It's noteworthy that alternative estimates of the CCEF are similar at treatment levels common to types in $\mathbf{T}_{\mathtt{io}}$. For $T = 3$, for instance, the difference between $E[Y^*(3)|T^* \in \mathtt{in}(3)]$ and $E[Y^*(3)|T^* \in \mathtt{out}(3)]$ is less than a week, while the gap between $E[Y^*(4)|T^* \in \mathtt{in}(4)]$ and $E[Y^*(4)|T^* \in \mathtt{out}(4)]$ is even smaller. As can be seen in column 5 in the table, neither difference is significantly different from zero in a 5% test.

Table 7 summarizes the implications of Table 6 for individual treatment effects by switcher type. Estimated treatment effects are largest in magnitude for the two types induced to have two additional children by twinning, $T^* = (2,4)$ and $T^* = (3,5)$. These estimates, equality of which is tested in column 4, average -3.1 per child. Estimates for one-child switchers range from around -2 for $T^* = (3,4)$ to -3.8 for $T^* = (2,3)$ and average -3. A test of equality for these types, reported in column 4, is also consistent with constant effects. A test of equal per-child effects for all five switcher types clearly rejects the null hypothesis of full homogeneity. This is driven by the difference between -1.96 for $T^* = (3,4)$ and the per-child estimate of -3.3 for $T^* = (2,4)$. This difference aside, however, the estimates in Table 7 align around per-child effects of about -3.

4 Summary and Directions for Further Work

The theoretical framework developed here uses a single instrumental variable to identify average causal effects in diverse applications involving multifaceted treatments. Our identification results are founded in part on the concept of instrument-induced switchers, an extension of the ideas behind Angrist et al. (1996) compliers and Frangakis and Rubin (2002) principal strata. Switcher treatment status changes in response to instrument assignment. This includes changes in multinomial, multiple, and ordered treatments.

Our framework leverages restrictions on the distribution of switcher types to identify a single average causal effect for all switchers, a parameter we've dubbed SLATE. We also use a combination of restrictions on switcher types and average potential outcomes to identify a general CCEF. The CCEF characterizes average potential outcomes and average causal effects for specific switcher types conditional on covariates. SLATE and CCEF parameters are shown to be identified both by IV estimands and by a weighting scheme that generalizes Abadie (2003) κ-weights for multiple treatments.

The utility of these results is demonstrated in three applications. Building on Kline and Walters (2016), we first consider average causal effects of Head Start preschool attendance on test scores in a setting with multinomial preschool treatments. As in Kline and Walters (2016), the estimates reported here suggest Head Start generates substantially larger achievement gains for children drawn into Head Start from home care than for children drawn into Head Start from competing programs. At the same time, our estimates of Head Start effects for children who switch into Head Start from competing programs are notably more precise than those reported in Kline and Walters (2016). Estimates of CCEF specifications allowing for covariate interactions are easily computed using $\kappa^{\star}$-weighting to estimate parsimonious approximations to a higher-dimensional CCEF. These show larger gains from Head Start for switchers from home care than for switchers from competing programs across covariate groups. Moreover, consistently across switcher types, girls benefit more from Head Start than do boys.

Our second application allows for two distinct pathways through which post-secondary grant aid boosts BA completion. The resulting estimates suggest early engagement with four-year college programs is the primary channel by which financial aid boosts degree completion; loan reduction alone does not seem to matter. The third application identifies and estimates the CCEF for effects of childbearing, an ordered treatment, on mothers' labor supply. Estimates of the CCEF for effects of childbearing on weeks worked show roughly constant and approximately linear reductions in labor supply of around three weeks per child, across switcher types.

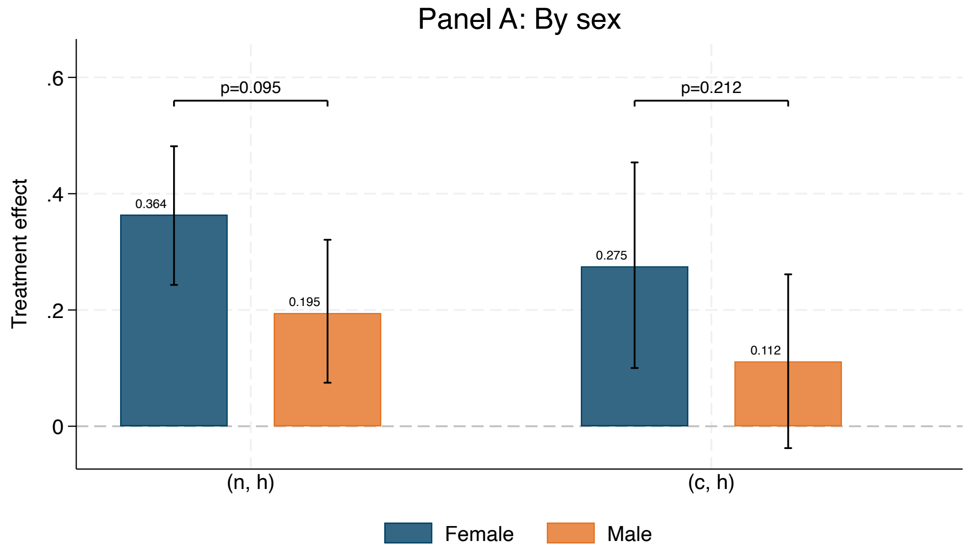
In future work, we plan to extend the framework developed here to exploit multiple instruments in models with multinomial and ordered treatments. Much work also remains to be done regarding estimation and specification testing in these settings. In addition to questions of efficiency, the homogeneity restrictions (RIO and ARIO) and restrictions on switcher types like D-dominance seem likely to lead to new tests of key identifying assumptions. These tests should be useful in a wide range of IV applications.

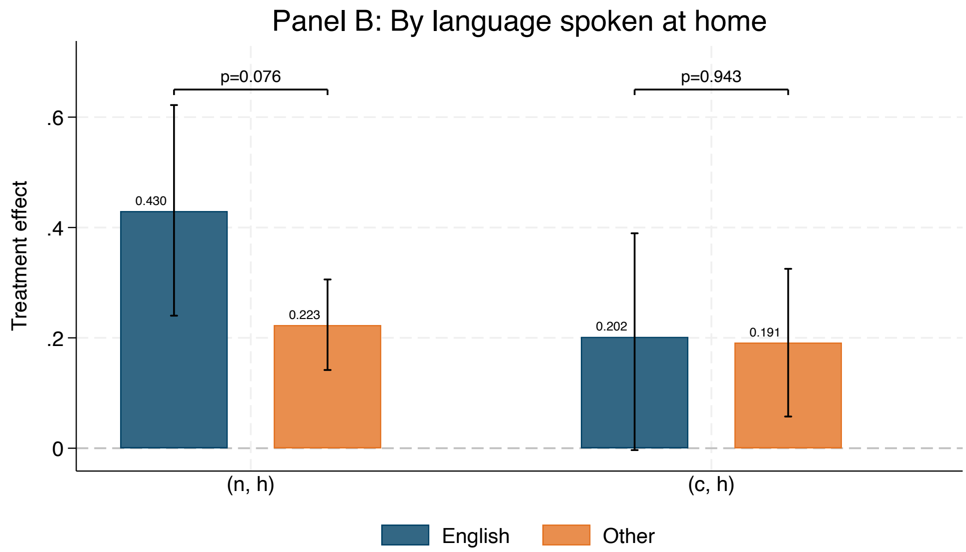
Table 1: Estimates of Head Start Effects on Test Scores

Treatment	2SLS First Stage (1)	2SLS Estimate (2)	SLATE $P(T^* \in \mathbf{S})$ (3)	SLATE Estimate (4)	CCEF (5)	Causal Effects Switcher Types (6)	Causal Effects $P(T^* = t^*)$ (7)	Causal Effects Estimate (8)
Panel A. IV-based estimates								
Head Start	0.679	0.250	0.680	0.250	0.230			
	(0.018)	(0.030)	(0.018)	(0.030)	(0.033)			
Competing preschools					0.094	(c, h)	0.238	0.136
					(0.057)		(0.015)	(0.054)
Home care					-0.081	(n, h)	0.443	0.312
					(0.048)		(0.020)	(0.039)
p-value for impact difference by switcher type								0.012
Panel B. $\kappa^\star$-based estimates								
Head Start			0.680	0.250	0.238			
			(0.018)	(0.030)	(0.024)			
Competing preschools					0.043	(c, h)	0.238	0.194
					(0.051)		(0.015)	(0.054)
Home care					-0.042	(n, h)	0.443	0.280
					(0.034)		(0.020)	(0.036)
p-value for impact difference by switcher type								0.166

Notes: This table reports 2SLS, SLATE, and CCEF estimates characterizing the effect of Head Start enrollment on test scores in the first year following random assignment. Parameters are identified by random assignment of Head Start offers. Treatment variable $T \in \{n, c, h\}$ denotes no preschool enrollment (n), other center enrollment (c), or Head Start (h). Column 1 reports conventional first stage estimates, with the corresponding 2SLS estimates instrumenting $1\{T = h\}$ shown in column 2. SLATE averages effects for the two switcher types (T^*) listed in column 6. SLATE (column 4) and the overall switching probability (column 3) in Panel A are estimated using Theorem 2.2 and setting $V = 1\{T = h\}$, as described in Appendix A.3. SLATE (column 4) and the overall switching probability (column 3) in Panel B are estimated using Theorem 2.3 as described in Appendix A.3. The CCEF (column 5) in Panel A is estimated by IV using Theorem 2.7, as described in Appendix A.3. The CCEF (column 5) in Panel B is estimated by $\kappa^\star$-weighting using Theorem 2.8, as described in Appendix A.3. Column 6 lists possible switcher types and column 7 shows the estimated probability of switcher types; these are computed using Lemma 2.2 as described in Appendix A.3. Column 8 reports estimated causal effects on test scores for each type, computed from the CCEF estimates in column 5. The p-value in Panel B is computed using the cluster bootstrap (clustering on center) as described in Appendix A.3. All models control for applicant sex, race, Spanish language use, language of testing, age cohort, teen mother, mother's marital status, mother's level of education, presence of both parents in the home, family size, special education status, income quartile dummies, urban status, a cubic function of baseline scores, and indicators for assigned center transportation and quality (quality measured by center characteristics and practices). The propensity score is estimated using probit. $P(T = t|Z, X)$ is estimated using multinomial probit. Standard errors in columns 1-2 (reported in parentheses) cluster on center. Standard errors in columns 3-5 and 7-8 (Panel A) are computed as described in Appendix A.3, also clustering on center. Standard errors in columns 3-5 and 8 in Panel B are computed using a cluster bootstrap (1,000 simulations) detailed in Appendix A.3. $N = 3,571$.

Figure 1: Estimates of Head Start Effects on Test Scores by Sex and Primary Language





Notes: This figure graphs estimated causal effects of Head Start on test scores by switcher type and sex (in Panel A) and by switcher type and language spoken at home (in Panel B). Causal effects are computed from a CCEF estimated by $\kappa^{\star}$–weighting using Theorem 2.8, as described in Appendix A.3. P-values are for tests of differences in impact by demographic group within switcher types. Confidence intervals and p-values are computed using a cluster bootstrap (clustering on center), as discussed in Appendix A.3. The propensity score is estimated using probit; $P(T = t|Z, X)$ is estimated using multinomial probit. Controls are as for Table 1. N = 3,571.

Table 2: Estimates of Financial Aid Effects on BAs Pooling Switchers

		2SLS (instrumenting D)		SLATE ($V = D$)	
Instrument	Switcher Types	First Stage	Estimate	$P(T^* \in \mathbf{S})$	Estimate
	(1)	(2)	(3)	(4)	(5)
Panel A. Randomization strata controls					
Aid award	(0,1), (2,3), (0,3)	0.245	0.329	0.246	0.325
		(0.012)	(0.050)	(0.012)	(0.049)
Panel B. Baseline covariates + randomization strata controls					
		0.245	0.346	0.246	0.342
		(0.011)	(0.047)	(0.011)	(0.047)

Notes: This table reports 2SLS and SLATE estimates for the effects of financial aid on BA completion. Parameters are identified by the random assignment of STBF aid awards. Treatment $T \in \{0, 1, 2, 3\}$ recodes the pair of Bernoulli treatments $(D, S) \in \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ into four levels, where D indicates reduced loans and S indicates early engagement. 2SLS estimates use aid offers to instrument D. SLATE averages effects for the three switcher types (T^*) listed in column 1. SLATE (column 5) and overall switching probability (column 4) are estimated using Theorem 2.2, as described in Appendix A.3, setting $V = D$. Baseline covariates include dummies for Pell eligibility, white race, above-median high school GPA, above-median ACT, community college alternate, Omaha resident, and first-generation college student. The propensity score is estimated using probit. 2SLS standard errors, reported in parentheses, are robust. Standard errors in columns 4-5 are computed as described in Appendix A.3. $N = 6,845$.

Table 3: Estimates of the CCEF for Aid Effects on BA Completion

Treatment Level	(D, S) (1)	$P(T = t)$ (2)	CCEF (3)	Causal Effects: Switcher Type (4)	Causal Effects: $P(T^* = t^*)$ (5)	Causal Effects: IV Estimates (6)
Panel A. Randomization strata controls						
0	(0, 0)	0.124	0.223	(0, 3)	0.103	0.626
			(0.049)		(0.012)	(0.053)
1	(1, 0)	0.238	-0.005	(0, 1)	-0.025	-0.228
			(0.276)		(0.010)	(0.281)
2	(0, 1)	0.304	0.790	(2, 3)	0.168	0.059
			(0.030)		(0.011)	(0.036)
3	(1, 1)	0.334	0.849			
			(0.020)			
Panel B. Baseline covariates + randomization strata controls						
0	(0, 0)	0.124	0.222	(0, 3)	0.107	0.630
			(0.048)		(0.011)	(0.052)
1	(1, 0)	0.238	0.023	(0, 1)	-0.028	-0.200
			(0.238)		(0.010)	(0.244)
2	(0, 1)	0.304	0.787	(2, 3)	0.167	0.065
			(0.030)		(0.011)	(0.036)
3	(1, 1)	0.334	0.852			
			(0.020)			

Notes: This table reports the estimated CCEF and corresponding treatment effects for effects of reduced loans (D) and early engagement (S) on BA completion. Parameters are identified by the random assignment of STBF aid awards. Column 2 gives sample proportions at each treatment level. Column 3 shows the CCEF, estimated by IV using Theorem 2.7, as described in Appendix A.3. Column 4 lists possible switcher types under D-dominance and column 5 shows the estimated probability of switcher types; these are computed as described in Appendix A.3. Column 6 reports estimated causal effects on BA completion for each type, computed from the CCEF estimates in column 3. Baseline covariates are as for Table 2. The propensity score is estimated using probit. Standard errors for the estimates in columns 3, 5, and 6 are computed as described in Appendix A.3. $N = 6,845$.

Table 4: Ordered Probit First-Stage Estimates for Twins & Samesex Instruments

	Twins		Samesex	
Parameter	(1)	(2)	(3)	(4)
π	1.084	1.315	0.147	0.170
	(0.0114)	(0.0150)	(0.0047)	(0.0049)
Marginal Effects	0.870	0.901	0.089	0.087
	(0.0106)	(0.0119)	(0.0029)	(0.0025)
c_3	0.314	-0.308	0.379	-0.210
	(0.0025)	(0.0251)	(0.0035)	(0.0252)
c_4	1.324	0.886	1.385	0.976
	(0.0035)	(0.0252)	(0.0044)	(0.0253)
c_5	2.009	1.708	2.068	1.795
	(0.0055)	(0.0255)	(0.0061)	(0.0257)
$c_4 - c_3$	1.010	1.193	1.005	1.185
$c_5 - c_3$	1.695	2.015	1.689	2.004
$c_5 - c_4$	0.685	0.822	0.684	0.819
max $c_{t+1} - c_t$	1.010	1.193	1.005	1.185
min $c_{t+1} - c_t$	0.685	0.822	0.684	0.819
Covariates	No	Yes	No	Yes

Notes: This table reports estimates of the ordered choice model detailed in Lemma A.4, assuming ε is normally distributed. The dependent variable is the number of children (T) for mothers of two or more. The second row of the table reports average marginal effects of each instrument on $E[T|Z, X]$. Data are from the Angrist and Evans (1998) 1980 Census sample of 254,652 married mothers. Treatment $T \in \{2, 3, 4, 5\}$ is the number of children censored at 5. Controls are age, age at first birth, age of first two children (in quarters), and indicators for boy 1st, boy 2nd, black, hispanic, and other race. Robust standard errors are reported in parenthesis.

Table 5: Estimates of Childbearing Effects Pooling Switchers Using Twins & Samesex Instruments

		2SLS		SLATE		
Instrument	Switcher Types (1)	First Stage (2)	Estimate (3)	V (4)	$P(T^* \in \mathbf{S})$ (5)	Estimate (6)
Twins	$(2,3), (2,4), (3,4),$ $(3,5), (4,5)$	0.824 (0.0116)	-3.22 (0.535)	$1\{T = 5\}$ $+1\{2 < T < 5\}Z$	0.976 (0.0005)	-2.71 (0.455)
Samesex	$(2,3), (3,4), (4,5)$	0.083 (0.0026)	-4.36 (1.007)	$T - 2$	0.083 (0.0026)	-4.36 (1.008)

Notes: This table reports 2SLS and SLATE estimates of effects of childbearing on weeks worked. Parameters are identified using twins and samesex instruments. Treatment $T \in \{2, 3, 4, 5\}$ is the number of children censored at 5. 2SLS estimates use twin and samesex dummies to instrument T. SLATE averages effects for the three switcher types listed in column 1; switcher types follow from Lemma A.4(iv) and the ordered probit estimates in Table 4. Formulas for V in Column 4 follow from Lemma A.4(ii) for twins and A.4(iii) for samesex. SLATE (column 5) and overall switching probabilities (column 4) are estimated using Theorem 2.2, as described in Appendix A.3. Covariates and sample are as for Table 4. The propensity score is estimated using probit. 2SLS standard errors, reported in parentheses, are robust. The probability of switching and SLATE standard errors are computed as described in Appendix A.3. $N = 254,652$.

Table 6: The CCEF for Childbearing Effects on Weeks Worked

Treatment Level	Switcher Types (1)	$P(T=t)$ (2)	CCEF $\texttt{in}(t)$ (3)	CCEF $\texttt{out}(t)$ (4)	col 2− col 3 (5)
2	$(2,3),(2,4)$	0.619	$\emptyset$	21.1	–
				(0.06)	
3	$(2,3),(3,4),(3,5)$	0.285	17.3	16.4	0.918
			(0.52)	(0.08)	(0.5240)
4	$(2,4),(3,4),(4,5)$	0.072	14.4	14.0	0.438
			(1.00)	(0.15)	(1.0115)
5	$(3,5),(4,5)$	0.023	10.6	$\emptyset$	–
			(2.86)		

Notes: This table reports the estimated CCEF for effects of childbearing on weeks worked. Parameters are identified using the twins instrument. Treatment $T \in \{2,3,4,5\}$ is the number of children censored at 5. Column 1 reports switcher types getting into or out of t. Column 2 gives sample proportions at each treatment level. Columns 3 and 4 report CCEF estimates for switchers in the sets $\texttt{in(t)}$ and $\texttt{out(t)}$; these sets are obtained from Lemma A.4 and the ordered probit estimates in Table 4. CCEF is estimated by IV using Theorem 2.7, as described in Appendix A.3. Column 5 computes and tests CCEF differences for levels common to $\texttt{in(t)}$ and $\texttt{out(t)}$. The propensity score is estimated using probit. Covariates and sample are as for Table 4. Standard errors for the estimates in columns 3, 4, and 5 are computed as described in Appendix A.3. $N = 254,652$.

Table 7: CCEF-Derived Estimates of Effects of Childbearing on Weeks Worked

Switcher Type	Causal Effect (1)	Per-child Effect (2)	Average Effects: Switchers Included	Average Effects: Average (3)	Average Effects: Equality Test (4)
(2, 3)	-3.75	same as (1)	$(2,3),(3,4),(4,5)$	-3.03	0.296
	(0.521)			(1.020)	df=2
(3, 4)	-1.96	same as (1)	$(2,4),(3,5)$	-3.10	0.782
	(1.004)			(0.756)	df=1
(4, 5)	-3.39	same as (1)	all switchers	-3.06	0.000
	(1.004)			(0.911)	df=4
(2, 4)	-6.63	-3.31			
	(1.002)	(0.501)			
(3, 5)	-5.79	-2.89			
	(2.861)	(1.431)			

Notes: This table reports estimates of causal effects of childbearing on weeks worked. Parameters are identified using the twins instrument. Treatment $T \in \{2,3,4,5\}$ is the number of children censored at 5. Switcher types follow from Lemma A.4(iv) and the ordered probit estimates in Table 4. Column 1 reports causal effects estimates based on the CCEF estimates in Table 6. Column 2 reports per-child effects. Column 3 reports, respectively, average incremental effect for switcher types $(2,3)$, $(3,4)$, and $(4,5)$, average of per-child effects for two-child switchers, and average of per-child effects for all switchers; corresponding standard errors are in parentheses. Column 4 reports the p-value for the test that the underlying effects in the corresponding row in column 3 are all equal. Covariates and sample are as for Table 4. The propensity score is estimated using probit. Standard errors are computed as described in Appendix A.3. $N = 254,652$.

A Appendix

Identification and estimation results are proved in separate appendices.

A.1 Proofs for Section 1 and 2

A.1.1 Proof of Proposition 1.1

Using that $Y = \sum_{t \in \mathbf{T}} Y^*(t) 1\{T = t\}$ and $T = T^*(0)(1 - Z) + T^*(1)Z$, we can obtain that

$$E[Y1\{T = t\}1\{T^* = t^*\}|X] = E[Y^*(t)1\{t^*(Z) = t\}1\{T^* = t^*\}|X] \qquad (44)$$
$$= E[Y^*(t)1\{T^* = t^*\}|X]P(t^*(Z) = t|X),$$

for any $t \in \{t^*(0), t^*(1)\}$, where the second equality follows from the law of iterated expectations and Lemma A.1 in Navjeevan et al. (2023). Furthermore, by the same arguments we also have

$$E[1\{T^* = t^*\}1\{T = t\}|X] = E[1\{T^* = t^*\}|X]P(t^*(Z) = t|X). \qquad (45)$$

Therefore, the theorem follows from dividing (44) by (45) and employing that $P(t^*(Z) = t|X) > 0$ for any $t \in \{t^*(0), t^*(1)\}$ due to $0 < P(Z = 1|X) < 1$. ■

A.1.2 Proof of Lemma 2.1

First note that since $P(T^* \in \mathbf{M}) = 1$ by Assumption 1.2(i), for any $z \in \{0, 1\}$ it follows

$$E[Y|Z = z, X] = \sum_{t^* \in \mathbf{M}} E[Y1\{T^* = t^*\}|Z = z, X] \qquad (46)$$
$$= \sum_{t^* \in \mathbf{M}} E[Y^*(t^*(z))1\{T^* = t^*\}|Z = z, X] = \sum_{t^* \in \mathbf{M}} E[Y^*(t^*(z))1\{T^* = t^*\}|X]$$

where the second equality follows from Assumption 1.1(ii) and the third from (Y^*, T^*) being independent of Z conditionally on X. Therefore, since $t^*(1) = t^*(0)$ for any $t^* \in \mathbf{M} \setminus \mathbf{S}$, result (46) and the law of iterated expectations yield

$$E[E[Y|Z = 1, X] - E[Y|Z = 0, X]] = \sum_{t^* \in \mathbf{S}} E[(Y^*(t^*(1)) - Y^*(t^*(0)))1\{T^* = t^*\}] \qquad (47)$$
$$= E[\tau(T^*)1\{T^* \in \mathbf{S}\}],$$

which establishes the first claim of the lemma. The second claim is immediate by dividing both sides of (47) by $P(T^* \in \mathbf{S})$, which is positive by Assumption 1.2(ii). ■

A.1.3 Proof of Theorem 2.1

First define the function $\ell(T^*) \equiv 1\{T^* \in \mathbf{S}\}$ and observe that by definition we have

$$P(T^* \in \mathbf{S}) = E[\ell(T^*)]. \qquad (48)$$

Corollary 4.3 in Navjeevan et al. (2023) provides necessary and sufficient conditions for the identification of moments with the structure in (48). In order to apply said Corollary, let $U_{\mathbf{M}}$ denote the uniform distribution on $\mathbf{M}$ and set $\mu \equiv \{P_{Y|X}\}^d U_{\mathbf{M}} P_Z P_X$. By lemma A.1,

$P_{Y^*T^*ZX} \ll \mu$ and therefore Assumption 1.1 implies that Assumptions 2.1 and 2.2 of Navjeevan et al. (2023) are satisfied with $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$. Moreover, since Lemma A.1 further establishes that $\mu \ll P_{Y^*T^*ZX}$, we may set $\bar{Q} = P_{Y^*T^*ZX}$. Condition (ii) of the Theorem therefore implies that $\bar{Q}^{\texttt{it}} \ll \bar{Q}$ with $d\bar{Q}^{\texttt{it}}/d\bar{Q}$ bounded as required by Corollary 4.3 in Navjeevan et al. (2023). Moreover, by definition of μ, we have

$$\frac{dP_{Z|X}}{d\mu_{Z|X}} = \frac{dP_{Z|X}}{dP_Z} \geq \varepsilon$$

because $P(Z = z|X) \geq \varepsilon$ for some $\varepsilon > 0$ and all $z \in \{0, 1\}$ by assumption. Finally, we note that since ℓ is bounded, condition (i) of the Theorem implies all the remaining requirements of Corollary 4.3(b) in Navjeevan et al. (2023) are satisfied. Therefore, Corollary 4.3(b) in Navjeevan et al. (2023) implies that $P(T^* \in \mathbf{S})$ is identified if and only if there exists a function f of (T, Z, X) satisfying $E[f(T, Z, X)|] < \infty$ and

$$\mu(\sum_{t\in\mathbf{T}} E_{\mu_{Z|X}}[f(t, Z, X)1\{T^*(Z) = t)] = \ell(T^*)) = 1 \tag{49}$$

where $E_{\mu_{Z|X}}$ denotes an expectation taken with respect to the conditional distribution of Z given X under μ. Next, note that $P(Z = 1) \in (0, 1)$, $\mu_{Z|X} = P_Z$, and $\mu_{T^*X} = P_{T^*X}$ by definition of μ imply that requirement (49) is equivalent to

$$P_{T^*X}(\sum_{t\in\mathbf{T}} \sum_{z\in\{0,1\}} f(t, z, X)1\{T^*(z) = t\} = \ell(T^*)) = 1 \tag{50}$$

holding for some function f of (T, Z, X). Moreover, since $P_{T^*}P_X \ll P_{T^*X}$ due to $P(T^* = t^*|X) \geq \varepsilon > 0$ for all $t^* \in \mathbf{M}$, condition (50) is itself equivalent to

$$P_{T^*}P_X(\sum_{t\in\mathbf{T}} \sum_{z\in\{0,1\}} f(t, z, X)1\{T^*(z) = t\} = \ell(T^*)) = 1. \tag{51}$$

Next, note that for any $(t, z) \in \mathbf{T} \times \{0, 1\}$, we obtain from T^* being independent of Z conditionally on X and $P(Z = z|X) \geq \varepsilon$ for some $\varepsilon > 0$ that

$$\begin{aligned} P(T = t, Z = z|X) &= P(T^*(z) = t|X)P(Z = z|X) \asymp P(T^*(z) = t|X) \\ &= \sum_{t^*\in\mathbf{M}:t^*(z)=t} P(T^* = t^*|X) \asymp \#\{t^* \in \mathbf{M} : t^*(z) = t\} \asymp P(T^*(z) = t), \end{aligned} \tag{52}$$

where the fourth result follows from $P(T^* = t^*|X) \geq \varepsilon$ for all $t^* \in \mathbf{M}$, and the final result from $P(T^*(z) = t) = 0$ if and only if $t^*(z) \neq t$ for all $t^* \in \mathbf{M}$ due to $P(T^* = t^*|X) \geq \varepsilon$. Moreover, by the law of iterated expectations and result (52) we also obtain

$$P(T = t, Z = z) = E[P(T = t, Z = z|X)] \asymp P(T^*(z) = t) \asymp P(T = t, Z = z|X). \tag{53}$$

In particular, note that $P(T = t, Z = z) \asymp P(T^*(z) = t)$ for all $(t, z) \in \mathbf{T} \times \{0, 1\}$ implies that condition (51), and hence identification of (48), is equivalent to

$$P_{T^*}P_X(\sum_{z\in\{0,1\}} \sum_{t:P(T=t,Z=z)>0} f(t, z, X)1\{T^*(z) = t\} = \ell(T^*)) = 1. \tag{54}$$

To conclude, note that if (5) holds, then (54), and hence identification of $P(T^* \in \mathbf{S})$, holds due to $T^* \in \mathbf{M}$ with probability one by Assumption 1.2 and $P(T = t, Z = z) \asymp P(T^*(z) = t)$ for all $(t, z) \in \mathbf{T} \times \{0, 1\}$. For the converse, suppose that (54) holds for some f satisfying $E[|f(T, Z, X)|] < \infty$. Then note that for any set $E \subseteq \mathbf{X}$, Bayes Rule and result (54) imply for any $(t, z) \in \mathbf{T} \times \{0, 1\}$ satisfying $P(T = t, Z = z) > 0$ that

$$P(X \in E) = \frac{P(X \in E|T = t, Z = z)P(T = t, Z = z)}{P(T = t, Z = z|X \in E)} \asymp P(X \in E|T = t, Z = z). \tag{55}$$

In particular, result (55) implies that $E[|f(t, z, X)|] \lesssim E[|f(T, Z, X)|]$ for any (t, z) satisying $P(T = t, Z = z) > 0$. Since $E[|f(T, Z, X)|] < \infty$ we may therefore define

$$\nu(t, z) = \begin{cases} E[|f(t, z, X)|] & \text{if } P(T = t, Z = z) > 0 \\ 0 & \text{if } P(T = t, Z = z) = 0 \end{cases} .$$

Finally, note that condition (54) implies that with P_{T^*} probability one we have that

$$\ell(T^*) = \sum_{z\in\{0,1\}} \sum_{t:P(T=t,Z=z)>0} \nu(t, z)1\{T^*(z) = t\} = \sum_{z\in\{0,1\}} \sum_{t\in\mathbf{T}} \nu(t, z)1\{T^*(z) = t\} \tag{56}$$

where the final equality follows from $P(T = t, Z = z) \asymp P(T^*(z) = t)$ by (53). Since $P(T^* = t^*|X) \geq \varepsilon$ for all $t^* \in \mathbf{M}$, result (56) implies that (5) holds with $\nu_0(t) = \nu(t, 0)$ and $\nu_1(t) = \nu(t, 1)$, and therefore that the identification of $P(T^* \in \mathbf{S})$ implies (5). ■

A.1.4 Proof of Theorem 2.2

To establish the result, we first use the law of iterated expectations to show that

$$\begin{aligned} &E[E[\nu_0(T)|Z = 0, X] + E[\nu_1(T)|Z = 1, X]] \\ &\quad = E[\nu_0(T)\frac{(1 - Z)}{P(Z = 0|X)} + \nu_1(T)\frac{Z}{P(Z = 1|X)}] \\ &\quad = \sum_{t\in\mathbf{T}} E[\nu_0(t)\frac{1\{T = t, Z = 0\}}{P(Z = 0|X)} + \nu_1(t)\frac{1\{T = t, Z = 1\}}{P(Z = 1|X)}]. \end{aligned} \tag{57}$$

Moreover, using the fact that response types T^* are independent of Z conditional on X yields

$$\begin{aligned} &\sum_{t\in\mathbf{T}} E[\nu_0(t)\frac{1\{T = t, Z = 0\}}{P(Z = 0|X)} + \nu_1(t)\frac{1\{T = t, Z = 1\}}{P(Z = 1|X)}] \\ &\quad = \sum_{t\in\mathbf{T}} E[\nu_0(t)P(T^*(0) = t|X)\frac{P(Z = 0|X)}{P(Z = 0|X)} + \nu_1(t)P(T^*(1) = t|X)\frac{P(Z = 1|X)}{P(Z = 1|X)}] \\ &\quad = \sum_{t\in\mathbf{T}} E[\nu_0(t)1\{T^*(0) = t\} + \nu_1(t)1\{T^*(1) = t\}]. \end{aligned} \tag{58}$$

Since $T^* \in \mathbf{M}$ with probability one, ISP implies that the last display in (58) equals $P(T^* \in \mathbf{S})$. The first claim of the Theorem follows from (57), (58), and $V \equiv \nu_1(T)Z - \nu_0(T)(1 - Z)$. For the second claim note that for any W with $E[|W|] < \infty$ we have

$$E[W\tilde{Z}] = E[\frac{WZ}{P(Z = 1|X)} - \frac{W(1 - Z)}{P(Z = 0|X)}] = E[E[Y|Z = 1, X] - E[Y|Z = 0, X]].$$

Therefore, the second claim of the Theorem follows from Lemma 2.1 and $P(T^* \in \mathbf{S}) = E(E[V|Z = 1, X] - E[V|Z = 0, X])$, as shown in the first part of the result. ■

A.1.5 Proof of Theorem 2.3

To establish part (i) note that for $\nu_0(T) = -1\{T \in \mathbf{T}_{\mathtt{i}}\}$ and $\nu_1(T) = 1\{T \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}\}$,

$$\sum_{t\in\mathbf{T}} \nu_0(t)1\{t^*(0) = t\} + \nu_1(t)1\{t^*(1) = t\}$$
$$= \sum_{t\in\mathbf{T}}(-1\{t \in \mathbf{T}_{\mathtt{i}}\}1\{t^*(0) = t\} + 1\{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}\}1\{t^*(1) = t\}) = 1\{t^* \in \mathbf{S}\},$$

where the second equality follows by noting that if $t^* \in \mathbf{M} \setminus \mathbf{S}$ then $t^*(0) = t^*(1) \notin \mathbf{T}_{\mathtt{io}}$ by definition of $\mathbf{T}_{\mathtt{io}}$ and IOS, while if $t^* \in \mathbf{S}$ then $1\{t^*(1) = t\} = 1$ for some $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}$ and $1\{t^*(0) = t\} = 0$ for all $t \in \mathbf{T}_{\mathtt{i}}$. Hence, IOS implies ISP as claimed. Moreover, we have

$$\kappa^\star(T, Z, X) = \tilde{Z}V - \tilde{Z}(1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{o}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\})P(Z = 0|X) \tag{59}$$
$$= \tilde{Z}V - \frac{(Z - P(Z = 1|X))}{P(Z = 1|X)}1\{T \in \mathbf{T}_{\mathtt{s}}\}$$

by definition of $\kappa^\star$, V, and direct calculation. Since we have $1\{T^* \in \mathtt{stay}(\mathrm{t})$ for some $t \in \mathbf{T}_{\mathtt{s}}\} = 1\{T \in \mathbf{T}_{\mathtt{s}}\}$ by definition of $\mathbf{T}_{\mathtt{s}}$ and T^* is independent of Z conditional on X by Assumption 1.1(iii), result (59) and Theorem 2.2 imply $E[\kappa^\star(T, Z, X)] = E[\tilde{Z}V] = P(T^* \in \mathbf{S})$, and therefore part (i) of the theorem follows.

In order to establish part (ii) of the Theorem, note that Theorem A.1 implies that

$$(\beta_0, \beta_1) \equiv \arg\min_{(b_0,b_1)} E[(Y - b_0\frac{h_0(X)}{E[h_0(X)|T^* \in \mathbf{S}]} - b_1\frac{Vh_1(X)}{E[h_1(X)|T^* \in \mathbf{S}]})^2\kappa^\star(T, Z, X)]$$
$$= E[(Y - b_0\frac{h_0(X)}{E[h_0(X)|T^* \in \mathbf{S}]} - b_1\frac{Vh_1(X)}{E[h_1(X)|T^* \in \mathbf{S}]})^2 1\{T^* \in \mathbf{S}\}].$$

Next, note that $1\{T^*(Z) \in \mathbf{T}_{\mathtt{i}}, T^* \in \mathbf{S}\} = 1\{T^*(1) \in \mathbf{T}_{\mathtt{i}}, T^* \in \mathbf{S}\}Z$ by definition of $\mathbf{T}_{\mathtt{i}}$. Therefore, using that $V = 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}, Z = 1\}$ we are able to conclude that

$$V1\{T^* \in \mathbf{S}\} = 1\{T^*(1) \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}\}Z1\{T^* \in \mathbf{S}\} = Z1\{T^* \in \mathbf{S}\}. \tag{60}$$

Therefore, result (60) and Z being independent of T^* conditional on X yields that

$$E[Vh_0(X)h_1(X)1\{T^* \in \mathbf{S}\}] = E[\frac{1\{T^* \in \mathbf{S}\}}{P^2(Z = 0|X)}]. \tag{61}$$

Combining result (61) with the law of iterated expectations then enables us to conclude

$$\frac{E[Vh_0(X)h_1(X)1\{T^* \in \mathbf{S}\}]/E[h_1(X)|T^* \in \mathbf{S}]E[h_0(X)|T^* \in \mathbf{S}])}{E[h_0^2(X)1\{T^* \in \mathbf{S}\}]/(E[h_0(X)|T^* \in \mathbf{S}])^2} = \frac{E[h_0(X)|T^* \in \mathbf{S}]}{E[h_1(X)|T^* \in \mathbf{S}]} \tag{62}$$

Thus, the Frisch-Waugh-Lovell theorem yields that β_1 is the minimizer to the problem

$$\beta_1 = \arg\min_{b_1} E[(Y - b_1(\frac{Vh_1(X)}{E[h_1(X)|T^* \in \mathbf{S}]} - \frac{h_0(X)}{E[h_1(X)|T^* \in \mathbf{S}]}))^2 1\{T^* \in \mathbf{S}\}] \tag{63}$$
$$= \arg\min_{b_1} E[(Y - \frac{b_1}{E[h_1(X)|T^* \in \mathbf{S}]}(Vh_1(X) - h_0(X)))^2 1\{T^* \in \mathbf{S}\}].$$

Next observe that result (60), Z being independent of (Y^*, T^*) conditional on X, the definitions of h_0 and h_1, and the law of iterated expectations imply

$$\begin{aligned} &E[Y(Vh_1(X) - h_0(X))1\{T^* \in \mathbf{S}\}] \\ &\quad = E[(YZ + Y(1-Z))(Zh_1(X) - h_0(Z))1\{T^* \in \mathbf{S}\}] \\ &\quad = E[(Y^*(T^*(1))Z(h_1(X) - h_0(X)) - Y^*(T^*(0))(1-Z)h_0(X))1\{T^* \in \mathbf{S}\}] \\ &\quad = E[(Y^*(T^*(1)) - Y^*(T^*(0)))1\{T^* \in \mathbf{S}\}]. \end{aligned} \tag{64}$$

Moreover, again using result (60) and the law of iterated expectations further implies

$$\begin{aligned} &E[(Vh_1(X) - h_0(X))^2 1\{T^* \in \mathbf{S}\}] \\ &\quad = E[(Zh_1^2(X) - Zh_1(X)h_0(Z) + h_0^2(X))1\{T^* \in \mathbf{S}\}] = E[h_1(X)1\{T^* \in \mathbf{S}\}]. \end{aligned} \tag{65}$$

Therefore, by combining results (63), (64), and (65) we are able to conclude that

$$\begin{aligned} \beta_1 &= \frac{E[Y(Vh_1(X) - h_0(X))1\{T^* \in \mathbf{S}\}]/E[h_1(X)|T^* \in \mathbf{S}]}{E[(Vh_1(X) - h_0(X))^2 1\{T^* \in \mathbf{S}\}]/(E[h_1(X)|T^* \in \mathbf{S}])^2} \\ &= \frac{E[(Y^*(T^*(1)) - Y^*(T^*(0)))1\{T^* \in \mathbf{S}\}]E[h_1(X)|T^* \in \mathbf{S}]}{E[h_1(X)1\{T^* \in \mathbf{S}\}]} \\ &= \tau(\mathbf{S}), \end{aligned} \tag{66}$$

which establishes the second claim of the Theorem. ■

A.1.6 Proof of Lemma 2.2

The claim for $t \in \mathbf{T_i} \cup \mathbf{T_o}$ follows from Lemma A.2 and the law of iterated expectations. The claim for $t \in \mathbf{T_{io}}$ follows by noting that if $t \in \mathbf{T_{io}}$, then $1\{Z = 1\}1\{T = t\} = 1\{Z = 1\}1\{T^* \in \texttt{in}(t)\}$ and $1\{Z = 0\}1\{T = t\} = 1\{T^* \in \texttt{out}(t)\}$. Therefore, we have

$$\begin{aligned} E[1\{T = t\}|Z = 1, X] &= E[1\{T^* \in \texttt{in}(t)\}|Z = 1, X] = P(T* \in \texttt{in}(t)\,|X) \\ E[1\{T = t\}|Z = 0, X] &= E[1\{T^* \in \texttt{out}(t)\}|Z = 0, X] = P(T* \in \texttt{out}(t)\,|X) \end{aligned}$$

because T^* is independent of Z conditional on X by Assumption 1.1(iii). Marginalization over the distribution of X proceeds as in the end of the proof of Theorem 2.2. ■

A.1.7 Proof of Corollary 2.1

For any $t^* \in \mathbf{S}$ and $t \in \{t^*(0), t^*(1)\}$ we have $P(T = t, T^* = t^*|X) > 0$ by Assumption 1.1(iv) and $P(T^* = t^*|X) > 0$ for all $t^* \in \mathbf{S}$. Therefore, Proposition 1.1 and RIO imply

$$E[Y^*(t)|T^* = t^*, X] = C(t, X) \tag{67}$$

for any $t^* \in \mathbf{S}$ and $t \in \{t^*(0), t^*(1)\}$. In particular, (67) implies $E[Y^*(t)|T^* = t^*, X]$ is constant in $t^* \in \texttt{in}(t) \cup \texttt{out}(t)$ and therefore that RIO implies (20) holds.

Conversely, if condition (20) is satisfied, then we may set $C(t, X)$ to equal

$$C(t, X) \equiv E[Y^*(t)|T^* \in \texttt{in}(t) \cup \texttt{out}(t), X], \tag{68}$$

where we have used that the right-hand side of (68) depends only on (t, X). Property (20) and

Proposition 1.1 then imply RIO holds. ■

A.1.8 Proof of Theorem 2.4

We first show that if the CCEF is identified, then IOS holds. We establish this direction by showing that, if the CCEF is identified, then at every $t \in \mathbf{T}$ one of the sets $\texttt{stay}(\mathrm{t})$, $\texttt{in}(\mathrm{t})$, and $\texttt{out}(\mathrm{t})$ is empty. To this end, note that if $t \in \mathbf{T}_{\mathtt{s}}$, then by definition $\texttt{in}(\mathrm{t})$ and $\texttt{out}(\mathrm{t})$ are empty. We therefore focus on $t \notin \mathbf{T}_{\mathtt{s}}$. At any such t either $\texttt{in}(\mathrm{t})$ or $\texttt{out}(\mathrm{t})$ is non-empty, and therefore there exists a $t^* \in \mathbf{S}$ such that $t \in \{t^*(0), t^*(1)\}$. Hence, because RIO holds and the CCEF is identified, the law of iterated expectations implies

$$\sum_{t^* \in \mathbf{S}} E[(Y - C(t,X))1\{T = t\}1\{T^* = t^*\}] = 0. \tag{69}$$

Next, note that $T = T^*(0)(1-Z) + T^*(1)Z$ and $(Y^*, T^*) \perp\!\!\!\perp Z|X$ by Assumption 1.1(iii) imply that

$$\begin{aligned}\sum_{t^* \in \mathbf{S}} & E[(Y - C(t,X))1\{T = t\}1\{T^* = t^*\}] \qquad (70)\\ &= \sum_{t^* \in \mathbf{S}} E[(Y^*(t) - C(t,X))1\{T^* = t^*\}1\{t^*(Z) = t\}]\\ &= E[(Y^*(t) - C(t,X))\{\sum_{t^* \in \mathbf{S}} 1\{T^* = t^*\}P(t^*(Z) = t|X)\}].\end{aligned}$$

Moreover, because the probability $P(t^*(Z) = t|X)$ is identified to equal either $P(Z = 0|X)$ (if $t^* \in \texttt{out}(\mathrm{t})$) or $P(Z = 1|X)$ (if $t^* \in \texttt{in}(\mathrm{t})$) it follows that the function

$$\ell_t(T^*, X) \equiv \sum_{t^* \in \mathbf{S}} 1\{T^* = t^*\}P(t^*(Z) = t|X) \tag{71}$$

is identified. Hence, so is the function $(Y^*(t) - C(t,X))\ell_t(T^*, X)$ because $C(T,X)$ is identified by hypothesis. Also note that results (69) and (70) then imply that

$$E[(Y^*(t) - C(t,X))\ell_t(T^*, X)] \tag{72}$$

is identified.

Corollary 4.4 in Navjeevan et al. (2023) provides necessary and sufficient conditions for identification of moments with the structure $E[\rho(Y^*(t))\ell(T^*, X)]$ to be identified when ρ and ℓ are identified. The arguments in Corollary 4.4 in Navjeevan et al. (2023), however, apply verbatim when ρ is allowed to depend on X as well, which maps into (72). To verify that the conditions of the corollary are satisfied, let $U_{\mathbf{M}}$ be a uniform distribution on $\mathbf{M}$ and set $\mu \equiv \{P_{Y|X}\}^d U_{\mathbf{M}} P_{ZX}$. By Lemma A.1, $P_{Y^*T^*ZX} \ll \mu$ and, therefore Assumption 1.1 implies that Assumptions 2.1 and 2.2 of Navjeevan et al. (2023) are satisfied with $\mathcal{Q} = \mathbf{Q} = L^1(\mu)$. Lemma A.1 further implies that $\mu \ll P_{Y^*T^*ZX}$, and we may therefore set $\bar{Q} = P_{Y^*T^*ZX}$ in Corollary 4.4 of Navjeevan et al. (2023). Condition (iii) of the theorem therefore verifies the requirements that $\bar{Q}^{\texttt{io}} \ll \bar{Q}$ and $d\bar{Q}^{\texttt{io}}/d\bar{Q}$ be bounded, while condition (i) of the theorem implies $\bar{Q}(T^* = t^*) > 0$ for all $t^* \in \mathbf{M}$. Since $(Y^*(t) - C(t,X))$ is bounded by condition (ii) of the theorem, ℓ_t is bounded by construction, and $\mathrm{Var}\{Y^*(t) - C(t,X)|X\} \geq \varepsilon$ by condition (ii) of the theorem, it follows that all the requirements of Corollary 4.4 in Navjeevan et al. (2023) are satisfied. Therefore, we may apply said corollary to conclude that identification of (72) implies that there is a function f of (Z, X) satisfying $E[|f(Z, X)|] < \infty$ and

$$P(E_{Z|X}[f(Z, X)1\{T^*(Z) = t\}] = \ell_t(T^*, X)) = 1, \tag{73}$$

where $E_{Z|X}$ denotes an expectation taken with respect to the distribution of Z given X.

Next, note that (73) holding with probability one under P_{T^*X}, the definition of ℓ_t in (71), and $P(T^* = t^*|X) > 0$ for all $t^* \in \mathbf{M}$ imply that (73) requires that

$$\begin{gathered} f(1, X)P(Z = 1|X) = P(Z = 1|X) \text{ if } \texttt{in}(t) \neq \emptyset \\ f(0, X)P(Z = 0|X) = P(Z = 0|X) \text{ if } \texttt{out}(t) \neq \emptyset \\ f(0, X)P(Z = 0|X) + f(1, X)P(Z = 1|X) = 0 \text{ if } \texttt{stay}(t) \neq \emptyset \end{gathered} \tag{74}$$

almost surely in X. However, since $P(Z = z|X) \geq \varepsilon$ for any $z \in \{0, 1\}$, there exists no solution to (74) whenever $\texttt{in}(t)$, $\texttt{out}(t)$, and $\texttt{stay}(t)$ are simultaneously non-empty. Hence, identification of the CCEF implies that one of $\texttt{in}(\mathrm{t})$, $\texttt{out}(\mathrm{t})$, and $\texttt{stay}(\mathrm{t})$ is empty and therefore that IOS holds.

To conclude, it remains to show that if IOS holds, then the CCEF is identified. To this end, note that since $P(T^* = t^*|X) > 0$ for any $t^* \in \mathbf{M}$, Assumption 1.1(iv) implies that $P(T = t, T^* = t|X) > 0$ for any $t \in \{t^*(0), t^*(1)\}$ and $t^* \in \mathbf{M}$. It is also helpful to note that for any $t^* \in \mathbf{S}$ and $t \in \{t^*(0), t^*(1)\}$, RIO and Corollary 2.1 imply that

$$E[Y|T = t, T^* = t^*, X] = E[Y|T = t, T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t}), X]. \tag{75}$$

We establish identification of the CCEF by considering the cases $t \in \mathbf{T}_{\texttt{s}}$, $t \in \mathbf{T}_{\texttt{io}}$, $t \in \mathbf{T}_{\texttt{i}}$, and $t \in \mathbf{T}_{\texttt{o}}$ separetely. For any $t \in \mathbf{T}_{\texttt{s}}$, the only $t^* \in \mathbf{M}$ that selects t is the stayer $t^* = (t^*(0), t^*(1)) = (t, t)$. Therefore, $1\{T = t\} = 1\{T^* = (t, t)\}$, which implies

$$\frac{E[Y1\{T = t\}|X]}{E[1\{T = t\}|X]} = \frac{E[Y1\{T = t, T^* = (t, t)\}|X]}{E[1\{T = t, T^* = (t, t)\}|X]} = E[Y|T = t, T^* = (t, t), X]. \tag{76}$$

Similarly, for any $t \in \mathbf{T}_{\texttt{io}}$, IOS implies $\texttt{stay}(\mathrm{t}) = \emptyset$. Therefore, $1\{T = t\} = 1\{T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})\}$ and result (75) allows us to conclude that

$$\begin{aligned} \frac{E[Y1\{T = t\}|X]}{E[1\{T = t\}|X]} &= \frac{E[Y1\{T = t, T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})\}|X]}{E[1\{T = t, T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})\}|X]} \\ &= E[Y|T = t, T^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t}), X] = E[Y|T = t, T^* = t^*, X] \end{aligned} \tag{77}$$

for any $t^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})$. Next, note that for any $t \in \mathbf{T}_{\texttt{i}}$, Lemma A.2, the law of iterated expectations, Corollary 2.1, and $\texttt{out}(\mathrm{t}) = \emptyset$ imply for any $t^* \in \texttt{in}(\mathrm{t})$ that

$$\begin{aligned} \frac{E[Y1\{T = t\}\tilde{Z}|X]}{E[1\{T = t\}\tilde{Z}|X]} &= \frac{E[Y^*(t)1\{T^* \in \texttt{in}(\mathrm{t})\}|X]}{P(T^* \in \texttt{in}(\mathrm{t})\,|X)} \\ &= E[Y^*(t)|T^* = t^*, X] = E[Y|T = t, T^* = t^*, X], \end{aligned} \tag{78}$$

where the final equality follows from Proposition 1.1. Moreover, since $\texttt{out}(\mathrm{t}) = \emptyset$ we have

$1\{T=t, Z=0\} = 1\{T^* \in \texttt{stay}(\mathrm{t}), Z=0\}$ and hence Assumption 1.1(iii) yields

$$\frac{E[Y1\{T=t,Z=0\}|X]}{E[1\{T=t,Z=0\}|X]} = \frac{E[Y^*(t)1\{T^* \in \texttt{stay}(\mathrm{t}), Z=0\}|X]}{E[1\{T^* \in \texttt{stay}(\mathrm{t}), Z=0\}|X]} \\ = E[Y^*(t)|T^* \in \texttt{stay}(\mathrm{t}), X] = E[Y|T=t, T^*=(t,t)], \tag{79}$$

where the final equality follows from Proposition 1.1 and $\texttt{stay}(\mathrm{t}) = \{(t,t)\}$. The identification of the CCEF at any $t \in \mathbf{T}_{\mathtt{o}}$ follows by applying the same arguments but replacing $\tilde{Z}$ with $-\tilde{Z}$ in (78) and $Z=0$ with $Z=1$ in (79). Hence, IOS implies identification of the CCEF and the theorem follows. ■

A.1.9 Proof of Theorem 2.5

Parts (i) and (ii) of the Theorem follow from Theorem A.1. To establish part (iii) of the Theorem, first note that by definition of $\tilde{Z}$ we have that

$$\tilde{Z}(1\{T \in \mathbf{T}_{\mathtt{io}}, Z=1\}P(Z=1|X) - 1\{T \in \mathbf{T}_{\mathtt{io}}, Z=0\}P(Z=0|X)) \\ = 1\{T \in \mathbf{T}_{\mathtt{io}}, Z=1\} + 1\{T \in \mathbf{T}_{\mathtt{io}}, Z=0\} = 1\{T \in \mathbf{T}_{\mathtt{io}}\}. \tag{80}$$

Also note that $\mathbf{T} = \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} + \mathbf{T}_{\mathtt{io}}$ because $\mathbf{T}_{\mathtt{s}}$ is empty by hypothesis. Therefore, since $V \equiv 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{\mathbf{T} \in \mathbf{T}_{\mathtt{io}}, Z=1\}$ it follows that $1 - V = 1\{T \in \mathbf{T}_{\mathtt{o}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}, Z=0\}$. Hence, the definition of $\kappa^\star$ and result (80) yield that

$$\kappa^\star(T,Z,X) = V\tilde{Z}P(Z=1|X) - (1-V)\tilde{Z}P(Z=0|X). \tag{81}$$

Part (iii) of the theorem then follows from (81) and the definition of $\tilde{Z}$. ■

A.1.10 Proof of Corollary 2.2

For the first claim, note that $P(T=t, T^*=t^*|X) > 0$ for any $t^* \in \mathbf{S}$ and $t \in \{t^*(0), t^*(1)\}$ by Assumption 1.1(iv) and $P(T^*=t^*|X) > 0$ for all $t^* \in \mathbf{S}$. We first show that condition (27) implies ARIO. To this end, note that Proposition 1.1 and $1\{T=t, T^*=t^*, T^* \in \mathbf{S}\}$ equaling zero whenever $t^* \notin \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})$ imply

$$E[Y|T,T^*,X]1\{T^* \in \mathbf{S}\} \\ = \sum_{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}} \sum_{t^* \in \texttt{in}(\mathrm{t}) \cup \texttt{out}(\mathrm{t})} E[Y^*(t)|T^*=t^*, X]1\{T=t, T^*=t^*\} \tag{82}$$

with probability one, and where we used $\texttt{in}(\mathrm{t}) = \emptyset$ and $\texttt{out}(\mathrm{t}) = \emptyset$ for all $t \in \mathbf{T}_{\mathtt{s}}$. Therefore, if condition (27) holds, then result (82) implies that with probability one

$$E[Y|T,T^*,X]1\{T^* \in \mathbf{S}\} = \sum_{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}} E[Y^*(t)|T^* \in \texttt{in}(\mathrm{t}), X]1\{T=t, T^* \in \texttt{in}(\mathrm{t})\} \\ + \sum_{t \in \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}} E[Y^*(t)|T^* \in \texttt{out}(\mathrm{t}), X]1\{T=t, T^* \in \texttt{out}(\mathrm{t})\}.$$

Hence, using that $1\{T = t, T^* \in \texttt{in}(\mathrm{t})\} = 1\{T = t, Z = 1, T^* \in \mathbf{S}\}$ and similarly $1\{T = t, T^* \in \texttt{out}(\mathrm{t})\} = 1\{T = t, Z = 0, T^* \in \mathbf{S}\}$ allows us to conclude that

$$\begin{aligned} E[Y|T, T^*, X]1\{T^* \in \mathbf{S}\} = &\sum_{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}} E[Y^*(t)|T^* \in \texttt{in}(\mathrm{t}), X]1\{T = t, Z = 1\}1\{T^* \in \mathbf{S}\} \\ &+ \sum_{t \in \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}} E[Y^*(t)|T^* \in \texttt{out}(\mathrm{t}), X]1\{T = t, Z = 0\}1\{T^* \in \mathbf{S}\}, \end{aligned}$$

which implies ARIO holds. For the converse, suppose (27) holds and note that for any $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}$ and $t^* \in \texttt{in}(\mathrm{t})$ we have that $1\{T = t, T^* = t^*\}$ implies $Z = 1$. Hence,

$$\begin{aligned} &E[Y|T = t, T^* = t^*, X]1\{T = t, T^* = t^*\} \\ &\qquad = E[Y^*(t)|T^* = t^*, X]1\{T = t, T^* = t^*, Z = 1\} = C_a(t, 1, X). \end{aligned} \tag{83}$$

Since (83) holds for any $t^* \in \texttt{in}(\mathrm{t})$, we can conclude that the first display in (27) holds. Similarly, since $1\{T = t, T^* = t^*\}$ implies $Z = 0$ for any $t^* \in \texttt{out}(\mathrm{t})$ we have

$$\begin{aligned} &E[Y|T = t, T^* = t^*, X]1\{T = t, T^* = t^*\} \\ &\qquad = E[Y^*(t)|T^* = t^*, X]1\{T = t, T^* = t^*, Z = 0\} = C_a(t, 0, X). \end{aligned} \tag{84}$$

for any $t^* \in \texttt{out}(\mathrm{t})$. Hence, it follows that the second display in (27) holds and therefore that ARIO implies (27).

For the second claim of the corollary, note that (83) holding for any $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}$ and $t^* \in \texttt{in}(\mathrm{t})$ and (84) holding for any $t \in \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}$ and $t^* \in \texttt{out}(\mathrm{t})$ imply that

$$E[Y^*(T^*(1)) - Y^*(T^*(0))|T^* = t^*, X] = C_a(t^*(1), 1, X) - C_a(t^*(0), 0, X) \tag{85}$$

for any $t^* = (t^*(0), t^*(1)) \in \mathbf{S}$. Finally, note that $T \in \mathbf{T}_{\mathtt{i}}$ and $T^* \in \mathbf{S}$ imply $Z = 1$ because $\texttt{out}(\mathrm{t}) = \emptyset$ for all $t \in \mathbf{T}_{\mathtt{i}}$. Similarly, $T \in \mathbf{T}_{\mathtt{o}}$ and $T^* \in \mathbf{S}$ imply $Z = 0$ because $\texttt{in}(\mathrm{t}) = \emptyset$ for all $t \in \mathbf{T}_{\mathtt{o}}$. Therefore, we can conclude that

$$\begin{aligned} &C_a(T, Z, X)1\{T \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}\}1\{T^* \in \mathbf{S}\} \\ &\qquad = (C_a(T, 1, X)1\{T \in \mathbf{T}_{\mathtt{i}}\} + C_a(T, 0, X)1\{T \in \mathbf{T}_{\mathtt{o}}\})1\{T^* \in \mathbf{S}\}, \end{aligned}$$

which establishes the final claim of the corollary. ■

A.1.11 Proof of Theorem 2.6

The result follows by setting $g(Y, T, Z, X) = (Y - F_a(T, Z, X; b))^2$ in Theorem A.1 and noting that $h = g$ in the notation of that theorem due to (30). ■

A.1.12 Proof of Theorem 2.7

We first derive some results that are helpful for establishing both parts (i) and (ii). Note that for any $t \in \mathbf{T}_{\mathtt{i}}$ we have from Lemma A.2 and the law of iterated expectations that

$$E[f(Y^*(t))1\{T^* \in \texttt{in}(\mathrm{t})\}|X] = E[f(Y)1\{T = t\}\tilde{Z}|X]. \tag{86}$$

Similarly, for any $t \in \mathbf{T}_{\mathtt{o}}$, Lemma A.2 and the law of iterated expectations imply that

$$E[f(Y^*(t))1\{T^* \in \mathtt{out}(\mathrm{t})\}|X] = -E[f(Y)1\{T = t\}\tilde{Z}|X].$$

Finally, Assumption 1.1(iii), and $1\{T = t, Z = 1\} = 1\{Z = 1, T^* \in \mathtt{in}(\mathrm{t})\}$ and $1\{T = t, Z = 0\} = 1\{Z = 1, T^* \in \mathtt{out}(\mathrm{t})\}$ for any $t \in \mathbf{T}_{\mathtt{io}}$ imply for any function f that

$$\begin{aligned} E[f(Y^*(t))1\{T^* \in \mathtt{in}(\mathrm{t})\}|X] &= E[f(Y)\frac{1\{T = t, Z = 1\}}{P(Z = 1|X)}|X] \\ E[f(Y^*(t))1\{T^* \in \mathtt{out}(\mathrm{t})\}|X] &= E[f(Y)\frac{1\{T = t, Z = 0\}}{P(Z = 0|X)}|X] \end{aligned} \tag{87}$$

for any $t \in \mathbf{T}_{\mathtt{io}}$. Also note $Z(Z - P(Z = 1|X)) = Z(1 - P(Z = 1|X))$ implies that

$$E[f(Y)\frac{1\{T = t, Z = 1\}}{P(Z = 1|X)}|X] = E[f(Y)1\{T = t\}Z\tilde{Z}|X]. \tag{88}$$

Similarly, since $(1 - Z)(Z - P(Z = 1|X)) = (1 - Z)(-P(Z = 1|X))$ we have that

$$E[f(Y)\frac{1\{T = t, Z = 0\}}{P(Z = 0|X)}|X] = -E[f(Y)1\{T = t\}(1 - Z)\tilde{Z}|X]. \tag{89}$$

We can now establish part (i) of the Theorem. To this end note that $P(T^* \in \mathtt{in}(\mathrm{t})\,|X) > 0$ for any $t \in \mathbf{T}_{\mathtt{i}}$ and $P(T^* \in \mathtt{out}(\mathrm{t})\,|X) > 0$ for any $t \in \mathbf{T}_{\mathtt{o}}$ because $P(T^* = t^*|X) > 0$ for any $t^* \in \mathbf{S}$ by assumption. Therefore, evaluating results (86) and (87) at $f(Y) = Y$ and $f(Y) = 1$ we are able to conclude that

$$\frac{E[\tilde{Z} \times 1\{T = t\}Y|X]}{E[\tilde{Z} \times 1\{T = t\}|X]} = \begin{cases} E[Y^*(t)|T^* \in \mathtt{in}(\mathrm{t}), X] & \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ E[Y^*(t)|T^* \in \mathtt{out}(\mathrm{t}), X] & \text{if } t \in \mathbf{T}_{\mathtt{o}} \end{cases}$$

Similarly, evaluating results (87), (88), and (89) at $f(Y) = Y$ and $f(Y) = 1$ implies that

$$\begin{aligned} E[Y^*(t)|T^* \in \mathtt{in}(\mathrm{t}), X] &= \frac{E[\tilde{Z} \times 1\{T = t\}ZY|X]}{E[\tilde{Z} \times 1\{T = t\}Z|X]} \\ E[Y^*(t)|T^* \in \mathtt{out}(\mathrm{t}), X] &= \frac{E[\tilde{Z} \times 1\{T = t\}(1 - Z)Y|X]}{E[\tilde{Z} \times 1\{T = t\}(1 - Z)|X]} \end{aligned} \tag{90}$$

for any $t \in \mathbf{T}_{\mathtt{io}}$. The first part of the theorem then follows from ARIO, (88) and (90).

For the second part of the theorem, we take expectations (over X) of results (86) and (87) evaluated at both $f(Y) = Y$ and $f(Y) = 1$ to obtain that

$$\frac{E[\tilde{Z} \times 1\{T = t\}Y]}{E[\tilde{Z} \times 1\{T = t\}]} = \begin{cases} E[Y^*(t)|T^* \in \mathtt{in}(\mathrm{t})] & \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ E[Y^*(t)|T^* \in \mathtt{out}(\mathrm{t})] & \text{if } t \in \mathbf{T}_{\mathtt{o}} \end{cases}. \tag{91}$$

Similarly, evaluating results (87), (88), and (89) at $f(Y) = Y$ and $f(Y) = 1$ implies that

$$\begin{aligned} E[Y^*(t)|T^* \in \mathtt{in}(\mathrm{t})] &= \frac{E[\tilde{Z} \times 1\{T = t\}ZY]}{E[\tilde{Z} \times 1\{T = t\}Z]} \\ E[Y^*(t)|T^* \in \mathtt{out}(\mathrm{t})] &= \frac{E[\tilde{Z} \times 1\{T = t\}(1 - Z)Y]}{E[\tilde{Z} \times 1\{T = t\}(1 - Z)]}. \end{aligned} \tag{92}$$

The second part of the theorem therefore follows from (92) and condition (33). ■

A.1.13 Proof of Theorem 2.8

Given a CCEF specification F_a satisfying (30), Corollary 2.2(ii) and Theorem A.1 taken together imply

$$\beta = \arg\min_b E[(C_a(T,Z,X) - F_a(T,Z,X;b))^2 1\{T^* \in \mathbf{S}\}] \tag{93}$$
$$= \arg\min_b E[(C_a(T,Z,X) - F_a(T,Z,X;b))^2 \kappa^\star(T,Z,X)] \ .$$

Using the fact that $\kappa^\star(t,Z,X) = 0$ for any $t \in \mathbf{T}_{\mathtt{s}}$ and the law of iterated expectations yields

$$E[(C_a(T,Z,X) - F_a(T,Z,X;b))^2 \kappa^\star(T,Z,X)] \tag{94}$$
$$= \sum_{t\in\mathbf{T}\setminus\mathbf{T}_{\mathtt{s}}} \sum_{z\in\{0,1\}} E[(C_a(t,z,X) - F_a(t,z,X;b))^2 1\{T=t, Z=z\}\kappa^\star(t,z,X)]$$
$$= \sum_{t\in\mathbf{T}\setminus\mathbf{T}_{\mathtt{s}}} \sum_{z\in\{0,1\}} E[(C_a(t,z,X) - F_a(t,z,X;b))^2 P(T=t, Z=z|X)\kappa^\star(t,z,X)] \ .$$

Note also that for any $t \in \mathbf{T}_{\mathtt{i}}$ we have that $1\{T^*(1) = t\} = 1\{T^* \in \mathtt{stay}(\mathrm{t}) \cup \mathtt{in}(\mathrm{t})\}$ and $1\{T^*(0) = t\} = 1\{T^* \in \mathtt{stay}(\mathrm{t})\}$. Therefore, Assumption 1.1(iii) implies

$$P(T=t|Z=1,X) - P(T=t|Z=0,X) \tag{95}$$
$$= P(T^*(1)=t|X) - P(T^*(0)=t|X) = P(T^* \in \mathtt{in}(\mathrm{t})\,|X) > 0 \ .$$

Likewise, for any treatment level $t \in \mathbf{T}_{\mathtt{o}}$,

$$P(T=t|Z=0,X) - P(T=t|Z=1,X) = P(T^* \in \mathtt{out}(\mathrm{t})\,|X) > 0 \ . \tag{96}$$

The facts that $C_a(T,Z,X) = C_1(T,X)1\{T \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}\} + C_2(T,Z,X)1\{T \in \mathbf{T}_{\mathtt{io}}\}$ by Corollary 2.2(ii) and that F_a satisfies (30) imply that, for any $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}$,

$$\sum_{z\in\{0,1\}} E[(C_a(t,z,X) - F_a(t,z,X;b))^2 P(T=t,Z=z|X)\kappa^\star(t,z,X)] \tag{97}$$
$$= E[(C_1(t,X) - F_1(t,X;b))^2 \sum_{z\in\{0,1\}} P(T=t,Z=z|X)\kappa^\star(t,z,X)]$$
$$= E[(C_1(t,X) - F_1(t,X;b))^2 \omega_t(X)] \ ,$$

where the final equality follows from (95) and (96), the definition of $\kappa^\star$, and direct calculation. Similarly, since $\kappa^\star(t,z,X) = 1$ when $t \in \mathbf{T}_{\mathtt{io}}$, we have that for any $t \in \mathbf{T}_{\mathtt{io}}$

$$\sum_{z\in\{0,1\}} E[(C_a(t,z,X) - F_a(t,z,X;b))^2 P(T=t,Z=z|X)] \tag{98}$$
$$= \sum_{z\in\{0,1\}} E[(C_2(t,z,X) - F_2(t,z,X;b))^2 \omega_{t,z}(X)] \ .$$

Hence, combining (93), (94), (97), and (98), and using $\mathbf{T} \setminus \mathbf{T}_{\mathtt{s}} = \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \cup \mathbf{T}_{\mathtt{io}}$, yields

$$\beta = \arg\min_b \Big\{ E[\sum_{t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}} (C_1(t,X) - F_1(t,X;b))^2 \omega_t(X)] \\ + E[\sum_{t \in \mathbf{T}_{\mathtt{io}}} \sum_{z \in \{0,1\}} (C_2(t,z,X) - F_2(t,z,X;b))^2 \omega_{t,z}(X)] \Big\} . \tag{99}$$

Finally, note that Corollary 2.2, the definition of $\tilde{Y}_t$, and Theorem 2.7(i) imply that

$$C_1(t,X) = E[Y^*(t)|T^* = t^*, X] = E[\tilde{Y}_t|X] \text{ for any } t \in \mathbf{T}_{\mathtt{i}},\ t^* \in \mathtt{in}(\mathrm{t}) \\ C_1(t,X) = E[Y^*(t)|T^* = t^*, X] = E[\tilde{Y}_t|X] \text{ for any } t \in \mathbf{T}_{\mathtt{o}},\ t^* \in \mathtt{out}(\mathrm{t}) \ . \tag{100}$$

Similarly, Corollary 2.2, the definition of $\tilde{Y}_{t,z}$, and Theorem 2.7(i) together imply

$$C_2(t,0,X) = E[Y^*(t)|T^* = t^*, X] = E[\tilde{Y}_{t,0}|X] \text{ for any } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \mathtt{out}(\mathrm{t}) \\ C_2(t,1,X) = E[Y^*(t)|T^* = t^*, X] = E[\tilde{Y}_{t,1}|X] \text{ for any } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \mathtt{in}(\mathrm{t}) . \tag{101}$$

The corollary then follows from (99), (100), and (101). ■

A.1.14 Supporting Results

Lemma A.1. *Suppose Assumptions 1.1, 1.2 hold, $P_{Y^*|T^*X} \ll \{P_{Y|X}\}^d \ll P_{Y^*|T^*X}$, and that for some $\varepsilon > 0$, with probability one, we have $P(T^* = t^*|X) \geq \varepsilon$ for all $t^* \in \mathbf{M}$ and $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$. Letting $U_{\mathbf{M}}$ be the uniform distribution on $\mathbf{M}$ and setting $\mu \equiv \{P_{Y|X}\}^d U_{\mathbf{M}} P_Z P_X$, it then follows that $P_{Y^*T^*ZX} \ll \mu \ll P_{Y^*T^*ZX}$.*

Proof. First note that because (Y^*, T^*) is independent of Z conditionally on X by Assumption 1.1(iii) and $T^* \in \mathbf{M}$ with probability one by Assumption 1.2 we obtain that

$$P_{Y^*T^*ZX} = P_{Y^*|T^*ZX} P_{T^*|ZX} P_{ZX} \ll P_{Y^*|T^*X} U_{\mathbf{M}} P_{ZX} \ll \mu, \tag{102}$$

where the final result follows from the definition of μ, $P_{ZX} \ll P_Z P_X$, and $P_{Y^*|T^*X} \ll \{P_{Y|X}\}^d$ by hypothesis. Next, note that $P(Z = z|X) \geq \varepsilon$ for some $\varepsilon > 0$ and all $z \in \{0,1\}$ implies that $P_Z P_X \ll P_{ZX}$. Therefore, since $\{P_{Y|X}\}^d \ll P_{Y^*|T^*X}$ by assumption and $P(T^* = t^*|X) \geq \varepsilon$ for all $t^* \in \mathbf{M}$ we also obtain that

$$\mu \equiv \{P_{Y|X}\}^d U_{\mathbf{M}} P_Z P_X \ll P_{Y^*|T^*X} P_{T^*|X} P_{ZX} = P_{Y^*T^*ZX}, \tag{103}$$

where the final result follows from (Y^*, T^*) being independent of Z conditionally on X by Assumption 1.1(iii). The lemma therefore follows from (102) and (103). ■

Theorem A.1. *Suppose Assumptions 1.1, 1.2, and* IOS *hold, and that, for some $\varepsilon > 0$, $\varepsilon \leq P(Z = 1|X) \leq 1 - \varepsilon$ with probability one. For any function g of (Y, T, Z, X) define*

$$h(Y,t,z,X) = \begin{cases} g(Y,t,1,X) & \text{for any } t \in \mathbf{T}_{\mathtt{i}} \text{ and } z \in \{0,1\} \\ g(Y,t,0,X) & \text{for any } t \in \mathbf{T}_{\mathtt{o}} \text{ and } z \in \{0,1\} \\ g(Y,t,z,X) & \text{otherwise} \end{cases} . \tag{104}$$

If h satisfies $E[|h(Y,T,Z,X)|] < \infty$, then it follows that

$$E[h(Y,T,Z,X)\kappa^\star(T,Z,X)] = E[g(Y,T,Z,X)1\{T^* \in \mathbf{S}\}]. \tag{105}$$

Proof. First, note the definition of $\kappa^\star$ in (18) and $P(Z = z|X) \geq \varepsilon > 0$ for $z \in \{0,1\}$ by hypothesis, imply that $\kappa^\star \in L^\infty(P)$. Hence, it follows that $\kappa^\star h \in L^1(P)$ because $h \in L^1(P)$ by hypothesis. Next, observe that Lemma A.2(i), the law of iterated expectations, and $h(Y^*(t),t,1,X) = h(Y^*(t),t,0,X)$ when $t \in \mathbf{T}_{\mathtt{i}}$ yield

$$\begin{aligned}\sum_{t\in\mathbf{T}_{\mathtt{i}}} & E[h(Y,T,Z,X)1\{T=t\}\tilde{Z}P(Z=1|X)] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{i}}} E[h(Y^*(t),t,Z,X)1\{T^*\in\mathtt{in}(t)\}P(Z=1|X)] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{i}}} E[h(Y,T,Z,X)1\{T^*\in\mathtt{in}(t)\}1\{T=t\}],\end{aligned} \tag{106}$$

where the second equality follows from Lemma A.2(ii). Moreover, by similar arguments

$$\begin{aligned}\sum_{t\in\mathbf{T}_{\mathtt{o}}} & E[h(Y,T,Z,X)1\{T=t\}(-\tilde{Z})P(Z=0|X)] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{o}}} E[h(Y^*(t),t,Z,X)1\{T^*\in\mathtt{out}(t)\}P(Z=0|X)] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{o}}} E[h(Y,T,Z,X)1\{T^*\in\mathtt{out}(t)\}1\{T=t\}],\end{aligned} \tag{107}$$

where the first and second equalities again rely on Lemmas A.2(i) and A.2(ii). On the other hand, by IOS, any $t \in \mathbf{T}_{\mathtt{io}}$ satisfies $\mathtt{stay}(t) = \emptyset$, and hence

$$\begin{aligned}1\{T=t\}1\{Z=0\} &= 1\{T=t\}1\{T^*\in\mathtt{out}(t)\} \\ 1\{T=t\}1\{Z=1\} &= 1\{T=t\}1\{T^*\in\mathtt{in}(t)\}\end{aligned} \tag{108}$$

for any $t \in \mathbf{T}_{\mathtt{io}}$. Therefore, we can conclude from result (108) and $Z \in \{0,1\}$ that

$$\begin{aligned}\sum_{t\in\mathbf{T}_{\mathtt{io}}} & E[h(Y,T,Z,X)1\{T=t\}] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{io}}} E[h(Y,T,Z,X)1\{T=t\}(1\{Z=0\}+1\{Z=1\})] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{io}}} E[h(Y,T,Z,X)1\{T=t\}1\{T^*\in\mathtt{out}(t)\cup\mathtt{in}(t)\}].\end{aligned} \tag{109}$$

Next, note that for any $t \in \mathbf{T}_{\mathtt{i}}$ and $t^* \in \mathbf{S}$, $1\{T^* = t^*\}1\{T = t\} = 1$ implies that $T^* \in \mathtt{in}(t)$ due to $\mathtt{out}(t) = \emptyset$ when $t \in \mathbf{T}_{\mathtt{i}}$. Therefore, using that $\mathtt{in}(t) \subseteq \mathbf{S}$, we obtain

$$\begin{aligned}\sum_{t\in\mathbf{T}_{\mathtt{i}}} & E[h(Y,T,Z,X)1\{T^*\in\mathtt{in}(t)\}1\{T=t\}] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{i}}}\sum_{t^*\in\mathbf{S}} E[h(Y,T,Z,X)1\{T^*\in\mathtt{in}(t)\}1\{T^*=t^*\}1\{T=t\}] \\ &= \sum_{t\in\mathbf{T}_{\mathtt{i}}}\sum_{t^*\in\mathbf{S}} E[h(Y,T,Z,X)1\{T^*=t^*\}1\{T=t\}].\end{aligned} \tag{110}$$

Similarly, since for any $t \in \mathbf{T}_{\mathtt{o}}$ and $t^* \in \mathbf{S}$ it follows that $1\{T^* = t^*\}1\{T = t\} = 1$ implies $T^* \in \mathtt{out}(t)$, we can conclude from the set inclusion $\mathtt{out}(t) \subseteq \mathbf{S}$ that

$$\sum_{t \in \mathbf{T}_{\mathtt{o}}} E[h(Y,T,Z,X)1\{T^* \in \mathtt{out}(t)\}1\{T = t\}] \\ = \sum_{t \in \mathbf{T}_{\mathtt{o}}} \sum_{t^* \in \mathbf{S}} E[h(Y,T,Z,X)1\{T^* = t^*\}1\{T = t\}]. \tag{111}$$

Furthermore, because $1\{T = t^*\}1\{T = t\} = 1$ implies $T^* \in \mathtt{in}(t) \cup \mathtt{out}(t)$ for any $t \in \mathbf{T}_{\mathtt{io}}$ and $t^* \in \mathbf{S}$, it also follows from the set inclusion $\mathtt{in}(t) \cup \mathtt{out}(t) \subseteq \mathbf{S}$ that

$$\sum_{t \in \mathbf{T}_{\mathtt{io}}} E[h(Y,T,Z,X)1\{T = t\}1\{T^* \in \mathtt{out}(t) \cup \mathtt{in}(t)\}] \\ = \sum_{t \in \mathbf{T}_{\mathtt{io}}} \sum_{t^* \in \mathbf{S}} E[h(Y,T,Z,X)1\{T^* = t^*\}1\{T = t\}]. \tag{112}$$

To conclude, note $1\{T = t\}1\{T^* = t^*\} = 0$ for any $t^* \in \mathbf{S}$ and $t \in \mathbf{T}_{\mathtt{s}}$. Since $\mathbf{T}_{\mathtt{s}}$, $\mathbf{T}_{\mathtt{i}}$, $\mathbf{T}_{\mathtt{o}}$, and $\mathbf{T}_{\mathtt{io}}$ form a partition of $\mathbf{T}$, we can combine the definition of $\kappa^\star$ in (18) with results (106), (107), (109), (111), and (112) together with Lemma A.3 to obtain that

$$E[h(Y,T,Z,X)\kappa^\star(T,Z,X)] = \sum_{t \in \mathbf{T}} \sum_{t^* \in \mathbf{S}} E[h(Y,T,Z,X)1\{T^* = t^*\}1\{T = t\}] \\ = \sum_{t^* \in \mathbf{S}} E[g(Y,T,Z,X)1\{T^* = t^*\}], \tag{113}$$

which verifies the claim of the theorem. ■

Lemma A.2. *If Assumptions 1.1(i)-(iii) hold, then: (i) The following equalities hold:*

$$E[1\{T = t\}\tilde{Z}|Y^*,T^*,X] = 1\{T^* \in \mathtt{in}(t)\} \text{ for all } t \in \mathbf{T}_{\mathtt{i}}$$
$$-E[1\{T = t\}\tilde{Z}|Y^*,T^*,X] = 1\{T^* \in \mathtt{out}(t)\} \text{ for all } t \in \mathbf{T}_{\mathtt{o}};$$

and (ii) For any function h of (Y,T,Z,X) satisfying $h(Y,t,0,X) = h(Y,t,1,X)$ for any $t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}}$ and $E[|h(Y,T,Z,X)|] < \infty$ it follows that

$$\sum_{t \in \mathbf{T}_{\mathtt{i}}} E[1\{T^* \in \mathtt{in}(t)\}(1\{T = t\}h(Y,T,Z,X) - P(Z = 1|X)h(Y^*(t),t,Z,X)] = 0$$
$$\sum_{t \in \mathbf{T}_{\mathtt{o}}} E[1\{T^* \in \mathtt{out}(t)\}(1\{T = t\}h(Y,T,Z,X) - P(Z = 0|X)h(Y^*(t),t,Z,X)] = 0.$$

Proof. To establish (i), first note $\tilde{Z} \in L^1(P)$ by the law of iterated expectations. Letting $E_{Z|X}$ denote expectations with respect to the conditional distribution of Z given X, it then follows from Corollary A.1 in Navjeevan et al. (2023) that for any $t \in \mathbf{T}_{\mathtt{i}}$

$$E[1\{T = t\}\tilde{Z}|Y^*,T^*,X] = E_{Z|X}[\frac{Z - P(Z = 1|X)}{P(Z = 1|X)P(Z = 0|X)}1\{T^*(Z) = t\}] \\ = 1\{T^*(1) = t\} - 1\{T^*(0) = t\} = 1\{T^* \in \mathtt{in}(t)\}, \tag{114}$$

where the first equality holds by definition of $\tilde{Z}$, and the final equality follows from $\mathtt{out}(t) = \emptyset$ for any $t \in \mathbf{T}_{\mathtt{i}}$. Moreover, again relying on Corollary A.1 in Navjeevan et al. (2023) and the definition of $\tilde{Z}$ yields, for any $t \in \mathbf{T}_{\mathtt{o}}$, that

$$-E[1\{T=t\}\tilde{Z}|Y^*,T^*,X] = E_{Z|X}[\frac{P(Z=1|X)-Z}{P(Z=1|X)P(Z=0|X)}1\{T^*(Z)=t\}] \tag{115}$$
$$= 1\{T^*(0)=t\} - 1\{T^*(1)=t\} = 1\{T^* \in \mathtt{out}(t)\},$$

where in the final equality we used that $\mathtt{in}(t) = \emptyset$ for any $t \in \mathbf{T}_{\mathtt{o}}$. Therefore, part (i) of the lemma follows from results (114) and (115).

In order to establish part (ii) of the lemma, first note that for any $t \in \mathbf{T}$ we have

$$1\{T=t\}1\{T^* \in \mathtt{in}(t)\} = 1\{Z=1\}1\{T^* \in \mathtt{in}(t)\} \tag{116}$$
$$1\{T=t\}1\{T^* \in \mathtt{out}(t)\} = 1\{Z=0\}1\{T^* \in \mathtt{out}(t)\}$$

by definition of the sets $\mathtt{in}(t)$ and $\mathtt{out}(t)$. Hence, by employing (116) we can conclude

$$\sum_{t\in\mathbf{T}_{\mathtt{i}}} E[h(Y,T,Z,X)1\{T=t\}1\{T^* \in \mathtt{in}(t)\}] \tag{117}$$
$$= \sum_{t\in\mathbf{T}_{\mathtt{i}}} E[h(Y^*(t),t,1,X)1\{T^* \in \mathtt{in}(t)\}1\{Z=1\}]$$
$$= \sum_{t\in\mathbf{T}_{\mathtt{i}}} E[h(Y^*(t),t,Z,X)1\{T^* \in \mathtt{in}(t)\}P(Z=1|X)],$$

where the second equality follows from the law of iterated expectations, (Y^*,T^*) being independent of Z conditionally on X by Assumption 1.1(iii), and $h(Y^*(t),t,1,X) = h(Y^*(t),t,0,X)$ for any $t \in \mathbf{T}_{\mathtt{i}}$ by hypothesis. Similarly, we also have

$$\sum_{t\in\mathbf{T}_{\mathtt{o}}} E[h(Y,T,Z,X)1\{T=t\}1\{T^* \in \mathtt{out}(t)\}] \tag{118}$$
$$= \sum_{t\in\mathbf{T}_{\mathtt{o}}} E[h(Y^*(t),t,0,X)1\{T^* \in \mathtt{out}(t)\}1\{Z=0\}]$$
$$= \sum_{t\in\mathbf{T}_{\mathtt{o}}} E[h(Y^*(t),t,Z,X)1\{T^* \in \mathtt{out}(t)\}P(Z=0|X)]$$

and therefore part (ii) of the lemma follows from results (117) and (118). ■

Lemma A.3. *Let Assumptions 1.1, 1.2 hold. For any function g of (Y,T,Z,X) set*

$$h(Y,t,z,X) = \begin{cases} g(Y,t,1,X) & \text{for any } t \in \mathbf{T}_{\mathtt{i}} \text{ and } z \in \{0,1\} \\ g(Y,t,0,X) & \text{for any } t \in \mathbf{T}_{\mathtt{o}} \text{ and } z \in \{0,1\} \\ g(Y,t,z,X) & \text{otherwise} \end{cases}.$$

Then, it follows that g and h satisfy the following equality with probability one

$$\sum_{t^*\in\mathbf{S}} h(Y,T,Z,X)1\{T^*=t^*\} = \sum_{t^*\in\mathbf{S}} g(Y,T,Z,X)1\{T^*=t^*\}.$$

Proof. First note that $\mathbf{T}_{\mathtt{s}}, \mathbf{T}_{\mathtt{io}}, \mathbf{T}_{\mathtt{i}}$, and $\mathbf{T}_{\mathtt{o}}$ form a partition of the set of possible treatments $\mathbf{T}$.

Therefore, by definition of h we can conclude that

$$\sum_{t^*\in\mathbf{S}} g(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{io}}\cup\mathbf{T}_{\mathtt{s}},T^*=t^*\}$$
$$=\sum_{t^*\in\mathbf{S}} h(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{io}}\cup\mathbf{T}_{\mathtt{s}},T^*=t^*\}. \tag{119}$$

Moreover, note that if $t\in\mathbf{T}_{\mathtt{i}}$, then $\mathtt{out}(\mathrm{t})=\emptyset$, which implies $1\{T\in\mathbf{T}_{\mathtt{i}},T^*=t^*\}=1\{T\in\mathbf{T}_{\mathtt{i}},T^*=t^*,Z=1\}$ for any $t^*\in\mathbf{S}$. Hence, we can conclude that

$$\sum_{t^*\in\mathbf{S}} g(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{i}},T^*=t^*\} \tag{120}$$
$$=\sum_{t^*\in\mathbf{S}} g(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{i}},T^*=t^*,Z=1\}$$
$$=\sum_{t^*\in\mathbf{S}} h(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{i}},T^*=t^*\},$$

where the final equality holds by definition of h. By similar arguments we also obtain

$$\sum_{t^*\in\mathbf{S}} g(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{o}},T^*=t^*\}=\sum_{t^*\in\mathbf{S}} h(Y,T,Z,X)1\{T\in\mathbf{T}_{\mathtt{o}},T^*=t^*\}, \tag{121}$$

and therefore the lemma follows form (119), (120), (121), and $\mathbf{T}_{\mathtt{s}},\mathbf{T}_{\mathtt{io}},\mathbf{T}_{\mathtt{i}}$, and $\mathbf{T}_{\mathtt{o}}$ forming a partition of $\mathbf{T}$. ■

A.2 Ordered Choice, ISP, and IOS

This appendix shows that latent-index ordered choice models generalizing (15) necessarily satisfy ISP and that these models satisfy IOS under additional, easily-verified restrictions. The auxiliary treatment variable V is defined here as in Theorem 2.2, while ε denotes an arbitrary scalar random variable. Although the lemma below holds regardless, when combined with ISP or IOS in pursuit of causal effects, ε is assumed to be independent of Z so that $T^*(z)$ is independent of Z.

Lemma A.4. *Assume discrete ordered treatment $T\in\mathbf{T}=\{J_l,\ldots,J^u\}$ and that treatment status is determined according to*

$$T^*(z)=\begin{cases} J_l & \text{if } z\pi+\varepsilon\le c_{J_l+1}\\ J_l+1 & \text{if } c_{J_l+1}<z\pi+\varepsilon\le c_{J_l+2}\\ \vdots & \\ J^u & \text{if } c_{J^u}<z\pi+\varepsilon\end{cases}$$

with $\pi>0$ and $c_{J_l+1}<c_{J_l+2}<\ldots<c_{J^u}$. Also set $c_{J_l}=-\infty$ and $c_{J^u+1}=+\infty$ and define the set $\mathcal{A}\equiv\{J_l\le t\le J^u: P(c_t<\varepsilon\le c_{t+1}-\pi)>0\}$. Given this choice model:

(i) ISP *holds with functions ν_0 and ν_1 given by $\nu_1(t)=|\{s\in\mathcal{A}:s<t\}|$ and*

$$\nu_0(t)=\begin{cases} 1-\nu_1(t) & \text{if } t\notin\mathcal{A}\\ -\nu_1(t) & \text{if } t\in\mathcal{A}\end{cases}. \tag{122}$$

(ii) If $c_{t+1}-c_t\le\pi$ for all $J_l+1\le t\le J^u-1$, then IOS *holds. If in addition ε is continuously*

distributed with support including the interval $[(c_{J_l+1} \wedge (c_{J_l+2} - \pi)) - \delta, c_{J^u} + \delta]$ *for some* $\delta > 0$*, then* IOS *holds if and only if* $c_{t+1} - c_t \leq \pi$ *for all* $J_l + 1 \leq t \leq J^u - 1$*, in which case*

$$V = 1\{T \in \mathbf{T}_{\mathtt{i}}\} + 1\{T \in \mathbf{T}_{\mathtt{io}}\}Z = 1\{T = J^u\} + 1\{J_l < T < J^u\}Z.$$

(iii) If Assumption 1.1 holds, $\pi \leq c_{t+1} - c_t$ *for all* $J_l + 1 \leq t \leq J^u - 1$ *and the conditional distribution of* ε *given* X *is continuous with support including* $[(c_{J_l+1} - \pi) - \delta, c_{J^u} + \delta]$ *for some* $\delta > 0$ *(with probability one over* X*), then*

$$\tau(\mathbf{S}) = \frac{E(E[Y|Z=1,X] - E[Y|Z=0,X])}{E(E[T|Z=1,X] - E[T|Z=0,X])}.$$

(iv) If ε *is continuously distributed with support including the interval* $[(c_{J_l+1} \wedge (c_{J_l+2} - \pi)) - \delta, (c_{J^u-1} \vee (c_{J^u} - \pi)) + \delta]$ *for some* $\delta > 0$*, then*

$$\mathbf{S} = \{(t, t') \in \mathbf{T} \times \mathbf{T} : c_{t'} - c_{t+1} < \pi < c_{t'+1} - c_t\}$$

and $P(T^* = t^*) > 0$ *for all* $t^* \in \mathbf{S}$.

Proof. First, note that for any $J_l \leq t \leq J^u$ it follows that $(T^*(0), T^*(1)) = (t, t)$ if and only if $c_t < \varepsilon$ and $\varepsilon \leq c_{t+1} - \pi$. Therefore, the ordered model is such that $\mathtt{stay}(\mathrm{t}) \neq \emptyset$ if and only if $P(c_t < \varepsilon \leq c_{t+1} - \pi) > 0$. Hence, by definition of $\mathcal{A}$ it follows that

$$\mathcal{A} = \{t \in \mathbf{T} : \mathtt{stay}(\mathrm{t}) \neq \emptyset\}. \tag{123}$$

Also observe that any switching type $T^* \in \mathbf{S}$ with $T^* = (T^*(0), T^*(1))$, corresponds to an ε satisfying $\varepsilon \leq c_{T^*(0)+1}$ and $\varepsilon + \pi > c_{T^*(1)}$ implying $\pi > c_{T^*(1)} - c_{T^*(0)+1}$. Since $c_{J_l+1} < c_{J_l+2} < \ldots < c_{J^u}$, the definition of $\mathcal{A}$ implies that for any switcher $T^* \in \mathbf{S}$ we have

$$\{s \in \mathcal{A} : T^*(0) < s < T^*(1)\} = \emptyset. \tag{124}$$

To verify ISP here, consider $T^* \notin \mathbf{S}$. Any such T^* satisfies $(T^*(0), T^*(1)) = (t, t)$ for some $t \in \mathcal{A}$ by result (123). Therefore, the definition of ν_0 in (122) implies

$$\nu_0(T^*(0)) + \nu_1(T^*(1)) = 0 \text{ for any } T^* \notin \mathbf{S}. \tag{125}$$

Next, consider any $T^* \in \mathbf{S}$, which implies $T^*(0) < T^*(1)$. If $T^*(0) \notin \mathcal{A}$, then (124) and the definition of ν_1 imply $\nu_1(T^*(0)) = \nu_1(T^*(1))$. Since $\nu_0(T^*(0)) = 1 - \nu_1(T^*(0))$ because $T^*(0) \notin \mathcal{A}$, it follows that $\nu_0(T^*(0)) + \nu_1(T^*(1)) = 1$. On the other hand, if $T^*(0) \in \mathcal{A}$, then $\nu_1(T^*(1)) = \nu_1(T^*(0)) + 1$ by (124) and the definition of ν_1. Since $\nu_0(T^*(0)) = -\nu_1(T^*(0))$ because $T^*(0) \in \mathcal{A}$, we have $\nu_0(T^*(0)) + \nu_1(T^*(1)) = 1$ and

$$\nu_0(T^*(0)) + \nu_1(T^*(1)) = 1 \text{ for any } T^* \in \mathbf{S}, \tag{126}$$

Together, results (125) and (126) verify ISP holds and hence part (i) follows.

To establish part (ii), note that $\pi > 0$ implies $\mathtt{in}(J_l) = \emptyset$ and $\mathtt{out}(J^u) = \emptyset$. Therefore, IOS holds if and only if for every $J_l < t < J^u$ at least one of $\mathtt{in}(\mathrm{t})$, $\mathtt{out}(\mathrm{t})$, or $\mathtt{stay}(\mathrm{t})$ is empty. However, $c_{t+1} - c_t \leq \pi$ for all $J_l + 1 \leq t \leq J^u - 1$ implies $P(c_t < \varepsilon \leq c_{t+1} - \pi) = 0$ for all $J_l + 1 \leq t \leq J^u - 1$. Therefore, result (123) implies $\mathtt{stay}(\mathrm{t}) = \emptyset$ for all $J_l + 1 \leq t \leq J^u - 1$, and therefore IOS holds.

Now, suppose ε is continuously distributed and its support includes $[(c_{J_l+1} \wedge (c_{J_l+2} - \pi)) - \delta, c_{J^u} + \delta]$ for some $\delta > 0$. For any $J_l + 1 \leq t \leq J^u - 1$, then note that a $T^* \in \texttt{out}(\text{t})$ corresponds to an ε satisfying $c_t < \varepsilon \leq c_{t+1}$ and $\varepsilon > c_{t+1} - \pi$. Hence, since

$$P(c_t \vee (c_{t+1} - \pi) < \varepsilon \leq c_{t+1}) > 0 \tag{127}$$

for all $J_l + 1 \leq t \leq J^u - 1$ due to our assumption on the support of ε, we obtain that $\texttt{out}(\text{t}) \neq \emptyset$ for all $J_l + 1 \leq t \leq J^u - 1$. Similarly, since any $T^* \in \texttt{in}(\text{t})$ corresponds to an ε satisfying $c_t < \varepsilon + \pi \leq c_{t+1}$ and $\varepsilon \leq c_t$, and our assumptions on ε imply

$$P(c_t - \pi < \varepsilon \leq c_t \wedge (c_{t+1} - \pi)) > 0 \tag{128}$$

for any $J_l + 1 \leq t \leq J^u - 1$, it follows that $\texttt{in}(\text{t}) \neq \emptyset$ for any $J_l + 1 \leq t \leq J^u - 1$. Therefore, IOS holds if and only if $\texttt{stay}(\text{t}) = \emptyset$ for all $J_l + 1 \leq t \leq J^u - 1$. By result (123) and our assumptions on ε, this condition is equivalent to $c_{t+1} - c_t \leq \pi$ for all $J_l + 1 \leq t \leq J^u - 1$, and therefore the lemma follows.

For part (iii), note that the fact that ε is continuously distributed with support including $[(c_{J_l+1} - \pi) - \delta, c_{J^u} + \delta]$ for some $\delta > 0$ implies that $\mathcal{A} = \{J_l, \ldots, J^u\}$. By part (i), therefore, $\nu_1(t) = t - J_l$ and $\nu_0(t) = -t + J_l$. Equivalently, $\tilde{\nu}_1(t) = -\tilde{\nu}_0(t) = t$ also satisfy ISP, so we can drop the additive shift, J_l. Part (iii) then follows from Theorem 2.2.

Finally, to establish part (iv) of the lemma, consider type $t^* = (t, t') \in \mathbf{T} \times \mathbf{T}$ with $t \neq t'$. This types has ε satisfying $c_t < \varepsilon \leq c_{t+1}$ and $c_{t'} < \pi + \varepsilon \leq c_{t'+1}$. Under the support condition in the statement of part (iv), $\max\{c_t, c_{t'} - \pi\} < \min\{c_{t'+1} - \pi, c_{t+1}\}$ if, and only if, $P(T^* = t^*) > 0$. Given $c_{J_l+1} < \cdots < c_{J^u}$, $\max\{c_t, c_{t'} - \pi\} < \min\{c_{t'+1} - \pi, c_{t+1}\}$ if, and only if, $c_t < c_{t'+1} - \pi$ and $c_{t'} - \pi < c_{t+1}$. ■

A.3 Estimation and Inference

Section 3 discusses estimates of SLATE and the CCEF constructed using the IV estimands in Theorems 2.2 and 2.7, respectively, $\kappa^\star$-based estimation of SLATE, as well as the CCEF weighted least squares estimand in Theorem 2.8. These are two-step estimates in which the first step is to estimate a (possibly parametric) estimate of the conditional distribution of (T, Z) (or Z alone) given X. Although the causal estimands of interest here are novel, the properties of the estimators they generate follow from standard results on the distribution of generalized method of moments (GMM) estimators. These properties are discussed below.

SLATE Estimators Using Theorems 2.2 and 2.3

The IV estimand for SLATE in Theorem 2.2 is

$$\tau(\mathbf{S}) = \frac{E[\tilde{Z}Y]}{E[\tilde{Z}V]}.$$

As for the estimates in Section 3, the propensity score is assumed to be estimated by a parametric model like probit or logit. This modek is written $P_\alpha(Z = 1|X)$ to denote dependence on parameter vector, α. Estimates of α, denoted $\hat{\alpha}$, may be obtained by maximum likelihood, non-linear least squares, or covariate balancing as described in Imai and Ratkovic (2014).

Given a first-step estimate of $\hat{\alpha}$, and defining

$$R_i(\hat{\alpha}) \equiv \frac{Z_i - P_{\hat{\alpha}}(Z=1|X_i)}{P_{\hat{\alpha}}(Z=1|X_i)(1-P_{\hat{\alpha}}(Z=1|X_i))},$$

$\tilde{Z}$ is consistently estimated by

$$\tilde{Z}_i(\hat{\alpha}) \equiv R_i(\hat{\alpha}) - \frac{1}{N}\sum_{j=1}^{N} R_j(\hat{\alpha}). \tag{129}$$

When the sample mean of $R_i(\hat{\alpha})$ is zero, $\tilde{Z}_i(\hat{\alpha})$ equals $R_i(\hat{\alpha})$ and second-step estimates are invariant to additive shifts in Y or V.[25]

Setting $V_i \equiv \nu_1(T_i)Z_i - \nu_0(T_i)(1-Z_i)$ generates second-step estimators of SLATE and the probability of switching that can be written:

$$\hat{\tau}(\mathbf{S}) \equiv \frac{\sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})Y_i}{\sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})V_i} \qquad \text{and} \qquad \hat{P}(T^* \in \mathbf{S}) \equiv \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})V_i. \tag{130}$$

The asymptotic properties of these estimators follow from the fact that they have a GMM representation. Specifically, $\hat{\tau}(\mathbf{S})$ is the value of b_1 that solves

$$\frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} Y - b_0 - b_1 V_i \\ (Y - b_0 - b_1 V_i)R_i(\hat{\alpha}) \end{pmatrix} = 0. \tag{131}$$

When $\hat{\alpha}$ also has a GMM representation, we can account for first-step estimation of $\hat{\alpha}$ by adding the relevant moment conditions to (131). Similarly, the value of b_1 that solves

$$\frac{1}{N}\sum_{i=1}^{N} \begin{pmatrix} R_i(\hat{\alpha}) - b_0 \\ V_i(R_i(\hat{\alpha}) - b_0) - b_1 \end{pmatrix} = 0, \tag{132}$$

is a GMM estimator of $\hat{P}(T^* \in \mathbf{S})$. The GMM framework easily accommodates clustered data, as in our Head Start application.

Under IOS, Theorem 2.3(ii) identifies SLATE as the solution to a weighted least squares problem, with weights equal to κ^*. Estimated SLATE in Panel B of Table 1 uses $\kappa^\star$ constructed by plugging (129) into (18). The corresponding switcher probability is estimated as the sample mean of $\kappa^\star$. The terms $E[h_0(X)|T^* \in \mathbf{S}]$ and $E[h_1(X)|T^* \in \mathbf{S}]$ in Theorem 2.3 are then obtained by applying Theorem 2.5(i). Finally, SLATE is estimated by solving the first order conditions for the sample analog of the minimization problem in Theorem 2.3(ii). Standard errors and p-values are obtained by bootstrapping the procedure sketched in this paragraph.

[25] Maximum likelihood does not impose a sample mean of $R_i(\hat{\alpha})$ equal to zero. Słoczyński et al. (2025) discusses normalization of weighting estimators in the context of basic LATE.

CCEF IV Estimators Using Theorem 2.7(ii)

Under the assumptions of Theorem 2.7(ii), the CCEF depends only on the treatment level and whether $T^* \in \texttt{in}(\mathrm{t})$ or $T^* \in \texttt{out}(\mathrm{t})$. Denote these values of the CCEF by

$$\begin{aligned} \gamma_t &\equiv E[Y^*(t)|T^*=t^*] && \text{if } t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \\ \gamma_{t,\mathtt{o}} &\equiv E[Y^*(t)|T^*=t^*] && \text{if } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \texttt{out}(\mathrm{t}) \\ \gamma_{t,\mathtt{i}} &\equiv E[Y^*(t)|T^*=t^*] && \text{if } t \in \mathbf{T}_{\mathtt{io}},\ t^* \in \texttt{in}(\mathrm{t})\,. \end{aligned}$$

Theorem 2.7(ii) shows that these parameters are identified by the IV estimands:

$$\begin{aligned} \gamma_t &= \frac{E[\tilde{Z} \times 1\{T=t\}Y]}{E[\tilde{Z} \times 1\{T=t\}]} && \text{for any } t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \\ \gamma_{t,\mathtt{o}} &= \frac{E[\tilde{Z} \times 1\{T=t\}(1-Z)Y]}{E[\tilde{Z} \times 1\{T=t\}(1-Z)]} && \text{for any } t \in \mathbf{T}_{\mathtt{io}} \\ \gamma_{t,\mathtt{i}} &= \frac{E[\tilde{Z} \times 1\{T=t\}ZY]}{E[\tilde{Z} \times 1\{T=t\}Z]} && \text{for any } t \in \mathbf{T}_{\mathtt{io}}. \end{aligned} \tag{133}$$

As with SLATE, the CCEF is estimated using sample analogs of the corresponding IV estimands with $\tilde{Z}_i(\hat{\alpha})$ (as defined in (129)) as instrument. These CCEF estimators are:

$$\begin{aligned} \hat{\gamma}_t &= \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}Y_i}{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}} && \text{for any } t \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{o}} \\ \hat{\gamma}_{t,\mathtt{o}} &= \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}(1-Z_i)Y_i}{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}(1-Z_i)} && \text{for any } t \in \mathbf{T}_{\mathtt{io}} \\ \hat{\gamma}_{t,\mathtt{i}} &= \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}Z_iY_i}{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha}) 1\{T_i=t\}Z_i} && \text{for any } t \in \mathbf{T}_{\mathtt{io}}. \end{aligned}$$

Similarly, $P(T^* \in \texttt{in}(\mathrm{t}))$ and $P(T^* \in \texttt{out}(\mathrm{t}))$ in Lemma 2.2 are consistently estimated by the corresponding sample analogs of these parameters:

$$\hat{P}(T^* \in \texttt{in}(\mathrm{t})) = \begin{cases} \frac{1}{N}\sum_{i=1}^N 1\{T_i=t\}\tilde{Z}_i(\hat{\alpha}) & \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ \frac{1}{N}\sum_{i=1}^N 1\{T_i=t\}Z_i\tilde{Z}_i(\hat{\alpha}) & \text{if } t \in \mathbf{T}_{\mathtt{io}} \end{cases} \tag{134}$$

and

$$\hat{P}(T^* \in \texttt{out}(\mathrm{t})) = \begin{cases} -\frac{1}{N}\sum_{i=1}^N 1\{T_i=t\}\tilde{Z}_i(\hat{\alpha}) & \text{if } t \in \mathbf{T}_{\mathtt{o}} \\ -\frac{1}{N}\sum_{i=1}^N 1\{T_i=t\}(1-Z_i)\tilde{Z}_i(\hat{\alpha}) & \text{if } t \in \mathbf{T}_{\mathtt{io}}. \end{cases} \tag{135}$$

Limiting distributions again follow from the GMM representation of IV estimation. This includes inference for causal effects, since these are linear combinations of CCEF levels.

CCEF and and SLATE are compatible in the sense that SLATE can be obtained as a weighted average of the elements of the CCEF. The next lemma shows this holds both for parameters, estimands, and for the IV estimators detailed in this appendix.

Lemma A.5. SLATE *and* CCEF *parameters and estimators are related as follows.*

(i) Suppose Assumptions 1.1 and 1.2 hold. Then SLATE *and the* CCEF *satisfy:*

$$\tau(\mathbf{S}) = \sum_{t \in \mathbf{T_i} \cup \mathbf{T_{io}}} E[Y^*(t)|T^* \in \mathtt{in}(t)] \times \frac{P(T^* \in \mathtt{in}(t))}{P(T^* \in \mathbf{S})} - \sum_{t \in \mathbf{T_o} \cup \mathbf{T_{io}}} E[Y^*(t)|T^* \in \mathtt{out}(t)] \times \frac{P(T^* \in \mathtt{out}(t))}{P(T^* \in \mathbf{S})}. \quad (136)$$

(ii) Suppose Assumptions 1.1, 1.2 hold, and $E[V\tilde{Z}] \neq 0$. Then (136) holds replacing SLATE *by the corresponding estimand in Theorem 2.2, replacing the* CCEF *by the corresponding estimand in Theorem 2.7(ii), and replacing switcher probabilities by the corresponding estimands in Lemma 2.2. In terms of estimands, we have:*

$$\frac{E[Y\tilde{Z}]}{E[V\tilde{Z}]} = \sum_{t \in \mathbf{T_i}} \frac{E[\tilde{Z}1\{T=t\}Y]}{E[V\tilde{Z}]} + \sum_{t \in \mathbf{T_{io}}} \frac{E[\tilde{Z}1\{T=t\}ZY]}{E[V\tilde{Z}]} + \sum_{t \in \mathbf{T_o}} \frac{E[\tilde{Z}1\{T=t\}Y]}{E[V\tilde{Z}]} + \sum_{t \in \mathbf{T_{io}}} \frac{E[\tilde{Z}1\{T=t\}(1-Z)Y]}{E[V\tilde{Z}]}. \quad (137)$$

(iii) Suppose $\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha}) \neq 0$ and $\sum_{i=1}^N Y_i\tilde{Z}_i(\hat{\alpha})1\{T_i \in \mathbf{T_s}\} = 0$ (for example, when $\mathbf{T_s} = \emptyset$). Then (137) holds for estimators in the form of sample analogs:

$$\frac{\sum_{i=1}^N Y_i\tilde{Z}_i(\hat{\alpha})}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} = \sum_{t \in \mathbf{T_i}} \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}Y_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} + \sum_{t \in \mathbf{T_{io}}} \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}Z_iY_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} + \sum_{t \in \mathbf{T_o}} \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}Y_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} + \sum_{t \in \mathbf{T_{io}}} \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}(1-Z_i)Y_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})}. \quad (138)$$

Part (i) of this lemma aligns CCEF and SLATE parameters. Part (ii) shows the equality holds for estimands as well as parameters. This result requires neither ISP or IOS, and thus holds under weaker conditions than the results that justify the estimands. Part (iii) says that our IV estimators of SLATE and the CCEF are similarly compatible under minimal assumptions. This result requires only a non-zero first stage for treatment V and that $\sum_{i=1}^N Y_i\tilde{Z}_i(\hat{\alpha})1\{T_i \in \mathbf{T_s}\} = 0$. The latter condition assuredly holds whenever $\mathbf{T_s}$ is empty.

Some specifications of $\mathbf{M}$ facilitate identification of the distribution of response types T^*. In this case, Lemma A.5(iii) implies that the SLATE estimator equals a weighted average of switcher-specific effects. In the Head Start application, for instance, $\mathbf{T_i} = \{h\}$, $\mathbf{T_o} = \{n, c\}$, and $\mathbf{T_{io}}$ and $\mathbf{T_s}$ are empty. Since $\mathtt{out}(c) = \{(c,h)\}$ and $\mathtt{out}(n) = \{(n,h)\}$, we have

$$P(T^* = (c,h)) = P(T^* \in \mathtt{out}(c)) \qquad P(T^* = (n,h)) = P(T^* \in \mathtt{out}(n)). \quad (139)$$

By Lemma 2.2 the probability of each of these switcher type is identified. Using (135), set

$$\hat{P}(T^* = (c,h)) = -\frac{1}{N}\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i = c\} \qquad \hat{P}(T^* = (n,h)) = -\frac{1}{N}\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i = n\}$$

and note that

$$\begin{aligned}\hat{P}(T^* \in \texttt{in}(h)) &= \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})1\{T_i = h\}\\ &= \frac{1}{N}\sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})(1 - 1\{T_i = c\} - 1\{T_i = n\})\\ &= \hat{P}(T^* = (n,h)) + \hat{P}(T^* = (c,h)),\end{aligned}$$

because the mean of $\tilde{Z}_i(\hat{\alpha})$ is zero by construction. Lemma A.5(iii) therefore implies

$$\begin{aligned}\frac{\sum_{i=1}^N Y_i\tilde{Z}_i(\hat{\alpha})}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} &= \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i = h\}Y_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})} + \sum_{t\in\{c,n\}} \frac{\sum_{i=1}^N \tilde{Z}_i(\hat{\alpha})1\{T_i = t\}Y_i}{\sum_{i=1}^N V_i\tilde{Z}_i(\hat{\alpha})}\\ &= \hat{\gamma}_h\frac{\hat{P}(T^* \in \texttt{in}(h))}{\hat{P}(T^* \in \mathbf{S})} - (\hat{\gamma}_c\frac{\hat{P}(T^* \in \texttt{out}(c))}{\hat{P}(T^* \in \mathbf{S})} + \hat{\gamma}_n\frac{\hat{P}(T^* \in \texttt{out}(n))}{\hat{P}(T^* \in \mathbf{S})})\\ &= (\hat{\gamma}_c - \hat{\gamma}_n)\frac{\hat{P}(T^* = (n,c))}{\hat{P}(T^* \in \mathbf{S})} + (\hat{\gamma}_c - \hat{\gamma}_h)\frac{\hat{P}(T^* = (c,h))}{\hat{P}(T^* \in \mathbf{S})}.\end{aligned}$$

This explains why estimated SLATE for Head Start in Section 3.1 equals a weighted average of the corresponding CCEF estimates.

Proof of Lemma A.5. In order to establish part (i), we first equivalently express SLATE as

$$\tau(\mathbf{S}) = \frac{1}{P(T^* \in \mathbf{S})} \times \sum_{t^*\in\mathbf{S}} E[(Y^*(t^*(1)) - Y^*(t^*(0)))1\{T^* = t^*\}]. \tag{140}$$

Since for any response type $t^* \in \mathbf{S}$ we must have $t^*(1) \in \mathbf{T}_{\mathtt{i}} \cup \mathbf{T}_{\mathtt{io}}$, it therefore follows that

$$\begin{aligned}&\sum_{t^*\in\mathbf{S}} E[Y^*(t^*(1))1\{T^* = t^*\}] = \sum_{t^*\in\mathbf{S}}\sum_{t\in\mathbf{T}_{\mathtt{i}}\cup\mathbf{T}_{\mathtt{io}}} E[Y^*(t)1\{T^* = t^*\}1\{t^*(1) = t\}]\\ &= \sum_{t\in\mathbf{T}_{\mathtt{i}}\cup\mathbf{T}_{\mathtt{io}}} E[Y^*(t)(\sum_{t^*\in\mathbf{S}} 1\{T^* = t^*\}1\{t^*(1) = t\})] = \sum_{t\in\mathbf{T}_{\mathtt{i}}\cup\mathbf{T}_{\mathtt{io}}} E[Y^*(t)1\{T^* \in \texttt{in}(\mathrm{t})\}], \end{aligned}\tag{141}$$

where the final equality follows by definition of $\texttt{in}(\mathrm{t})$. By similar arguments we have that

$$\sum_{t^*\in\mathbf{S}} E[Y^*(t^*(0))1\{T^* = t^*\}] = \sum_{t\in\mathbf{T}_{\mathtt{o}}\cup\mathbf{T}_{\mathtt{io}}} E[Y^*(t)1\{T^* \in \texttt{out}(\mathrm{t})\}] \tag{142}$$

and therefore the second claim of the lemma follows from results (140)-(142).

To establish part (ii) just note that $\mathbf{T}_{\mathtt{i}}$, $\mathbf{T}_{\mathtt{o}}$, $\mathbf{T}_{\mathtt{io}}$ and $\mathbf{T}_{\mathtt{s}}$ being a partition of $\mathbf{T}$ imply

$$\begin{aligned}\sum_{t\in\mathbf{T}_{\mathtt{i}}\cup\mathbf{T}_{\mathtt{o}}} E[\tilde{Z}1\{T = t\}Y] &+ \sum_{t\in\mathbf{T}_{\mathtt{io}}} (E[\tilde{Z}1\{T = t\}ZY] + E[\tilde{Z}1\{T = t\}(1 - Z)Y]) \\ &= E[\tilde{Z}Y] - \sum_{t\in\mathbf{T}_{\mathtt{s}}} E[\tilde{Z}1\{T = t\}Y]\\ &= E[\tilde{Z}Y] - \sum_{t\in\mathbf{T}_{\mathtt{s}}} E[\tilde{Z}1\{T^* \in \texttt{stay}(\mathrm{t})\}Y^*(t)]\\ &= E[\tilde{Z}Y],\end{aligned}\tag{143}$$

where the second equality follows from $1\{T = t\} = 1\{T^* \in \texttt{stay}(\mathrm{t})\}$, and the third from $E[\tilde{Z}|X] = 0$ and $\tilde{Z}$ being independent of (T^*, Y^*) conditionally on X by Assumption 1.1(iii).

Part (ii) of the lemma therefore follows from (143) and $E[V\tilde{Z}] \neq 0$.

Finally, to establish part (iii) we follow the same manipulations as in (143) to obtain

$$\sum_{i=1}^{N}\Big(\sum_{t\in\mathbf{T_i}\cup\mathbf{T_o}} \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}Y_i + \sum_{t\in\mathbf{T_{io}}} (\tilde{Z}_i(\hat{\alpha})1\{T_i=t\}Z_iY_i + \tilde{Z}_i(\hat{\alpha})1\{T_i=t\}(1-Z_i)Y_i)\Big)$$
$$= \sum_{i=1}^{N} \tilde{Z}_i(\hat{\alpha})Y_i - \sum_{i=1}^{N}\sum_{t\in\mathbf{T_s}} \tilde{Z}_i(\hat{\alpha})Y_i 1\{T_i=t\} = \sum_{i=1}^{N}\tilde{Z}_i(\hat{\alpha})Y_i, \quad (144)$$

where the final equality follows from $\sum_{i=1}^{N} Y_i\tilde{Z}_i(\hat{\alpha})1\{T_i \in \mathbf{T_s}\} = 0$. The third claim of the lemma therefore follows form (144) and $\sum_{i=1}^{N}\tilde{Z}_i(\hat{\alpha})V_i \neq 0$. ■

CCEF Estimators Based on Theorem 2.8

Theorem 2.8 identifies the CCEF as the solution to a population weighted least squares problem. Distribution theory for estimators based on this result is more involved than for SLATE and CCEF estimates using IV.

Note first that weights and outcomes in the least squares problem here depend on the joint conditional probability $P(T=t, Z=z|X)$ rather than the instrument-assignment propensity score. As with the propensity score, this conditional probability is estimated using a parametric model, $P_\alpha(T=t, Z=z|X)$, that depends on parameter vector α. Again, let $\hat{\alpha}$ denote the corresponding first-step estimator.

First-step estimator $\hat{\alpha}$ is used to construct plug-in estimators $\tilde{Y}_t(\hat{\alpha})$ for $\tilde{Y}_t$, $\omega_t(X;\hat{\alpha})$ for $\omega_t(X)$ (for all $t \in \mathbf{T_i}\cup\mathbf{T_o}$), $\tilde{Y}_{t,z}(\hat{\alpha})$ for $\tilde{Y}_{t,z}$, and $\omega_{t,z}(X;\hat{\alpha})$ for $\omega_{t,z}(X)$ (for all $t\in\mathbf{T_{io}}$ and $z\in\{0,1\}$). Specifically, factoring $P_{\hat{\alpha}}(T=t, Z=z|X)$, we set

$$\tilde{Y}_t(\hat{\alpha}) \equiv \frac{(Z - P_{\hat{\alpha}}(Z=1|X))1\{T=t\}Y}{P_{\hat{\alpha}}(Z=1, T=t|X) - P_{\hat{\alpha}}(Z=1|X)P_{\hat{\alpha}}(T=t|X)}$$
$$\tilde{Y}_{t,0}(\hat{\alpha}) \equiv \frac{1\{T=t\}(1-Z)Y}{P_{\hat{\alpha}}(T=t, Z=0|X)}$$
$$\tilde{Y}_{t,1}(\hat{\alpha}) \equiv \frac{1\{T=t\}ZY}{P_{\hat{\alpha}}(T=t, Z=1|X)}$$

and

$$\omega_t(X;\hat{\alpha}) = P_{\hat{\alpha}}(Z=1|X)(P_{\hat{\alpha}}(T=t|Z=1,X) - P_{\hat{\alpha}}(T=t|Z=0,X))_+ \quad \text{if } t\in\mathbf{T_i}$$
$$\omega_t(X;\hat{\alpha}) = P_{\hat{\alpha}}(Z=0|X)(P_{\hat{\alpha}}(T=t|Z=0,X) - P_{\hat{\alpha}}(T=t|Z=1,X))_+ \quad \text{if } t\in\mathbf{T_o}$$
$$\omega_{t,z}(X;\hat{\alpha}) = P_{\hat{\alpha}}(T=t, Z=z|X) \quad \text{if } t\in\mathbf{T_{io}}.$$

CCEF estimator $\hat{\beta}$ minimizes objective function Q_n over parameter space $\mathcal{B}$, where

$$Q_N(b) \equiv \frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{t\in\mathbf{T_i}\cup\mathbf{T_o}} (\tilde{Y}_{t,i}(\hat{\alpha}) - F_1(t, X_i; b))^2\omega_t(X_i;\hat{\alpha})$$
$$+ \sum_{t\in\mathbf{T_{io}}, z\in\{0,1\}} (\tilde{Y}_{t,z,i}(\hat{\alpha}) - F_2(t,z,X_i;b))^2\omega_{t,z}(X_i;\hat{\alpha})\Big).$$

Under conditions stated below, $\hat{\beta}$ is consistent for the β that minimizes Q over $\mathcal{B}$, where

$$Q(b) \equiv E\Big[\sum_{t \in \mathbf{T_i} \cup \mathbf{T_o}} (\tilde{Y}_t(\alpha) - F_1(t, X; b))^2 \omega_t(X; \alpha) + \sum_{t \in \mathbf{T_{io}}, z \in \{0,1\}} (\tilde{Y}_{t,z}(\alpha) - F_2(t, z, X; b))^2 \omega_{t,z}(X; \alpha) \Big].$$

When the model for the conditional probability $P(T = t, Z = z|X)$ is correctly specified, $\tilde{Y}_t(\alpha)$, $\tilde{Y}_{t,z}(\alpha)$, $\omega_t(X; \alpha)$, and $\omega_{t,z}(X; \alpha)$ equal $\tilde{Y}_t$, $\tilde{Y}_{t,z}$, $\omega_t(X)$, and $\omega_{t,z}(X)$, respectively. In this case, minimizer β of Q minimizes the objective function given by (36).

The asymptotic properties of first-step estimator $\hat{\alpha}$ are derived under standard conditions for GMM estimators. Specifically, let $W \equiv (Y, T, Z, X)$ denote the vector of observed variables, and suppose the first-step estimator $\hat{\alpha}$ is chosen to solve

$$\frac{1}{N} \sum_{i=1}^{N} m_\alpha(W_i, \hat{\alpha}) = 0, \tag{145}$$

with probability approaching one, where $m_\alpha(W, \hat{\alpha})$ is of the same dimension as $\hat{\alpha}$. For any matrix Ω, also let $\|\Omega\|_o$ denote the square root of the largest eigenvalue of $\Omega'\Omega$ (i.e., the largest singular value of Ω). Estimator $\hat{\alpha}$ is then assumed to satisfy:

Assumption A.1. *(i) $\hat{\alpha} \xrightarrow{p} \alpha$; (ii) $E[m_\alpha(W, \alpha)] = 0$ and $E[\|m_\alpha(W, \alpha)\|^2] < \infty$; (iii) With probability approaching one (145) holds; (iv) There is a neighborhood $\mathcal{N}_\alpha$ of α such that $m_\alpha(w, a)$ is differentiable in $a \in \mathcal{N}_\alpha$ and $E[\sup_{a \in \mathcal{N}_\alpha} \|\nabla_a m_\alpha(W, a)\|_o] < \infty$; (v) The matrix $E[\nabla_a m_\alpha(W, \alpha)]$ is invertible.*

Assumption A.1(i) requires $\hat{\alpha}$ to be consistent for α. Assumptions A.1(ii)-(v) are standard regularity conditions that guarantee asymptotic normality of $\hat{\alpha}$.

Consistency of $\hat{\beta}$ follows from Assumption A.1(i) and the following two assumptions:

Assumption A.2. *(i) β is the unique minimizer of Q in $\mathcal{B}$; (ii) $\mathcal{B}$ is compact; (iii) $F_1(t, X; \cdot)$ is continuous and $E[\sup_{b \in \mathcal{B}} F_1^2(t, X; b)] < \infty$ for all $t \in \mathbf{T_i} \cup \mathbf{T_o}$; (iv) $F_2(t, z, X; \cdot)$ is continuous and $E[\sup_{b \in \mathcal{B}} F_1^2(t, z, X; b)] < \infty$ for all $t \in \mathbf{T_{io}}$, $z \in \{0, 1\}$.*

Assumption A.3. *(i) There is a neighborhood $\mathcal{N}_\alpha$ of α and $M < \infty$ with $\omega_t(X; a)$ and $\tilde{Y}_t(a)$ continuous in $a \in \mathcal{N}_\alpha$, $E[\sup_{a \in \mathcal{N}_\alpha} \tilde{Y}_t^2(a)] < \infty$, $P(\sup_{a \in \mathcal{N}_\alpha} |\omega_t(X; a)| \leq M) = 1$ for all $t \in \mathbf{T_i} \cup \mathbf{T_o}$; (ii) $\omega_{t,z}(X; a)$ and $\tilde{Y}_{t,z}(a)$ are continuous in $a \in \mathcal{N}_\alpha$ and satisfy $E[\sup_{a \in \mathcal{N}_\alpha} \tilde{Y}_{t,z}^2(a)] < \infty$ and $P(\sup_{a \in \mathcal{N}_\alpha} |\omega_{t,z}(X; a)| \leq M)$ for all $t \in \mathbf{T_{io}}$, $z \in \{0, 1\}$.*

Assumption A.3 requires the weights $\omega_t(X; \alpha)$ (for $t \in \mathbf{T_i} \cup \mathbf{T_o}$) and $\omega_{t,z}(X; \alpha)$ (for $t \in \mathbf{T_{io}}$) to be bounded. This condition is sure to be satisfied in first-step specifications for which $P_\alpha(T = t, Z = z|X)$ belongs to $[0, 1]$.

Assumptions yielding asymptotic normality of $\hat{\beta}$ use the notation

$$m_\beta(W, b, a) \equiv \sum_{t \in \mathbf{T_i} \cup \mathbf{T_o}} (\tilde{Y}_t(a) - F_1(t, X; b)) \omega_t(X; a) \nabla_b F_1(t, X; b) + \sum_{t \in \mathbf{T_{io}}, z \in \{0,1\}} (\tilde{Y}_{t,z}(a) - F_2(t, z, X; b)) \omega_{t,z}(X; a) \nabla_b F_2(t, z, X; b),$$

so that $E[m_\beta(W,b,\alpha)]$ equals the derivative of Q at b. We then require:

Assumption A.4. *(i) There is a neighborhood $\mathcal{N}_\beta$ of β with $\mathcal{N}_\beta \subset \mathcal{B}$; (ii) $F_1(t,X;b)$ is differentiable in $b \in \mathcal{N}_\beta$, $E[\|\nabla_b F_1(t,X;\beta)\|^2] < \infty$ for all $t \in \mathbf{T}_\mathtt{i} \cup \mathbf{T}_\mathtt{o}$; (iii) $F_2(t,z,X;b)$ is differentiable in $b \in \mathcal{N}_\beta$, $E[\|\nabla_b F_2(t,z,X;\beta)\|^2] < \infty$ for all $t \in \mathbf{T}_\mathtt{io}$, $z \in \{0,1\}$; (iv) $m_\beta(W,b,a)$ is differentiable in $b \in \mathcal{N}_\beta$ for all $a \in \mathcal{N}_\alpha$, $E[\sup_{b\in\mathcal{N}_\beta, a\in\mathcal{N}_a} \|\nabla_b m_\beta(W,b,a)\|_o] < \infty$; (v) $E[\|m_\beta(W,\beta,\alpha)\|^2] < \infty$ and $E[\nabla_b m_\beta(W,\beta,\alpha)]$ are invertible.*

Assumption A.5. *(i) $\sup_{b\in\mathcal{N}_\beta} \|m_\beta(W,b,a_1) - m_\beta(W,b,a_2)\| \le D(W)\|a_1 - a_2\|$ for all $a_1, a_2 \in \mathcal{N}_\alpha$ and D with $E[D^2(W)] < \infty$; (ii) $\tilde{Y}_t(a)$ is differentiable in $a \in \mathcal{N}_\alpha$ for all $t \in \mathbf{T}_\mathtt{i} \cup \mathbf{T}_\mathtt{o}$ and $E[\sup_{a\in\mathcal{N}_\alpha} \|\nabla_b F_1(t,X;\beta)(\nabla_a \tilde{Y}_t(a))'\|_o] < \infty$; (iii) $\tilde{Y}_{t,z}(a)$ is differentiable in $a \in \mathcal{N}_\alpha$ for all $t \in \mathbf{T}_\mathtt{io}$, $z \in \{0,1\}$, and $E[\sup_{a\in\mathcal{N}_\alpha} \|\nabla_b F_2(t,z,X;\beta)(\nabla_a \tilde{Y}_{t,z}(a))'\|_o] < \infty$.*

Assumption A.4 imposes standard conditions under which $\hat{\beta}$ is asymptotically normal when α is known. Assumption A.5 allows us to account for the impact of estimated $\hat{\alpha}$ on the asymptotic distribution of $\hat{\beta}$. Assumption A.5 requires $\tilde{Y}_t(\alpha)$ and $\tilde{Y}_{t,z}(\alpha)$ to be differentiable, but does not impose differentiability on weights $\omega_t(X;\alpha)$ and $\omega_{t,z}(X;\alpha)$.

The contribution of $\hat{\alpha}$ to the influence function of $\hat{\beta}$ depends on

$$G_y \equiv \sum_{t\in\mathbf{T}_\mathtt{i}\cup\mathbf{T}_\mathtt{o}} \omega_t(X;\alpha)\nabla_b F_1(t,X;\beta)(\nabla_a \tilde{Y}_t(\alpha))' + \sum_{t\in\mathbf{T}_\mathtt{io}, z\in\{0,1\}} \omega_{t,z}(X;\alpha)\nabla_b F_2(t,z,X;\beta)(\nabla_a \tilde{Y}_{t,z}(\alpha))'$$

and

$$G_\omega \equiv E\Big[\sum_{t\in\mathbf{T}_\mathtt{i}\cup\mathbf{T}_\mathtt{o}} (\tilde{Y}_t(\alpha) - F_1(t,X;\beta))\nabla_b F_1(t,X;\beta)(\nabla_a \omega_t(X;\alpha))' + \sum_{t\in\mathbf{T}_\mathtt{io}, z\in\{0,1\}} (\tilde{Y}_{t,z}(\alpha) - F_2(t,z,X;\beta))\nabla_b F_2(t,z,X;\beta)(\nabla_a \omega_{t,z}(X;\alpha))'\Big].$$

Define matrices $M_\alpha \equiv E[\nabla_a m_\alpha(W,\alpha)]$ and $M_\beta \equiv E[\nabla_b m_\beta(W,\beta,\alpha)]$. The next result uses this notation to establish the asymptotic normality of $\hat{\beta}$.

Theorem A.2. *Assume A.1, A.2, A.3 , A.4, A.5 hold and let $\{W_i\}_{i=1}^N$ be an i.i.d. sequence of random variables.*

(i) Suppose that $E[\tilde{Y}_t(\alpha)|X] = F_1(t,X;\beta)$ for all $t \in \mathbf{T}_\mathtt{i}\cup\mathbf{T}_\mathtt{o}$ and $E[\tilde{Y}_{t,z}(\alpha)|X] = F_2(t,z,X;\beta)$ for all $t \in \mathbf{T}_\mathtt{io}$ and $z \in \{0,1\}$. Then,

$$\sqrt{N}\{\hat{\beta} - \beta\} \xrightarrow{d} N(0, \mathrm{Var}\{M_\beta^{-1}(m_\beta(W,\beta,\alpha) + G_y M_\alpha^{-1} m_\alpha(W,\alpha))\}).$$

(ii) Suppose that $E[m_\beta(W,\beta,a)]$ is continuously differentiable in $a \in \mathcal{N}_\alpha$. Then,

$$\sqrt{N}\{\hat{\beta} - \beta\} \xrightarrow{d} N(0, \mathrm{Var}\{M_\beta^{-1}(m_\beta(W,\beta,\alpha) + (G_y + G_\omega) M_\alpha^{-1} m_\alpha(W,\alpha))\}).$$

Part (i) of Theorem A.2 derives the asymptotic distribution of $\hat{\beta}$ assuming F matches the conditional mean of $\tilde{Y}_t(\alpha)$ (for $t \in \mathbf{T}_\mathtt{i} \cup \mathbf{T}_\mathtt{o}$) and of $\tilde{Y}_{t,z}(\alpha)$ (for $t \in \mathbf{T}_\mathtt{io}$). When the first step is correctly specified, these implies F matches the CCEF. In this case, the fact that weights are estimated does not change the asymptotic variance of $\hat{\beta}$ – i.e.,weighting by $\omega_t(X;\hat{\alpha})$ is asymptotically equivalent to weighting by $\omega_t(X;\alpha)$.

Part (ii) of Theorem A.2 derives the asymptotic distribution of $\hat{\beta}$ assuming without requiring the CCEF model be correctly specified. In this case, estimation error in the weights can affect the asymptotic variance of $\hat{\beta}$. To characterize this impact, Theorem A.2(ii) assumes $E[m_\beta(W,\beta,a)]$ is differentiable in a, which requires the weights themselves to be differentiable. The latter condition requires that, for some $\varepsilon > 0$,

$$\begin{aligned} &P_\alpha(T=t|Z=1,X) - P_\alpha(T=t|Z=0,X) \geq \varepsilon \qquad \text{if } t \in \mathbf{T}_{\mathtt{i}} \\ &P_\alpha(T=t|Z=0,X) - P_\alpha(T=t|Z=1,X) \geq \varepsilon \qquad \text{if } t \in \mathbf{T}_{\mathtt{o}}. \end{aligned} \tag{146}$$

If this condition fails, $\hat{\beta}$ remains $\sqrt{N}$-consistent but its asymptotic distribution may not be Gaussian (Fang and Santos, 2019).

Standard errors and confidence intervals for the estimates based on Theorem 2.8 in Table 1 use a bootstrap. We assume the CCEF model is correctly specified, so the asymptotic distribution of $\hat{\beta}$ is the same using either estimated or population weights. We therefore fix weights at the value estimated in the full sample across bootstrap replications. The quantities $\tilde{Y}_t, \tilde{Y}_{t,z}$, and parameter estimates $\hat{\alpha}, \hat{\beta}$ are recomputed in each bootstrap sample. When the CCEF model is potentially misspecified, weights should also be re-estimated in each draw. In this case, condition (146) must be imposed. The consistency of both bootstrap procedures follows from the asymptotic equivalence of our estimator and GMM estimators and the consistency of the bootstrap for the latter (Hall and Horowitz, 1996; Hahn, 1996; Andrews, 2002); see also Gonçalves et al. (2023) for recent results on bootstrap consistency for two-step extremum estimators.

Given estimates $\hat{\beta}^{(b)}$ for bootstrap replications $b = 1, \ldots, B$, standard errors are computed as the square-root of $B^{-1}\sum_{b=1}^{B}(\hat{\beta}_j^{(b)} - \hat{\beta}_j)^2$. Confidence intervals at a $1-\alpha$ level are computed as $[\hat{\beta}_j - q_{(1-\alpha)/2}, \hat{\beta}_j - q_{\alpha/2}]$, where q_τ is the τ-th quantile of the distribution of $(\hat{\beta}_j^{(b)} - \hat{\beta}_j)_{b=1}^{B}$. Finally, p-values are computed using the bootstrap distribution of the t-statistic centered at $\hat{\beta}_j$.

Proof of Theorem A.2. Lemmas A.6 and A.7 used here are proved at the end of this section. By Lemma A.6, $\hat{\beta} \xrightarrow{p} \beta$. Moreover, Assumption A.1 and standard arguments (e.g., Theorem 3.4 in Newey and McFadden (1994)) imply

$$\sqrt{N}\{\hat{\alpha} - \alpha\} = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M_\alpha^{-1} m_\alpha(W_i,\alpha) + o_P(1), \tag{147}$$

where recall that $M_\alpha \equiv E[\nabla_a m_\alpha(W,\alpha)]$. Next, note that the consistency of $\hat{\beta}$, Assumption A.4(i), and Q_n being differentiable by Assumptions A.4(ii)(iii) imply $\hat{\beta}$ satisfies

$$\frac{1}{N}\sum_{i=1}^{N} m_\beta(W_i,\hat{\beta},\hat{\alpha}) = 0. \tag{148}$$

Consistency of $\hat{\alpha}$ (by Assumption A.1(i)) and $\hat{\beta}$, and Lemma A.7 then imply that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_\beta(W_i,\hat{\beta},\hat{\alpha}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_\beta(W_i,\beta,\alpha) + \sqrt{N}E[(m_\beta(W,\hat{\beta},\hat{\alpha}) - m_\beta(W,\beta,\alpha))] + o_P(1). \tag{149}$$

In particular, combining results (148) and (149) and rearranging terms yields that

$$\sqrt{N}E[(m_\beta(W,\beta,\hat{\alpha}) - m_\beta(W,\hat{\beta},\hat{\alpha}))] = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_\beta(W_i,\beta,\alpha) + \sqrt{N}E[(m_\beta(W,\beta,\hat{\alpha}) - m_\beta(W,\beta,\alpha))] + o_P(1). \tag{150}$$

Next, note that $E[\|m_\beta(W,\beta,\alpha)\|^2] < \infty$ by Assumption A.4(v) and hence result (147) and Assumption A.1(ii) and Assumption A.5(i) together imply that the right hand side of (150) is $O_P(N^{-1/2})$. Therefore, the consistency of $(\hat{\beta},\hat{\alpha})$, the mean value theorem, Assumptions A.4(iv)(v), and result (150) allow us to conclude that

$$\sqrt{N}\{\hat{\beta}-\beta\} = -\frac{1}{\sqrt{N}}\sum_{i=1}^{N} M_\beta^{-1}(m_\beta(W_i,\beta,\alpha) + E[(m_\beta(W,\beta,\hat{\alpha}) - m_\beta(W,\beta,\alpha))]) + o_P(1) \tag{151}$$

where recall that $M_\beta \equiv E[\nabla_b m_\beta(W,\beta,\alpha)]$.

To establish the first part of the theorem, note that $E[\tilde{Y}_t(\alpha)|X] = F_1(t,X;\beta)$, $E[\tilde{Y}_{t,z}(\alpha)|X] = F_2(t,z,X;\beta)$, Assumptions A.3 and A.5(ii)(iii), and result (147) imply

$$\sqrt{N}E[(m_\beta(W,\beta,\hat{\alpha}) - m_\beta(W,\beta,\alpha))] = G_y M_\alpha^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_\alpha(W_i,\alpha) + o_P(1). \tag{152}$$

Hence, part (i) of the theorem follows from (151), (152), $E[m_\beta(W,\beta,\alpha)] = 0$ by β minimizing Q and Assumption A.4(ii), $E[m_\alpha(W,\alpha)] = 0$ by Assumption A.1(i), and the central limit theorem. If instead the conditions of part (ii) of the theorem hold, then

$$\sqrt{N}E[(m_\beta(W,\beta,\hat{\alpha}) - m_\beta(W,\beta,\alpha))] = (G_y + G_\omega)M_\alpha^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N} m_\alpha(W_i,\alpha) + o_P(1) \tag{153}$$

by result (147) and the delta method (see, e.g., Theorem 3.9.4 in van der Vaart and Wellner (1996)). Part (ii) of the theorem follows from (151), (153), and the central limit theorem. ■

Lemma A.6. *If Assumptions A.1(i), A.2, A.3 hold, and* $\{W_i\}_{i=1}^N$ *is i.i.d., then* $\hat{\beta} \xrightarrow{p} \beta$.

Proof. We first define a finite sample criterion function $\tilde{Q}_n(b,a)$ to be given by

$$\tilde{Q}_N(b,a) \equiv \frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{t\in\mathbf{T}_\mathrm{i}\cup\mathbf{T}_\mathrm{o}} (\tilde{Y}_{t,i}(a) - F_1(t,X_i;b))^2\omega_t(X_i;a) + \sum_{t\in\mathbf{T}_\mathrm{io}, z\in\{0,1\}} (\tilde{Y}_{t,z,i}(a) - F_2(t,z,X_i;b))^2\omega_{t,z}(X_i;a)\Big), \tag{154}$$

and note that $Q_N(b) = \tilde{Q}_N(b, \hat{\alpha})$. Similarly, we define the population analog to be

$$\tilde{Q}(b, a) \equiv E\Big[\sum_{t \in \mathbf{T_i} \cup \mathbf{T_o}} (\tilde{Y}_t(a) - F_1(t, X; b))^2 \omega_t(X; a) + \sum_{t \in \mathbf{T_{io}}, z \in \{0,1\}} (\tilde{Y}_{t,z}(a) - F_2(t, z, X; b))^2 \omega_{t,z}(X; a)\Big), \quad (155)$$

noting that $Q(b) = \tilde{Q}(b, \alpha)$. Assumptions A.2(ii)-(iv) and A.3(i)(ii) together with Lemma 2.4 in Newey and McFadden (1994) then allow us to conclude

$$\sup_{(b,a) \in \mathcal{B} \times \mathcal{N}_\alpha} |\tilde{Q}_N(b, a) - \tilde{Q}(b, a)| = o_P(1) \quad (156)$$

and that $\tilde{Q}(b, a)$ is continuous on $\mathcal{B} \times \mathcal{N}_\alpha$ for some neighborhood $\mathcal{N}_\alpha$ of α. Moreover, by Assumption A.1(i), $\hat{\alpha} \in \mathcal{N}_\alpha$ with probability approaching one, and therefore

$$\sup_{b \in \mathcal{B}} |Q_N(b) - Q(b)| \leq \sup_{b \in \mathcal{B}} |\tilde{Q}_N(b, \hat{\alpha}) - \tilde{Q}(b, \hat{\alpha})| + \sup_{b \in \mathcal{B}} |\tilde{Q}(b, \hat{\alpha}) - Q(b, \alpha)| = o_P(1), \quad (157)$$

where the final result follows from (156), $\tilde{Q}$ being continuous on $\mathcal{B} \times \mathcal{N}_\alpha$, and continuous functions being uniformly continuous on compact sets. The consistency of $\hat{\beta}$ then follows from result (157), $\tilde{Q}$ being continuous on $\mathcal{B} \times \mathcal{N}_\alpha$ implying Q is continuous on $\mathcal{B}$, Assumptions A.2(i)(ii), and Theorem 2.1 in Newey and McFadden (1994). ■

Lemma A.7. *If Assumption A.4(ii)-(iv) and A.5(i) are satisfied, then it follows that the class of functions $\mathcal{F} \equiv \{m_\beta(W, a, b) : (a, b) \in \mathcal{N}_\alpha \times \mathcal{N}_\beta\}$ is Donsker.*

Proof. By the mean value theorem (see, e.g., Proposition 7.3.2 in Luenberger (1969)), it follows that for any $a \in \mathcal{N}_\alpha$ and $b_1, b_2 \in \mathcal{N}_\beta$ we have the bound

$$\|m_\beta(W, b_1, a) - m_\beta(W, b_2, a)\| \leq \sup_{(b,a) \in \mathcal{N}_\beta \times \mathcal{N}_\alpha} \|\nabla_b m_\beta(W, b, a)\|_o \times \|b_1 - b_2\|. \quad (158)$$

Since $E[\sup_{(b,a) \in \mathcal{N}_\beta \times \mathcal{N}_\alpha} \|\nabla_b m_\beta(W, b, a)\|_o] < \infty$ by Assumption A.4(iv), result (158) together with Assumption A.5(i) imply there is a $B(W)$ satisfying $E[B^2(W)] < \infty$ and

$$\|m_\beta(W, b_1, a_1) - m_\beta(W, b_2, a_2)\| \leq B(W)\{\|a_1 - a_2\| + \|b_1 - b_2\|\}$$

for all $b_1, b_2 \in \mathcal{N}_\beta$ and $a_1, a_2 \in \mathcal{N}_\alpha$. The lemma then follows from Theorems 2.5.6 and 2.7.11 in van der Vaart and Wellner (1996). ■

References

Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231–263.

Abadie, A., J. Angrist, and G. Imbens (2002): "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings," *Econometrica*, 70, 91–117.

Abdulkadiroglu, A., J. D. Angrist, Y. Narita, and P. A. Pathak (2017): "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 85, 1373–1432.

Andresen, M. E. and M. Huber (2021): "Instrument-based Estimation with Binarised Treatments: Issues and Tests for the Exclusion Restriction," *The Econometrics Journal*, 24, 536–558.

Andrews, D. W. (2002): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119–162.

Angrist, I. D. and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," in *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, vol. 51, 401–434.

Angrist, J., D. Autor, and A. Pallais (2022): "Marginal Effects of Merit Aid for Low-income Students," *The Quarterly Journal of Economics*, 137, 1039–1090.

Angrist, J., B. Ferman, C. Gao, P. Hull, O. L. Tecchio, and R. W. Yeh (2024): "Instrumental Variables with Time-Varying Exposure: New Estimates of Revascularization Effects on Quality of Life," *National Bureau of Economic Research Working Paper Series*, 33296.

Angrist, J. and V. Lavy (2009): "The Effects of High Stakes High School Achievement Awards: Evidence from a Randomized Trial," *American Economic Review*, 99, 1384–1414.

Angrist, J., V. Lavy, and A. Schlosser (2010): "Multiple Experiments for the Causal Link between the Quantity and Quality of Children," *Journal of Labor Economics*, 28, 773–824.

Angrist, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *The American Economic Review*, 80, 313–336.

Angrist, J. D. and S. H. Chen (2011): "Schooling and the Vietnam-Era GI Bill: Evidence from the Draft Lottery," *American Economic Journal: Applied Economics*, 3, 96–118.

Angrist, J. D. and W. N. Evans (1998): "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size," *The American Economic Review*, 88, 450–477.

Angrist, J. D. and P. Hull (2023): "Instrumental Variables Methods Reconcile Intention-to-Screen Effects Across Pragmatic Cancer Screening Trials," *Proceedings of the National Academy of Sciences*, 120, e2311556120, https://doi.org/10.1073/pnas.2311556120.

Angrist, J. D. and G. W. Imbens (1995): "Two-stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–455.

Athey, S., R. Chetty, G. W. Imbens, and H. Kang (2025): "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely," *Review of Economic Studies*, rdaf087.

Behagel, L., B. Crepon, and M. Gurgand (2013): "Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial," Institute for the Study of Labor Discussion Paper No. 7447.

Blandhol, C., J. Bonney, M. Mogstad, and A. Torgovitsky (2022): "When is TSLS Actually LATE?" NBER Working Paper No. 29709.

Caetano, C., G. Caetano, and J. Carlos Escanciano (2023): "Regression Discontinuity Design with Multivalued Treatments," *Journal of Applied Econometrics*, 38, 840–856.

Chabrier, J., S. Cohodes, and P. Oreopoulos (2016): "What Can We Learn from Charter School Lotteries?" *Journal of Economic Perspectives*, 30, 57–84.

Chernozhukov, V., I. Fernández-Val, S. Han, and K. Wüthrich (2024): "Estimating Causal Effects of Discrete and Continuous Treatments with Binary Instruments," *arXiv preprint arXiv:2403.05850*.

de Chaisemartin, C. (2017): "Tolerating Defiance? Local Average Treatment Effects Without Monotonicity," *Quantitative Economics*, 8, 367–396.

D'Haultfœuille, X. and P. Février (2015): "Identification of Nonseparable Triangular Models with Discrete Instruments," *Econometrica*, 83, 1199–1210.

Dutz, D., I. Huitfeldt, S. Lacouture, M. Mogstad, A. Torgovitsky, and W. van Dijk (2021): "Selection in Surveys: Using Randomized Incentives to Detect and Account for Nonresponse Bias," Tech. rep., National Bureau of Economic Research.

Fang, Z. and A. Santos (2019): "Inference on Directionally Differentiable Functions," *The Review of Economic Studies*, 86, 377–412.

Feller, A., T. Grindal, L. Miratrix, and L. C. Page (2016): "Compared to What? Variation in the Impacts of Early Childhood Education by Alternative Care Type," *The Annals of Applied Statistics*, 10, 1245–1285.

Finkelstein, A., S. Taubman, B. Wright, M. Bernstein, J. Gruber, J. P. Newhouse, H. Allen, K. Baicker, and O. H. S. Group (2012): "The Oregon Health Insurance Experiment: Evidence from the First Year," *The Quarterly Journal of Economics*, 127, 1057–1106.

Frangakis, C. E. and D. B. Rubin (2002): "Principal Stratification in Causal Inference," *Biometrics*, 58, 21–29.

Frölich, M. (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics*, 139, 35–75.

Frölich, M. and M. Huber (2017): "Direct and Indirect Treatment Effects–Causal Chains and Mediation Analysis with Instrumental Variables," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 1645–1666.

Goff, L. (2024): "When Does IV Identification Not Restrict Outcomes?" *arXiv preprint arXiv:2406.02835*.

Gonçalves, S., U. Hounyo, A. J. Patton, and K. Sheppard (2023): "Bootstrapping Two-stage Quasi-maximum Likelihood Estimators of Time Series Models," *Journal of Business & Economic Statistics*, 41, 683–694.

Gray-Lobe, G., P. Pathak, and C. R. Walters (2023): "The Long-Term Effects of Universal Preschool in Boston," *The Quarterly Journal of Economics*, 138, 363–411.

Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12, 187–197.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica: Journal of the Econometric Society*, 891–916.

Heckman, J. J. and R. Pinto (2018): "Unordered Monotonicity," *Econometrica*, 86, 1–35.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006): "Understanding Instrumental Variables in Models with Essential Heterogeneity," *The Review of Economics and Statistics*, 88, 389–432.

Hull, P. (2018): "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons," Available at SSRN 2705108.

Imai, K., L. Keele, and T. Yamamoto (2010): "Identification, Inference and Sensitivity Analysis for Causal Mediation Effects," *Statistical Science*, 25, 51 – 71.

Imai, K. and M. Ratkovic (2014): "Covariate Balancing Propensity Score," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76, 243–263.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Kirkeboen, L. J., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *The quarterly journal of economics*, 131, 1057–1111.

Kline, P. and C. R. Walters (2016): "Evaluating Public Programs with Close Substitutes: The Case of Head Start," *The Quarterly Journal of Economics*, 131, 1795–1848.

Kwon, S. and J. Roth (2025): "Testing Mechanisms," *arXiv:2404.11739.*

Lee, S. and B. Salanié (2018): "Identifying Effects of Multivalued Treatments," *Econometrica*, 86, 1939–1963.

Luenberger, D. G. (1969): *Optimization by Vector Space Methods*, New York: Wiley.

Mogstad, M. and A. Torgovitsky (2024): "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects," in *Handbook of Labor Economics*, ed. by C. Dustmann and T. Lemieux, Elsevier, vol. 5, chap. 1, 1–114.

Mountjoy, J. (2022): "Community Colleges and Upward Mobility," *American Economic Review*, 112, 2580–2630.

Navjeevan, M., R. Pinto, and A. Santos (2023): "Identification and Estimation in a Class of Potential Outcomes Models," *arXiv 2310.05311.*

Newey, W. K. and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle and D. L. McFadden, Elsevier Science B.V., vol. IV, 2113–2245.

Nibbering, D. and M. Oosterveen (Forthcoming): "Instrument-Based Estimation of Full Treatment Effects with Partial Compliers," *The Review of Economics and Statistics.*

Pinto, R. (2021): "Beyond Intention to Treat: Using the Incentives in Moving to Opportunity to Identify Neighborhood Effects," *NBER Working Paper.*

Rose, E. K. and Y. Shem-Tov (2024): "On Recoding Ordered Treatments as Binary Indicators," *The Review of Economics and Statistics*, 1–32.

Słoczyński, T., S. D. Uysal, and J. M. Wooldridge (2025): "Abadie's Kappa and Weighting Estimators of the Local Average Treatment Effect," *Journal of Business & Economic Statistics*, 43, 164–177.

Tan, Z. (2006): "Regression and Weighting Methods for Causal Inference Using Instrumental Variables," *Journal of the American Statistical Association*, 101, 1607–1618.

Torgovitsky, A. (2015): "Identification of Nonseparable Models Using Instruments With Small Support," *Econometrica*, 83, 1185–1197.

United States Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation (2018): "Head Start Impact Study (HSIS), 2002-2006 [United States]," ICPSR29462-v7, http://doi.org/10.3886/ICPSR29462.v7.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: with Applications to Statistics*, New York: Springer.