IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons*

Peter Hull†

February 2018

Abstract

Instrumental variable (IV) estimates may be difficult to interpret when there are multiple alternatives to treatment. I consider identification of counterfactual-specific local average treatment effects via interactions of the instrument and stratifying controls. Deriving general IV estimands, I establish identification under mean-independence of complier effects vis-a-vis the strata. I then extend IV to a two-step procedure valid under conditional independence. The theory is illustrated in the context of GED certification, where individuals may otherwise graduate high school or drop out. I also show how it can be used to address differential attrition with an application to the Oregon Medicaid experiment

*I thank Josh Angrist, Kirill Borusyak, Mayara Felix, Amy Finkelstein, Jim Heckman, Sally Hudson, Jack Liebersohn, Jack Mountjoy, Kathleen Mullen, Parag Pathak, Miikka Rokkanen, Bernard Salanié, Chris Taber, Chris Walters, and seminar participants from the the Bank of Canada, MIT, and the University of Calgary and for valuable feedback. I gratefully acknowledge financial support from the National Institute on Aging (grant #T32-AG000186)

†The University of Chicago and Microsoft Research. Email: `hull@uchicago.edu`; website: `http://peterhull.net`

1 Introduction

Treatment effect heterogeneity is often a central consideration in economics. In instrumental variable (IV) designs, Imbens and Angrist (1994) famously show that regressions on a binary treatment variable identify average effects for compliers – those induced into treatment – given a binary instrument that is as-good-as-randomly assigned, monotonically affects treatment receipt, and has no direct effect on outcomes. Naturally, these local average treatment effects (LATEs) are also estimable within strata unaffected by the instrument. In a randomized trial with treatment assigned by lottery, for example, conditional IV regressions can reveal heterogeneity in average effects across compliers with different pre-randomization characteristics.

Often, however, it is not possible to stratify on key dimensions of heterogeneity. Suppose a researcher finds a quasi-experimental decline in General Educational Development (GED) testing standards that causes some individuals to obtain a GED. These compliers may be drawn from distinct fallbacks: while some would have dropped out of high school if not for the change, others may have earned a traditional diploma. Under the Imbens and Angrist (1994) assumptions, an IV wage regression with a single GED treatment estimates a causally-interpretable LATE across these two groups, but this parameter may be difficult to interpret economically. With typically-positive returns to schooling, the LATE will mix together positive and negative effects and represent an ambiguously-signed average. Separating these effects is challenging when one cannot directly observe and stratify on fallback schooling levels.

In some settings it may be difficult to even causally interpret effects with multiple fallbacks. Consider the issue of differential attrition in a randomized trial: outcomes are voluntarily provided via survey, and the instrument makes some individuals less likely to respond. Here neither a full-population IV regression – which includes survey nonresponders – nor one restricted to the subgroup of endogenous responders is likely to be causal. Again, while we wish to isolate effects for a subset of compliers (those with always-measured outcomes), we are unlikely to identify them directly.

This paper explores ways in which covariate stratifications, while not able to perfectly separate complier groups, may nevertheless disentangle counterfactual-specific treatment effects. The basic strategy is intuitive: if there exists a stratification across which the proportion of compliers with different fallbacks varies but, on average, their causal effects do not, differences in stratum-specific reduced-form effects may be attributed to differences in complier shares in such a way that identifies counterfactual-specific LATEs. I show that this deconvolution is automated by an IV regression with multiple endogenous variables and strata interaction instruments. When complier effect homogeneity is too restrictive, moreover, I show that IV informs bounds on particular weighted averages of counterfactual-specific effects and can be generalized to allow for heterogeneity through a rich set of observables. In this case I propose a flexible two-step estimator for the multiple LATEs.

Interacting an instrument with covariates in models with multiple endogenous variables has

a long history in econometrics (Wooldridge, 2002; Reardon and Raudenbush, 2013). This paper establishes minimal restrictions on treatment effect heterogeneity under which such coefficients are causal. In similar settings, Behaghel et al. (2013), Blackwell (2017), Lee and Salanié (2016), and Mountjoy (2017) consider nonparametric identification of multiple treatment channels and multiple treatment-specific instruments, while Kirkebøen et al. (2016) show how IV identifies counterfactual-specific LATEs when preferred treatment alternatives are directly measured.[1] This paper offers a different approach for when only a single binary instrument is available and treatment fallbacks are unobserved. Broadly, I use a principal stratification framework (Frangakis and Rubin, 2002) to extend the LATE framework to settings with multiple causal channels. Using observables to nonparametrically extend the IV result also relates to the extrapolation of LATEs across quasi-experiments and within regression discontinuity designs in Angrist and Fernandez-Val (2013) and Angrist and Rokkanen (2016).

I develop this econometric theory with the motivating GED example, and illustrate identification in a GED selection model inspired by Heckman and Urzúa (2010). Monte Carlo simulations show that the finite-sample bias of the counterfactual-specific LATE estimators becomes negligible with moderately-sized first-stage F-statistics, though they are widely distributed. This highlights the relatively narrow identifying variation of cross-strata comparisons and, correspondingly, likely less-precise estimation.

Lastly, I revisit Finkelstein et al. (2012)'s analysis of the 2008 Oregon Medicaid Experiment to show how the theory applies to nonrandom sample attrition. Rather than restricting analyses to individuals with endogenously provided outcomes, I show how a baseline stratification can isolate effects for individuals that always yield outcomes. This leverages the common surveying practice of limited intensive followup, where strata defined by random second-round surveying are plausibly uncorrelated with the distribution of complier treatment effects.[2] Despite significant differential attrition, the Finkelstein et al. (2012) results appear robust.

2 Theory

For each individual in a population we observe a binary instrument Z, an outcome Y, a binary covariate X, and a variable T equaling 1, a, or b. Here $T = 1$ indicates a treated individual, while someone with $T = a$ or $T = b$ is in one of two untreated states.[3] Indicators for being in each fallback are given by A and B; treatment is then $D = 1 - A - B$. For concreteness, we may imagine Z indicating exposure to a quasi-experimental reduction in GED passing standards, Y denoting adult earnings, and $D = 1$ if an individual is GED-certified. Other individuals may either be high school dropouts ($A = 1$) or graduates ($B = 1$).

[1] See Kline and Walters (2016) and Hull (2018) for recent semiparametric IV analyses of multiple treatment effects.

[2] DiNardo et al. (2006) and Behaghel et al. (2015) explore alternative parametric and partial identification methods leveraging randomized intensive followup for intent-to-treat estimation.

[3] It is straightforward to generalize to n untreated states and n-valued X. I develop the $n = 2$ case for simplicity.

Causal effects are defined in terms of potential outcomes. Treatment and fallback states given $Z = z$ are written D_z, A_z, and B_z, while Y_{zt} denotes realizations of Y when $Z = z$ and $T = t$. These latent variables are independent across individuals, satisfying a stable unit treatment value assumption (Rubin, 1974). To this we add:

A1 *Independence*: $((Y_{z1}, Y_{za}, Y_{zb}, A_z, B_z)_{z\in\{0,1\}}) \perp\!\!\!\perp Z|X$

A2 *Exclusion*: $\forall t \in \{1, a, b\}, Pr(Y_{0t} = Y_{1t}) = 1$

A3 *Monotonicity*: $Pr(A_1 \leq A_0) = Pr(B_1 \leq B_0) = 1$.

In the GED example, A1 states that variation in passing standards Z is as-good-as-randomly assigned with respect to potential outcomes within strata given by X. This identifies conditional effects of Z on Y, A, and B. Interpretation of the earnings effect by way of schooling further requires an exclusion restriction (A2), which defines the single-indexed potential outcomes $Y_t = Y_{zt}$. Finally, we assume the instrument's effect on treatment is monotone, in that Z induces nobody to either fallback. In the GED example, monotonicity implies no student is caused to drop out of or complete high school when a GED becomes easier to obtain.

Under A3 we may categorize individuals as one of four types: never-takers, a-compliers, b-compliers, and always-takers. These groups are defined in Figure 1. In the GED example, never-takers are those who either always drop out of high school or always obtain a traditional diploma, while always-takers are students that obtain a GED even when it is difficult to pass. Compliers are individuals induced to a GED when it becomes easier, but either drop out (a-compliers) or graduate high school (b-compliers) when it is harder. A-compliers are defined by $A_1 < A_0$ while b-compliers have $B_1 < B_0$.

Although stated with expanded notation, it is straightforward to verify that A1-A3 coincide with the Imbens and Angrist (1994) assumptions on $(Y, D, Z)|X$. The value of the extended setup is that we can now write the conditional IV estimand as an average of two counterfactual-specific LATEs for each of the a-complier and b-complier subpopulations. Specifically, we have the following:

Lemma 1 Suppose $Pr(D_1 > D_0|X) \neq 0$. Then under A1-A3 a conditional-on-X regression of Y on D instrumented by Z identifies

$$E[Y_1 - Y_a|A_1 < A_0, X]\omega(X) + E[Y_1 - Y_b|B_1 < B_0, X](1 - \omega(X)), \tag{1}$$

where

$$\omega(X) = \frac{Pr(A_1 < A_0|X)}{Pr(A_1 < A_0|X) + Pr(B_1 < B_0|X)}. \tag{2}$$

The proof of Lemma 1, derived in the appendix along with all other proofs, uses the aforementioned equivalence of A1-A3 to show the conditional IV estimand identifies a LATE. With two treatment fallbacks, this may in turn be written as a weighted average of effects across two complier groups, with weights proportional to their population shares. An IV regression on the GED treatment thus

weights together effects for students with a dropout counterfactual with effects for students who would have otherwise graduated. Here we are interested in extracting these counterfactual-specific LATEs.

Note that while this paper focuses on multiple fallbacks, we could also write A3 with the inequalities reversed and proceed with A and B corresponding to distinct treatments. This version of the theory is closely related to the literature on multiple treatment mediators (Reardon and Raudenbush, 2013) as well as the framework of Behaghel et al. (2013), who consider two treatment states and a tri-valued instrument $\tilde{Z} \in \{1, a, b\}$. With $Z = \mathbf{1}[\tilde{Z} = 1]$ and the remaining heterogeneity in $\tilde{Z}$ attributed to individuals, A1-A3 are implied by this framework. Under similar assumptions, Kirkebøen et al. (2016) discuss identification of fallback-specific LATEs when the econometrician is able to measure A_0 and B_0 directly. The following can thus be thought of as an alternative for when only one valid instrument is available and potential fallbacks are not perfectly observed. Instead, we leverage a stratification generating variation in average potential fallbacks.[4]

2.1 IV Identification

Intuition for the first identification result comes from equations (1) and (2). Instrument independence identifies the conditional first-stage effects of Z on $1 - A$ and $1 - B$, while monotonicity implies these equal $Pr(A_1 < A_0|X)$ and $Pr(B_1 < B_0|X)$. Thus the weighting scheme in (2) is identified under A1-A3. If moreover the two counterfactual-specific LATEs are the same across strata, (1) and (2) constitute a system of two equations (one for each stratum) and two unknown LATEs, and identification is achieved as long as $\omega(0)$ and $\omega(1)$ are not the same.

An IV regression with two endogenous variables and strata-interacted instruments implements this approach. Consider the second stage of

$$Y = \mu + \alpha(1 - A) + \beta(1 - B) + \gamma X + \epsilon, \tag{3}$$

with $1 - A$ and $1 - B$ instrumented by Z and $Z \times X$ controlling for X. Let

$$\alpha(x) = E[Y_1 - Y_a|A_1 < A_0, X = x] \tag{4}$$

$$\beta(x) = E[Y_1 - Y_b|B_1 < B_0, X = x] \tag{5}$$

denote the stratum- and counterfactual-specific LATEs and

$$\pi_a(x) = Pr(A_1 < A_0|X = x) \tag{6}$$

$$\pi_b(x) = Pr(B_1 < B_0|X = x). \tag{7}$$

be the corresponding shares of a- and b-compliers. We then have the following:

[4] Although this approach uses an IV regression with two instruments and two endogenous variables, as in Behaghel et al. (2013), the strata-interacted instruments here will generally not satisfy their assumptions. The approaches trivially coincide with eachother and Kirkebøen et al. (2016) when a-compliers and b-compliers are fully stratified.

Proposition 1 Suppose the matrix of complier shares

$$\Pi = \begin{bmatrix} \pi_a(0) & \pi_b(0) \\ \pi_a(1) & \pi_b(1) \end{bmatrix} \tag{8}$$

is nonsingular. Then under A1-A3 the coefficients in equation (3) are

$$\alpha = \omega\alpha(0) + (1-\omega)\alpha(1) + \delta_a(\beta(0) - \beta(1)) \tag{9}$$

$$\beta = (1-\omega)\beta(0) + \omega\beta(1) + \delta_b(\alpha(0) - \alpha(1)), \tag{10}$$

where

$$\omega = \left(1 - \frac{\pi_a(1)}{\pi_b(1)} \Big/ \frac{\pi_a(0)}{\pi_b(0)}\right)^{-1} \tag{11}$$

$$\delta_a = \left(\frac{\pi_a(0)}{\pi_b(0)} - \frac{\pi_a(1)}{\pi_b(1)}\right)^{-1} \tag{12}$$

$$\delta_b = \left(\frac{\pi_b(0)}{\pi_a(0)} - \frac{\pi_b(1)}{\pi_a(1)}\right)^{-1}. \tag{13}$$

Proposition 1 extends the Imbens and Angrist (1994) interpretation of single-treatment IV to regressions with multiple endogenous variables and an interacted instrument. The interaction must lead to variation in the complier composition, so that the first stage matrix Π is invertible. Although similarly derived, equations (9)-(13) are not as easily interpreted as in single-treatment IV, however. The coefficient on $1 - A$ equals a weighted average of a-complier effects across the strata plus a "bias" term scaled by the difference in b-complier effects. Since the scaling factor is identified by the Π, IV may inform bounds on $\omega\alpha(0) + (1-\omega)\alpha(1)$, with narrower intervals given by stratifications where $\pi_a(X)/\pi_b(X)$ is more heterogeneous. Note, however, that since these ratios are always positive under A3, $\omega \in (-\infty, 0) \cup (1, \infty)$ and the partially-identified average is nonconvex.

With $\delta_a, \delta_b \neq 0$, Proposition 1 shows that two-treatment IV models point-identify counterfactual-specific LATEs if and only if they are mean-independent of the stratification. That is, consider

A4 *LATE homogeneity*: $E[Y_1 - Y_a | A_1 < A_0, X]$ and $E[Y_1 - Y_b | B_1 < B_0, X]$ do not depend on X .

The form of equations (9)-(10) then makes the following immediate:

Corollary to Proposition 1 : If Π is invertible and A1-A4 hold, the coefficients in (3) are $\alpha = E[Y_1 - Y_a | A_1 < A_0]$ and $\beta = E[Y_1 - Y_b | B_1 < B_0]$.

Under LATE homogeneity, therefore, the multiple-treatment IV regression recovers fallback-specific LATEs.[5] In the stylized example, α and β are the average returns to GED certification for

[5]One may extend Proposition 1 to multivalued X and an over-identified IV regression with many stratum interactions. When A4 holds across all values of X, counterfactual-specific LATEs are identified by any two and a standard overidentification test is valid for A1-A4.

individuals who would have otherwise dropped out and completed high school, respectively, provided these do not vary systematically by X.[6]

What kinds of economic models generate A1-A4? Consider an individual deciding between the treatment and fallbacks to maximize state-specific latent utility ν_t:

$$D = \mathbf{1}[\nu_1 \geq \nu_a, \nu_1 \geq \nu_b] \tag{14}$$

$$A = \mathbf{1}[\nu_a \geq \nu_b, \nu_a \geq \nu_1] \tag{15}$$

$$B = \mathbf{1}[\nu_b \geq \nu_a, \nu_b \geq \nu_1]. \tag{16}$$

For A3 it is sufficient to have $\nu_1 = h(X, Z, \eta)$ with $h(x, 1, \eta) \geq h(x, 0, \eta)$ almost-surely. A2 and A4 hold if potential outcomes can be written

$$Y_t = f(X) + g(t) + \epsilon_t, \tag{17}$$

where

$$E[\epsilon_1 - \epsilon_a | A_1 < A_0, X] = E[\epsilon_1 - \epsilon_a | A_1 < A_0] \tag{18}$$

$$E[\epsilon_1 - \epsilon_b | B_1 < B_0, X] = E[\epsilon_1 - \epsilon_b | B_1 < B_0], \tag{19}$$

while A1 holds if $(\eta, \nu_a, \nu_b, \epsilon_1, \epsilon_a, \epsilon_b)'$ is independent of Z given X. Note that (17)-(19) neither requires X to be excludable nor independent from potential outcomes (in which case it may be thought of as an instrument); rather, we allow X to additively enter the outcome equation, with mean-independent differences in residual outcome determinants within complier subpopulations. Loosely, though there may be more a-compliers in the $X = 1$ stratum, they are not systematically different from others when A4 holds. In section 3 and the appendix I build a simple parametric model satisfying LATE homogeneity, in which X uniformly shifts the utility of one of the fallbacks.

2.2 Relaxing Independence and Homogeneity

Multi-treatment IV requires cross-stratum comparisons to inform a common pair of average treatment effects, which in some applications may be unduly restrictive. As in other settings, however, the IV approach may be extended to allow A4 to hold only conditional on a set of predetermined covariates (Hirano et al., 2003; Abadie, 2003). The general approach is again intuitive: one could imagine running two-treatment IV regressions at each point in the support of a discrete control W. When cross-stratum differences in complier LATEs are fully captured by W, averaging the resulting coefficients over the marginal complier distribution will recover population LATEs. This procedure is conceptually possible yet likely infeasible when W is continuous or high-dimensional. I next outline an alternative, flexible implementation.

[6] Proposition 1 also gives a way to indirectly validate A4 given a control C likely correlated with treatment effects. Setting $Y = C \times A$ in equation (3) yields a coefficient on $1 - B$ of $\delta_b(E[C|A_1 < A_0, X = 1] - E[C|A_1 < A_0, X = 0])$ under A1-A3, since $\beta(0) = \beta(1) = 0$. One could therefore test whether C systematically varies for a-compliers, and likewise for b-compliers.

I start with extensions of A1 and A4:

A1′ $((Y_{z1}, Y_{za}, Y_{zb}, A_z, B_z)_{z=0,1}) \perp\!\!\!\perp Z|W, X$

A4′ $E[Y_1 - Y_a|A_1 < A_0, W, X]$ and $E[Y_1 - Y_b|B_1 < B_0, W, X]$ do not depend on X

Here A1′ only requires the instrument to be as-good-as-randomly assigned once potential confounders in W and X are fixed, while A4′ allows for arbitrary cross-stratum heterogeneity in complier treatment effects through W. We then have the following:

Proposition 2 Suppose $Pr(Z = 1|W, X) \in (0, 1)$, $Pr(X = 1|W) \in (0, 1)$ and

$$\Pi(W) = \begin{bmatrix} Pr(A_1 < A_0|W, X = 0) & Pr(B_1 < B_0|W, X = 0) \\ Pr(A_1 < A_0|W, X = 1) & Pr(B_1 < B_0|W, X = 1) \end{bmatrix} \tag{20}$$

is nonsingular with probability one. Define

$$\lambda = \frac{E[Z|W, X] - Z}{E[Z|W, X](1 - E[Z|W, X])} \tag{21}$$

$$\mu_a = \frac{E[\lambda AX|W] - E[\lambda A|W]X}{E[X|W](1 - E[X|W])} \tag{22}$$

$$\mu_b = \frac{E[\lambda BX|W] - E[\lambda B|W]X}{E[X|W](1 - E[X|W])}. \tag{23}$$

Then, under A1′, A2, A3, and A4′,

$$E[Y_1 - Y_a|A_1 < A_0] = E\left[\left(\frac{E[\lambda A|W]}{E[\lambda A]}\frac{\lambda\mu_b}{E[\lambda\mu_b A|W]}\right)Y\right] \tag{24}$$

$$E[Y_1 - Y_b|B_1 < B_0] = E\left[\left(\frac{E[\lambda B|W]}{E[\lambda B]}\frac{\lambda\mu_a}{E[\lambda\mu_a B|W]}\right)Y\right]. \tag{25}$$

These weights are moreover identified by $E[X|W]$, $E[Z|W, X]$, $E[A|W, X, Z]$, and $E[B|W, X, Z]$.

The proof of Proposition 2 shows that conditional-on-W IV coefficients on $1 - A$ can be written as the ratio of $E[\lambda\mu_b Y|W]$ to $E[\lambda\mu_b A|W]$. Averaging these over the marginal distribution of W for a-compliers (with $E[\lambda A|W]/E[\lambda A]$ weights) thus identifies $E[Y_1 - Y_a|A_1 < A_0]$. Similar logic proves (25). This requires the conditional IV estimand to be well-defined throughout the support of W: $Z|W, X$ and $X|W$ must be nondegenerate and the conditional first-stage matrix must be invertible. Thus, while W must be rich enough to make Z ignorable and the stratum-specific LATEs homogeneous, it must still allow for cross-stratum variation in complier shares. Proposition 2 then suggests a two-step estimation procedure. In a first step, we flexibly estimate four conditional expectations: $E[X|W]$, $E[Z|W, X]$, $E[A|W, X, Z]$, and $E[B|W, X, Z]$. The appendix shows how these can then be used to form estimates of the weights in equations (24) and (25).[7]

[7]Inference for this two-step estimator may under appropriate regularity conditions be based either on a bootstrap procedure or on analytic expressions following Andrews (1991) and Newey (1994a, 1994b).

Finally, note that adding covariates to the treatment utility and potential outcome equations of the earlier selection model extends it along the lines of A1′, A2, A3, and A4′:

$$\nu_1 = h(W, X, Z, \eta) \tag{26}$$

$$Y_t = f(W, X) + g(W, t) + \epsilon_t, \tag{27}$$

where differences in ϵ_t are mean-independent of X given W among compliers, the structural error vector is independent of Z given (X, W), and $h(W, x, z, \eta)$ is almost-surely monotone in z given W. Complier treatment effects may now nonparametrically vary across X through the $g(W, t)$.

3 Empirical evaluation

3.1 GED Monte Carlo

I illustrate identification and estimation of counterfactual-specific LATEs with a model of GED selection and adult earnings. Heckman and Urzúa (2010) use such a model to demonstrate identification of GED effects under structural assumptions; I modify their specification to include a stratification consistent with A1-A4. Here again Y denotes an individual's log hourly earnings, D indicates GED certification, and A and B indicate the two fallbacks of dropping out and completing high school, respectively. As in Heckman et al. (2012), the instrument Z denotes a quasi-experimental reduction in GED passing standards which induces some to take the GED. These authors find that compliers who were older at the time of such a 1997 reform were more likely to be drawn from a dropout counterfactual, perhaps due to differential constraints from minimum schooling regulations. I correspondingly generate a stratification X representing age at the time of reform. While age may directly affect earnings, I assume X does not affect the marginal returns to schooling while uniformly shifting dropout utility, satisfying LATE homogeneity. An appendix section fully specifies the model.

Population first-stage and reduced-form effects from the modeled quasi-experiment are reported in Table 1. Panel A shows that a decrease in GED passing standards increases the share of GED-certified students by 5.5 percentage points and decreases hourly earnings by 1.2 percent. Under monotonicity, the former represents the total share of compliers in the population and the ratio of reduced-form to first-stage effects is the overall LATE (-0.217).

Stratum-specific quasi-experimental moments are reported in Panel B. Compliers in the dropout-constrained ($X = 0$) subsample are more likely to obtain a high school diploma when $Z = 0$, while older compliers are more likely to drop out. The model parameterizes wages such that students tend to see gains when shifted to the GED from the dropout fallback and losses when it replaces a traditional diploma. Consequently, the reduced form effect of an easier GED exam is higher when $X = 1$ (0.009) than when $X = 0$ (-0.019). The population two-treatment IV coefficients, $\alpha = 0.277$ and $\beta = -0.396$, are obtained by multiplying the inverse of the first stage matrix by the reduced form vector. By Proposition 1 these equal counterfactual-specific LATEs.

I next explore the finite-sample performance of IV estimates. Figure 2 summarizes 5,000 Monte Carlo simulations from the model, over a range of parameters governing the strength of the IV first stage. The black curve in panel A shows that the conventional LATE estimator becomes median unbiased at relatively low levels of the multivariate first stage F-statistic (recall as these models are just-identified, the estimators have no finite-sample moments). The counterfactual-specific LATE estimates, in contrast, retain modest bias until $F \approx 5$, with virtually all bias vanishing by $F = 10$. Interestingly, bias in this model is nonmonotone for the dropout-counterfactual LATE estimator, initially increasing in the range of 0.2-0.4. This increase is offset by a decline in the bias of the other LATE estimator.

Panel B of Figure 2 similarly plots the interquartile range of Monte Carlo estimates. These highlight the fact that while the two-treatment IV estimates are likely approximately unbiased they are also likely to be highly variable, with an interquartile range around 5-10 times wider than the corresponding single-treatment LATE estimate when $F = 5$. As counterfactual-specific LATEs are identified by narrower variation in cross-strata first stage differences, they are prone to less precision.

3.2 Differential Attrition

Identifying counterfactual-specific treatment effects is essential in a randomized trial with imperfect followup. Suppose program offers Z are randomly assigned to individuals, who may choose both whether or not to comply with the treatment and whether or not to participate in a subsequent survey of an outcomes Y. Individuals can then be said to select between three possible states: being treated and reporting outcomes (D), being untreated and reporting outcomes (A), and not reporting outcomes (B). Since outcomes are only measured in states D and A (suppose the researcher sets $Y = 0$ for anyone with $B = 1$), the estimable single-treatment LATE,

$$E[Y_1 - Y_0|D_1 > D_0] = E[Y_1 - Y_a|A_1 < A_0]\omega + E[Y_1|B_1 < B_0](1 - \omega) \tag{28}$$

$$\text{for} \quad \omega = \frac{Pr(A_1 < A_0)}{Pr(A_1 < A_0) + Pr(B_1 < B_0)}, \tag{29}$$

is not a weighted average of causal effects whenever there are any compliers drawn from B, or $Pr(B_1 < B_0) \neq 0$. Facing endogenous attrition, researchers often conduct analyses on a restricted subset of individuals with valid outcomes. Such a procedure, however, may also be difficult to interpret as conditioning on ex post outcomes will tend to introduce imbalance in distribution of the instrument when $Pr(B_1 < B_0) \neq 0$.[8]

As the form of equations (28) and (29) suggests, the differential attrition problem can be mapped to the multiple-counterfactual setting. For a given exogenous stratification X, independence of Z from potential outcomes (A1) is ensured by virtue of the randomized design, and in many settings

[8] Common solutions to differential attrition include parametric sample selection modeling (Gronau, 1974; Heckman, 1976) and nonparametric partial identification (Lee, 2009; Behaghel et al., 2009; Engberg et al., 2014). Weighting estimators have also been proposed under conditional ignorability assumptions (Frölich and Huber, 2014).

a program offer is likely to have no direct effect on latent outcomes and to not deter program participation. A2 and A3 would then be satisfied provided that Z has no direct effect on attrition behavior given treatment status and the indirect effect through treatment is monotone. These assumptions are the same as those typically used to estimate nonparametric bounds on causal parameters under differential attrition (Lee, 2009; Behaghel et al., 2009).

To solve the problem with this paper's methods requires a stratification inducing variation in response behavior while maintaining LATE homogeneity. One candidate X exploits the practice of randomized intensive followup, a common surveying technique recommended with high attrition rates (Duflo et al., 2008). Suppose upon initially measuring outcomes a researcher selects a random fraction p of nonresponders for additional surveying attempts. Denote this set and a random fraction p of initial responders by $X = 1$, with $X = 0$ for all others.[9] Since X is correlated with an individual's probability of facing additional surveying, this $X = 1$ strata likely contains a larger proportion of compliers with observable untreated outcomes ($A_1 < A_0$). Moreover since X is randomly assigned, LATEs for both types of compliers will be the same across strata to the extent the additional followup attempt draws second-round responses from individuals representative of the pool of initial nonresponders.

The key LATE homogeneity condition is however not guaranteed by randomized followup *per se*, and researchers hoping to use Propositions 1 or 2 to address differential attrition should carefully design their randomized intensive surveying scheme. As an example, suppose researchers randomly assign offers for a job-training program and initially conduct phone interviews on employment outcomes throughout the day. If treated individuals are more likely to be working, the offer will have an effect through treatment on the probability an individual will be home to answer the survey: these compliers will have $B_1 < B_0$. However, suppose the timing of interviews is as-good-as-random with respect to working hours (perhaps due to alphabetical or other quasi-random queuing), and that a random second interview round occurs in similar fashion on a later day. Then individuals with $X = 1$ will face a higher probability of being home when surveyed on either day one or day two, but those successfully interviewed on the second day will not vary systematically from those interviewed in the initial round. A4 would then hold, and Proposition 1 may be used.

I illustrate this approach by estimating the effects of Medicaid on self-reported outcomes in the first year following a lottery of roughly 90,000 low-income adults. Finkelstein et al. (2012) discuss the setting for the Oregon Health Insurance Experiment, which drew lottery winners in eight drawings from March through September 2008. Lottery winners became eligible for enrollment in a comprehensive Medicaid program, in which around 30% enrolled. Along with administrative outcomes, Finkelstein et al. (2012) conducted a mail survey in the summer of 2009. Survey outcomes suggest Medicaid increased complier healthcare utilization, decreased out-of-pocket expenditure and debt, and improved overall health. The relatively low rate of survey

[9] It will by Rao-Blackwell logic be more efficient to include all initial responders in both strata and weight them by either p or $1 - p$. I follow this approach in the application.

response (50%) and moderate 2 percentage point imbalance in the probability of response by eligibility status, however, suggest caution in interpreting these subsample IV estimates.[10]

I first use the experimental data to replicate the Finkelstein et al. (2012) analysis sample and restricted IV estimates. 30% of initial nonrespondents were selected for additional followup attempts by mail and phone; the average yield on intensive surveying was around 22%. I then let the stratum indicator $X = 1$ for those designated for intensive followup and for a proportionate random sample of initial respondents, and construct A, B, and D from survey response and treatment indicators as outlined above.[11]

IV estimates of the effect of Medicaid on a variety of health, financial, and medical care outcomes are reported in Table 3. Column 1 replicates the "restricted" IV specifications of Finkelstein et al. (2012) with a single treatment variable estimated over the subsample of individuals with measured outcomes.[12] Column 2 instead reports estimates of the coefficient on $1 - A$ in IV regressions of the form of equation (3) over the full experimental sample. Under A1-A4, these reflect Medicaid effects for compliers who would always provide survey outcomes. Interestingly, the two-treatment IV specification yields point estimates quite close to those from the restricted single-treatment model across virtually every outcome. Although, as expected, the former are somewhat less precisely-estimated, the estimate differences (reported in column 3) are tightly distributed around zero. This suggests that, despite apparent endogenous attrition, the results of Finkelstein et al. (2012) serve as reliable measures of the true causal effects.

4 Summary

This paper extends the theoretical framework of Imbens and Angrist (1994) to settings where more than one causal channel is needed to answer an economic or causal question with a single quasi-experiment. The ease by which Proposition 1 may be applied, using an estimator with familiar statistical properties, is apparent in the above applications. More involved, though still tractable estimators relax the key independence assumptions with controls. This framework thus complements more data-intensive or parametric approaches to multiple-treatment LATE estimation. Moverover, although the key assumption of LATE homogeneity may prove strong in some settings, researchers may sometimes be able to increase its plausibility by careful experimental design. This could prove especially useful in randomized trials with imperfect followup, where differential attrition concerns loom large.

[10] Finkelstein et al. (2012) address attrition by showing instrument balance in the respondent subsample and construct Lee (2009) bounds for intent-to-treat effects. They find generally robust results for healthcare use and financial strain while not able to reject no effect on self-reported health.

[11] Although equations (28)-(29) describe a Z that makes attrition less likely, in Finkelstein et al. (2012) offers increase the probability of nonresponse. In either case using $1 - A = 1 - (1 - C)R$ and $1 - B = R$ as the two endogenous variables in equation (3) (where C indicates treatment receipt and R denotes survey response) identifies the average causal effect of treatment among compliers who always respond as the coefficient on $1 - A$.

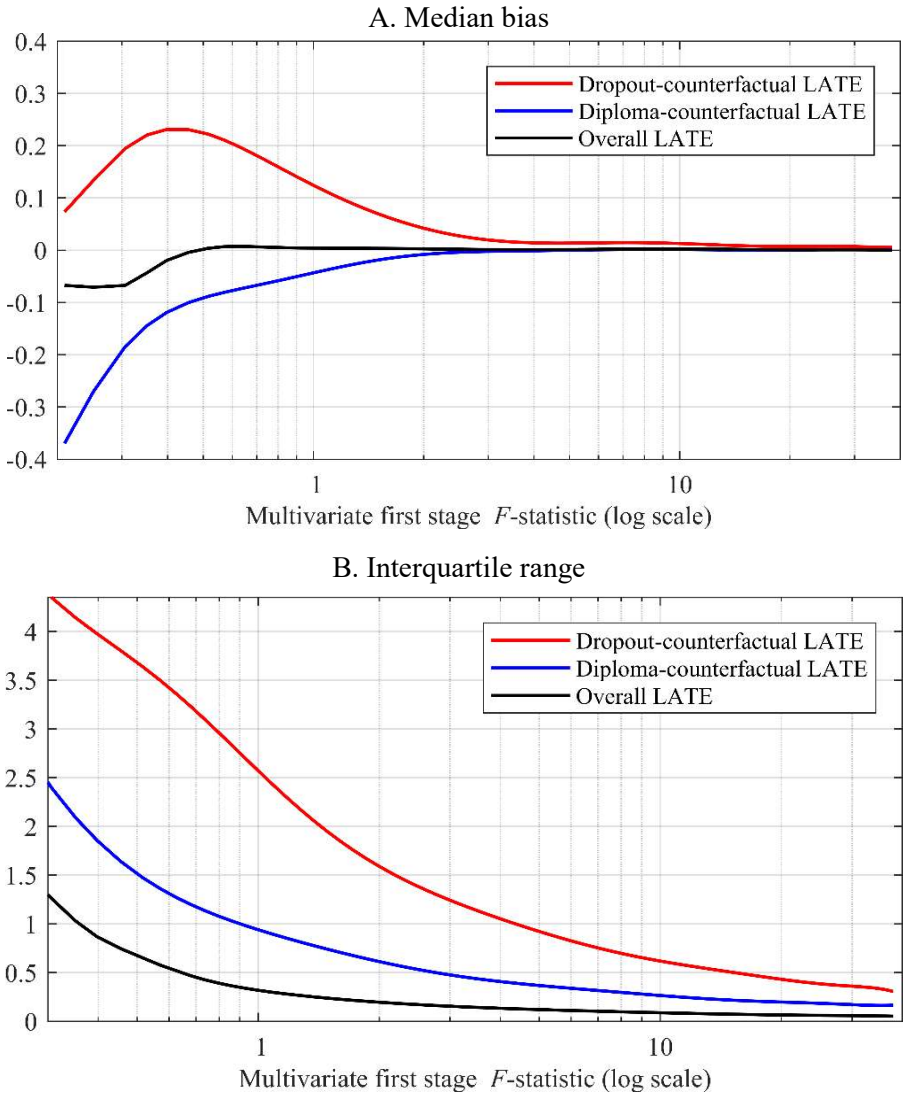
[12] I follow Finkelstein et al. (2012) in weighting all restricted IV estimates by the inverse probability of being included in the intensive followup group. This turns out to be of little empirical consequence.

Figure 1: Behavioral groups permitted by monotonicity

	$D_1 = 0$	$D_1 = 1$
$D_0 = 0$	**1. Never-takers** ($A_1 = 1$, $A_0 = 1$ $B_1 = 0$, and $B_0 = 0$, or $A_1 = 0$, $A_0 = 0$, $B_1 = 1$, and $B_0 = 1$)	**2. a-compliers** ($A_1 = 0$, $A_0 = 1$, $B_1 = 0$, and $B_0 = 0$) **3. b-compliers** ($A_1 = 0$, $A_0 = 0$, $B_1 = 0$, and $B_0 = 1$)
$D_0 = 1$		**4. Always-takers** ($A_1 = 0$, $A_0 = 0$, $B_1 = 0$, and $B_0 = 0$)

Notes: This figure lists the potential outcomes of the four behavioral groups implied by Assumption 3

Figure 2: Monte carlo simulations of GED LATE estimates



Notes: This figure summarizes the median bias (panel A) and interquartile range (panel B) of instrumental variable estimates of the effects of GED certification on different complier populations, based on 5,000 simulations of the model described in the text. Curves are smoothed by a tenth-order polynomial.

Table 1: Simulated treatment effects from a GED scoring change

	First stages			Reduced forms	LATEs
	GED certified	Not a HS dropout	Not a HS graduate	Log wages	
	(1)	(2)	(3)	(4)	(5)
		A. Single-treatment IV			
All individuals	0.055			-0.012	
All compliers					-0.217
		B. Two-treatment IV			
More dropout-constrained stratum		0.005	0.050	-0.019	
Less constrained stratum		0.073	0.029	0.009	
Dropout-counterfactual compliers					0.277
Diploma-counterfactual compliers					-0.396

Notes: This table reports population effects from the simulated model of GED certification and labor market returns described in the text. Columns 1-3 report first-stage effects of the instrument on GED certification (D), whether an individual is not a high school dropout (1-A), and whether an individual is not a high school graduate (1-B). Column 4 reports reduced-form effects of Z on log wages (Y). In panel A effects are for full sample, while in panel B the effects are reported for two strata differentiated by the average difficulty of dropping out of high school (X). The single-treatment IV coefficient is the ratio of reduced-form to first-stage effects in Panel A. The two-treatment IV coefficients vector is the inverse of the first-stage matrix in Panel B post-multiplied by the reduced-form vector. Under the assumptions in the text, these represent the local average treatment effect of GED certification for all compliers (panel A), and for compliers who would otherwise drop out of high school or otherwise earn a traditional diploma (panel B).

Table 2: Estimated first-year effects from the Oregon Health Insurance Experiment

	Restricted IV (1)	isoLATE IV (2)	Difference (3)
A. Healthcare utilization (extensive margin)			
Using prescription drugs currently	0.088	0.083	0.005
	(0.029)	(0.031)	(0.009)
Outpatient visits, last 6 months	0.212	0.197	0.015
	(0.025)	(0.039)	(0.022)
ER visits, last 6 months	0.022	0.015	0.007
	(0.023)	(0.034)	(0.017)
Inpatient hospital admissions, last 6 months	0.008	-0.007	0.014
	(0.014)	(0.022)	(0.013)
B. Compliance with preventative care			
Blood cholesterol checked (ever)	0.114	0.138	-0.023
	(0.026)	(0.041)	(0.027)
Blood tested for high blood sugar/diabetes (ever)	0.090	0.077	0.013
	(0.026)	(0.038)	(0.021)
Mammogram, last 12 months (women $\geq$ 40)	0.187	0.134	0.053
	(0.040)	(0.081)	(0.064)
Pap test, last 12 months (women)	0.183	0.170	0.013
	(0.034)	(0.052)	(0.029)
C. Financial strain			
Any out of pocket expenses, last 6 months	-0.200	-0.173	-0.026
	(0.026)	(0.042)	(0.027)
Owe money for medical expenses currently	-0.180	-0.187	0.007
	(0.026)	(0.038)	(0.019)
Trouble paying medical bills, last 6 months	-0.154	-0.122	-0.033
	(0.025)	(0.044)	(0.031)
Refused treatment due to debt, last 6 months	-0.036	-0.037	0.001
	(0.014)	(0.017)	(0.006)
D. Health outcomes			
Self-reported heath good/very good/excellent	0.133	0.098	0.035
	(0.026)	(0.050)	(0.038)
Self-reported health not poor	0.099	0.094	0.005
	(0.018)	(0.025)	(0.012)
Health about the same or better, last 6 months	0.113	0.089	0.024
	(0.023)	(0.038)	(0.026)
# of days physical health good, past 30 days	1.317	1.306	0.011
	(0.563)	(0.735)	(0.350)
# past 30 days poor health did not impair activity	1.585	1.773	-0.188
	(0.606)	(0.832)	(0.389)
# of days mental health good, past 30 days	2.082	1.857	0.225
	(0.640)	(0.929)	(0.521)
E. Access to care			
Have usual place of clinic-based care	0.339	0.367	-0.028
	(0.027)	(0.043)	(0.027)
Have personal doctor	0.280	0.282	-0.001
	(0.026)	(0.037)	(0.019)
Got all needed medical care, last 6 months	0.239	0.187	0.052
	(0.022)	(0.055)	(0.047)
Got all needed drugs, last 6 months	0.195	0.151	0.044
	(0.019)	(0.047)	(0.040)
Didn't use ER for nonemergency, last 6 months	-0.004	-0.004	0.000
	(0.015)	(0.022)	(0.011)

Notes: This table reports IV estimates of the effects of Medicaid on 12-month survey outcomes using randomized Medicaid offers from the Oregon Health Insurance Experiment as instruments. Column 1 uses a single treatment variable, restricts estimation to those individuals with valid survey responses for each outcome, and weights by the inverse probability of intensive follow-up, as in Finkelstein et al. (2012). Column 2 is estimated with two endogenous variables, as described in the text, using the full sample of 78,374 individuals. Column 3 reports differences in these estimates. All regression control for experimental strata. Robust standard errors, clustered by household, are reported in parentheses.

References

Abadie, A. (2003): "Semiparametric Instrumental Variables Estimation of Treatment Response Models," *Journal of Econometrics*, 113(2), 231–263.

Andrews, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59, 307–345.

Angrist, J. D. and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," Cambridge University Press, Advances in Economics and Econometrics.

Angrist, J. D. and M. Rokkanen (2016): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of the American Statistical Association*, forthcoming.

Behaghel, L., B. Crépon, and M. Gurgand (2013): "Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial," IZA Discussion Paper 7447.

Behaghel, L., B. Crépon, M. Gurgand, and T. L. Barbanchon (2009): "Sample Attrition Bias in Randomized Experiments: A Tale of Two Surveys," IZA Discussion Paper 4162.

——— (2015): "Please Call Again: Correcting Nonresponse Bias in Treatment Effect Models," *Review of Economics and Statistics*, 97, 1070–1080.

Blackwell, M. (2017): "Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments," *Journal of the American Statistical Association*, 112, 590–599.

Cragg, J. G. and S. G. Donald (1993): "Testing Identifiability and Specification in Instrumental Variable Models," *Econometric Theory*, 9, 222–240.

DiNardo, J., J. McCrary, and L. Sanbonmatsu (2006): "Constructive Proposals for Dealing with Attrition: An Empirical Example," Working Paper.

Duflo, E., R. Glennerster, and M. Kremer (2008): "Using Randomization in Development Economics Research: A Toolkit," Elsevier, Handbook of Development Economics 4, 3895–3962.

Engberg, J., D. Epple, J. Imbrogno, H. Sieg, and R. Zimmer (2014): "Education Programs That Have Lotteried Admission and Selective Attrition," *Journal of Labor Economics*, 32(1), 27–63.

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

Frölich, M. and M. Huber (2014): "Treatment Evaluation with Multiple Outcome Periods under Endogeneity and Attrition," *Journal of the American Statistical Association*, 109, 1697–1711.

Gronau, R. (1974): "Wage Comparisons – A Selectivity Bias," *Journal of Political Economy*, 82(6), 1119–1143.

Heckman, J. J. (1976): "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," NBER, vol. 5 of *Annals of Economic and Social Measurement*, 475–492.

Heckman, J. J., J. E. Humphries, P. A. LaFontaine, and P. L. Rodríguez (2012): "Taking the Easy Way Out: How the GED Testing Program Induces Students to Drop Out," *Journal of Labor Economics*, 30(3), 495–520.

Heckman, J. J. and S. Urzúa (2010): "Comparing IV with Structural Models: What Simple IV Can and Cannot Identify," *Journal of Econometrics*, 156, 27–37.

Hirano, K., G. W. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71, 1161–1189.

Hull, P. (2018): "Estimating Hospital Quality with Quasi-experimental Data," Working Paper.

Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475.

Kirkebøen, L., E. Leuven, and M. Mogstad (2016): "Field of Study, Earnings, and Self-Selection," *Quarterly Journal of Economics*, 101, 1057–1111.

Kline, P. and C. Walters (2016): "Evaluating Public Programs with Close Substitutes: The Case of Head Start," *The Quarterly Journal of Economics*, 131, 1795–1848.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *Review of Economic Studies*, 76, 1071–1102.

Lee, S. and B. Salanié (2016): "Identifying Effects of Multivalued Treatments," Working Paper.

Mountjoy, J. (2017): "Community Colleges and Upward Mobility," Working Paper.

Newey, W. K. (1994a): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

——— (1994b): "Series Estimation of Regression Functionals," *Econometric Theory*, 10, 1–28.

Reardon, S. F. and S. W. Raudenbush (2013): "Under What Assumptions Do Site-by-Treatment Instruments Identify Average Causal Effects?" *Sociological Methods Research*, 42.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688–701.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Online Appendix

Proof of Lemma 1

Consider the reduced-form regression of Y on Z, conditional on X. By the excludability of Z given treatment and fallback status (A2), we can write

$$Y = Y_1 + (Y_a - Y_1)A + (Y_b - Y_1)B,$$

so that the regression coefficient on Z identifies:

$$\begin{aligned}
E[Y|Z=1,X] - E[Y|Z=0,X] =& E[Y_1 + (Y_a - Y_1)A + (Y_b - Y_1)B|Z=1,X] \\
&- E[Y_1 + (Y_a - Y_1)A + (Y_b - Y_1)B|Z=0,X] \\
=& E[Y_1|Z=1,X] - E[Y_1|Z=0,X] \\
&+ E[(Y_a - Y_1)A_1|Z=1,X] - E[(Y_a - Y_1)A_0|Z=0,X] \\
&+ E[(Y_b - Y_1)B_1|Z=1,X] - E[(Y_b - Y_1)B_0|Z=0,X] \\
=& E[(Y_a - Y_1)(A_1 - A_0)|X] + E[(Y_b - Y_1)(B_1 - B_0)|X] \\
=& E[Y_1 - Y_a|A_1 < A_0, X]Pr(A_1 < A_0|X) \\
&+ E[Y_1 - Y_b|B_1 < B_0, X]Pr(B_1 < B_0|X),
\end{aligned}$$

where the third equality follows by independence of Z given X (A1) and the fourth by monotonicity (A3). Furthermore, the conditional first-stage regression of D on Z is

$$\begin{aligned}
E[D|Z=1,X] - E[D|Z=0,X] &= E[D_1 - D_0|X] \\
&= E[(1 - A_1 - B_1) - (1 - A_0 - B_0)|X] \\
&= E[A_0 - A_1|X] + E[B_0 - B_1|X] \\
&= Pr(A_1 < A_0|X) + Pr(B_1 < B_0|X).
\end{aligned}$$

This again follows by A1 and A3. The conditional IV coefficient on D is the ratio of reduced-form to first-stage expressions, completing the proof □

Proof of Proposition 1

The proof to Lemma 1 shows that under A1-A3 the conditional reduced form identifies

$$E[Y|Z=1,X] - E[Y|Z=0,X] = \alpha(X)\pi_a(X) + \beta(X)\pi_b(X).$$

Furthermore, the conditional first-stage regressions for $1-A$ and $1-B$ are

$$\begin{aligned} E[1-A|Z=1,X] - E[1-A|Z=0,X] &= -E[A_1 - A_0|X] \\ &= Pr(A_1 < A_0|X) \\ &= \pi_a(X) \end{aligned}$$

and

$$E[1-B|Z=1,X] - E[1-B|Z=0,X] = \pi_b(X).$$

As with Lemma 1, these follow from A1 and A3.

Consider the multiple-endogenous variable IV regression of equation (3). Let $\mathbf{Y}$ denote a vector of observations of Y, $\mathbf{X}$ a matrix of observations of $1-A$ and $1-B$, and $\mathbf{Z}$ a matrix of observations of Z and ZX. The endogenous regressor coefficients then satisfy:

$$\begin{aligned} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} &= p\lim\left((\widetilde{\mathbf{Z}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{Z}}'\mathbf{Y}\right) \\ &= p\lim\left(((\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{X}})^{-1}(\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\mathbf{Y}\right), \end{aligned}$$

where $\widetilde{\mathbf{Z}}$ and $\widetilde{\mathbf{X}}$ are matricies of residuals from regressing $\mathbf{Z}$ and $\mathbf{X}$ on X and a constant. By above,

$$\begin{aligned} p\lim\left((\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\widetilde{\mathbf{X}}\right) &= \begin{bmatrix} \pi_a(0) & \pi_b(0) \\ \pi_a(1) & \pi_b(1) \end{bmatrix} = \Pi \\ p\lim\left((\widetilde{\mathbf{Z}}'\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}'\mathbf{Y}\right) &= \begin{bmatrix} \alpha(0)\pi_a(0) + \beta(0)\pi_b(0) \\ \alpha(1)\pi_a(1) + \beta(1)\pi_b(1) \end{bmatrix}. \end{aligned}$$

When Π is invertible, the continuous mapping theorem and Slutsky's theorem imply:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \pi_a(0) & \pi_b(0) \\ \pi_a(1) & \pi_b(1) \end{bmatrix}^{-1} \begin{bmatrix} \alpha(0)\pi_a(0) + \beta(0)\pi_b(0) \\ \alpha(1)\pi_a(1) + \beta(1)\pi_b(1) \end{bmatrix}.$$

Proposition 1 follows from simplifying this expression. □

Proof of Proposition 2

Define for each $V \in \{A, B, Y\}$

$$\begin{aligned}\delta^V_{W,X} &= E[V|Z=0,W,X] - E[V|Z=1,W,X] = \frac{E[(1-Z)V|W,X]}{1-E[Z|W,X]} - \frac{E[ZV|W,X]}{E[Z|W,X]} \\ &= E\left[\frac{E[Z|W,X]-Z}{E[Z|W,X](1-E[Z|W,X])}V|W,X\right] \\ &= E[\lambda V|W,X],\end{aligned}$$

and note that by A1′ and A2 we have

$$\begin{aligned}\delta^A_{W,X} &= E[A_0 - A_1|W,X] \\ \delta^B_{W,X} &= E[B_0 - B_1|W,X] \\ \delta^Y_{W,X} &= E[(Y_a - Y_1)(A_0 - A_1) + (Y_b - Y_1)(B_0 - B_1)|W,X].\end{aligned}$$

Next define the random vector

$$\begin{bmatrix}\alpha_W \\ \beta_W\end{bmatrix} = \begin{bmatrix}\delta^A_{W,0} & \delta^B_{W,0} \\ \delta^A_{W,1} & \delta^B_{W,1}\end{bmatrix}^{-1}\begin{bmatrix}\delta^Y_{W,0} \\ \delta^Y_{W,1}\end{bmatrix}.$$

Here α_W and β_W are conditional analogues of the multiple endogenous variable IV specification used in Proposition 1. Focusing on the first, we may write

$$\alpha_W = \frac{\delta^Y_{W,0}\delta^B_{W,1} - \delta^Y_{W,1}\delta^B_{W,0}}{\delta^A_{W,0}\delta^B_{W,1} - \delta^A_{W,1}\delta^B_{W,0}}.$$

Furthermore,

$$\begin{aligned}\delta^Y_{W,0}\delta^B_{W,1} - \delta^Y_{W,1}\delta^B_{W,0} =& E[(Y_a - Y_1)(A_0 - A_1)|W,X=0]E[B_0 - B_1|W,X=1] \\ &+ E[(Y_b - Y_1)(B_0 - B_1)|W,X=0]E[B_0 - B_1|W,X=1] \\ &- E[(Y_a - Y_1)(A_0 - A_1)|W,X=1]E[B_0 - B_1|W,X=0] \\ &- E[(Y_b - Y_1)(B_0 - B_1)|W,X=1]E[B_0 - B_1|W,X=0] \\ =& E[Y_1 - Y_a|A_1 < A_0, W, X=0]Pr(A_1 < A_0|W,X=0)Pr(B_1 < B_0|W,X=1) \\ &+ E[Y_1 - Y_b|B_1 < B_0, W, X=0]Pr(B_1 < B_0|W,X=0)Pr(B_1 < B_0|W,X=1) \\ &- E[Y_1 - Y_a|A_1 < A_0, W, X=1]Pr(A_1 < A_0|W,X=1)Pr(B_1 < B_0|W,X=0) \\ &- E[Y_1 - Y_b|B_1 < B_0, W, X=1]Pr(B_1 < B_0|W,X=1)Pr(B_1 < B_0|W,X=0) \\ =& E[Y_1 - Y_a|A_1 < A_0, W](\delta^A_{W,0}\delta^B_{W,1} - \delta^A_{W,1}\delta^B_{W,0}),\end{aligned}$$

where the second equality follows by A3 and the third by A4′. Thus

$$E[Y_1 - Y_a | A_1 < A_0, W] = \alpha_W,$$

so that, by the Law of Iterated Expectations,

$$\begin{aligned} E[Y_1 - Y_a | A_1 < A_0] &= E\left[\frac{E[(Y_1 - Y_a)(A_1 - A_0)|W]}{Pr(A_1 < A_0)}\right] \\ &= E\left[\frac{Pr(A_1 < A_0|W)}{Pr(A_1 < A_0)}\alpha_W\right]. \end{aligned}$$

Finally, note that we can write

$$\begin{aligned} \delta^V_{W,0}\delta^B_{W,1} - \delta^V_{W,1}\delta^B_{W,0} &= E[\lambda V|W, X=0]E[\lambda B|W, X=1] - E[\lambda V|W, X=1]E[\lambda B|W, X=0] \\ &= \frac{E[\lambda V(1-X)|W]E[\lambda BX|W]}{(1-E[X|W])E[X|W]} - \frac{E[\lambda VX|W]E[\lambda B(1-X)|W]}{E[X|W](1-E[X|W])} \\ &= E\left[\lambda \frac{E[\lambda BX|W] - E[\lambda B|W]X}{E[X|W](1-E[X|W])} V|W\right] \\ &= E\left[\lambda \mu_b V|W\right], \end{aligned}$$

and

$$\begin{aligned} Pr(A_1 < A_0|W) &= E[E[A_0 - A_1|W, X]|W] \\ &= E[E[\lambda \cdot A|W, X]|W] \\ &= E[\lambda A|W]. \end{aligned}$$

Thus, once again applying the Law of Iterated Expectations,

$$E[Y_1 - Y_a | A_1 < A_0] = E\left[\frac{E[\lambda A|W]}{E[\lambda A]} \frac{\lambda \mu_b Y}{E\left[\lambda \mu_b A|W\right]}\right]$$

The same steps show the result for $E[Y_1 - Y_b|B_1 < B_0]$. □

Note that the function $\lambda(w, x, z)$ generating $\lambda = \lambda(W, X, Z)$ is identified by the conditional expectation function $E[Z|W, X]$ and that

$$\begin{aligned} E[\lambda A|W = w] = \sum_{x=0,1}\sum_{z=0,1} &\lambda(w, x, z)E[A|W = w, X = x, Z = z] \\ &\times Pr(Z = z|W = w, X = x)Pr(X = x|W = w) \end{aligned}$$

and similarly for $E[\lambda B|W]$. Moreover,

$$E[\lambda BX|W = w] = \sum_{z=0,1} \lambda(w, 1, z)E[B|W = w, X = 1, Z = z]Pr(Z = z|W = w, X = 1)E[X|W = w].$$

Thus both the weights $E[\lambda A|W=w]/E[\lambda A]$ and the function $\mu_b(w,x)$ generating $\mu_b = \mu_b(W,X)$ are identified by the conditional expectation functions $E[X|W]$, $E[Z|W,X]$, $E[A|W,X,Z]$, and $E[B|W,X,Z]$. Finally, note that

$$E[\lambda\mu_b A|W=w] = \sum_{x=0,1}\sum_{z=0,1} \lambda(w,x,z)\mu_a(w,z)E[A|W=w,X=x,Z=z] \\ \times Pr(Z=z|X=x,W=w)Pr(X=x|W=w)$$

We can thus form sample analogues of the weighting schemes identifying $E[Y_1 - Y_a|A_1 < A_0]$ from nonparametric estimates of these conditional expectation functions. Similar results follow for $E[Y_1 - Y_b|B_1 < B_0]$.

GED Selection Model

Section 3.1 simulates data on educational attainment and labor market returns using a model inspired by that of Heckman and Urzúa (2010). Potential log hourly earnings are given by

$$Y_{ti} = \mu_t + \gamma X_i + \epsilon_{it}, \tag{30}$$

where $X_i = 1$ is a cohort indicator and $t \in \{1, a, b\}$ indexes the individual's educational status: GED-certified, high school dropout, or traditional high school graduate. Individuals observe the schooling environment and chooses t that maximizes ν_{ti}, where

$$\nu_{1i} = \Phi(\pi_1 \tilde{Z}_{1i} - \eta_{1i}) \tag{31}$$

$$\nu_{ai} = \Phi(\pi_a \tilde{Z}_{ai} - \eta_{ai}) \mathbf{1}[X_i \geq \xi_i] \tag{32}$$

$$\nu_{bi} = \Phi(\pi_b \tilde{Z}_{bi} - \eta_{bi}), \tag{33}$$

and where $\Phi(\cdot)$ denotes the normal CDF. That is, individuals choose the schooling level that gives them the highest latent utility, subject to the constraint that some may not be allowed to drop out of high school by virtue of being too young ($\mathbf{1}[X_i < \xi_i]$). To simulate the model, I let $(\tilde{Z}_{1i}, \tilde{Z}_{ai}, \tilde{Z}_{bi}) \sim N(\mu_Z, \Sigma_Z)$ and $(\epsilon_{1i}, \epsilon_{ai}, \epsilon_{bi}, \eta_{1i}, \eta_{ai}, \eta_{bi}) \sim N(0, \Sigma_{\epsilon\nu})$ where

$$\Sigma_Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Sigma_{\epsilon\nu} = \begin{bmatrix} 0.64 & 0.16 & 0.16 & 0.024 & -0.32 & 0.016 \\ 0.16 & 1 & 0.2 & 0.02 & -0.3 & 0.01 \\ 0.16 & 0.2 & 1 & 0.02 & -0.4 & 0.04 \\ 0.024 & 0.02 & 0.02 & 1 & 0.6 & 0.1 \\ -0.32 & -0.3 & -0.4 & 0.6 & 1 & 0.2 \\ 0.016 & 0.01 & 0.04 & 0.1 & 0.2 & 1 \end{bmatrix}, \tag{34}$$

and where $(\mu_1, \mu_a, \mu_b) = (0.3, 0.1, 0.7)$ and $(\pi_1, \pi_a, \pi_b) = (0.2, 0.3, 0.1)$. With $X_i = \xi_i = 0$, this model is identical to that of Heckman and Urzúa (2010).

To generate cross-strata variation I let $\xi \sim N(0.5, 0.025)$ and $X \sim Bernoulli(0.5)$. Setting $\gamma = 0.2$ with allows an individual's cohort to affect the level of her adulthood wages, but not her relative returns to schooling, satisfying A4. As in Heckman and Urzúa (2010), I apply Proposition 1 with an instrument Z_i that represents an exogenous increase in $\tilde{Z}_{1i}$ by 0.75 standard deviations. Population moments of this model are reported in Table 1. For the Monte Carlo simulations in Figure 2, I draw 5,000 sets of 200,000 independent observations from the model while varying the first-stage parameter π_1 from 0.01 to 0.5. This leads to variation in Cragg and Donald (1993) first-stage F-statistics; the benchmark with $\pi_1 = 0.3$ corresponds to $F \approx 12.2$.