Interpreting Instrumental Variable Estimands with Unobserved Treatment Heterogeneity: The Effects of College Education*

Clint Harris†

October 1, 2022

Abstract

Many treatment variables used in empirical applications nest multiple unobserved versions of a treatment. I show that instrumental variable (IV) estimands for the effect of a composite treatment are IV-specific weighted averages of effects of unobserved component treatments. Differences between IVs in unobserved component compliance produce differences in IV estimands even without treatment effect heterogeneity. I describe a monotonicity condition under which IV estimands are positively-weighted averages of unobserved component treatment effects. Next, I develop a method that allows instruments that violate monotonicity to contribute to estimation of treatment effects by allowing them to place nonconvex, outcome-invariant weights on unobserved treatments across multiple outcomes. Finally, I apply the method to estimate lifecycle returns to college, finding returns that range from 7% to 30% over the life cycle. My findings emphasize the importance of leveraging instrumental variables that do not shift individuals between versions of treatment, as well as the importance of policies that encourage students to attend "high-return college" in addition to those that encourage "high-return students" to attend college.

JEL Codes: C36, I26, C38

Keywords: Returns to Education, Instrumental Variables, Local Average Treatment Effects, Unobserved Heterogeneity, Factor Models

1 Introduction

Treatment variables of social and empirical interest often come in multiple varieties. For instance, years of education encompasses educational time investments across subjects and levels of schooling, years of work experience encompasses time investments across a variety of tasks which develop different skills, and healthcare expenditures encompass a wide variety of medical services. When these differing components of composite treatments have different treatment effects, models which describe single effects of composite treatments on outcomes are misspecified, and the stable unit treatment value assumption (SUTVA) of Rubin (1974) is violated. This paper describes conditions under which standard instrumental variables (IV) methods identify convex weighted averages of unobserved treatments' effects, describes an alternative factor-augmented treatment effects model that allows IVs to place negative weights on unobserved treatments, and applies the new method to estimation of returns to college using the NLSY79.

For some applications involving treatments with unobserved treatment heterogeneity, estimated composite treatment effects may equally weight all component treatments, providing an estimate of the average effect of component treatments. In other cases, the estimated composite treatment effect may be inherited from a single unrepresentative component, which may not even share the sign of the average effect. Finally, if the variation in treatment used to estimate effects shifts individuals between unobserved component treatments sufficiently often, estimated composite treatment effects will reflect differences in effects between component treatments rather than a positively-weighted average effect of component treatments relative to nontreatment.

I begin with a theoretical discussion of standard instrumental variables methods when treatments come in multiple unobserved varieties. I show that instruments which shift individuals into (but never out of or between) multiple versions of treatment identify convex weighted averages of unobserved treatments' effects, essentially extending the insights of Imbens and Angrist (1994) and Heckman and Pinto (2018) to the setting with unobserved treatment heterogeneity. I show similar results for the case in which instruments shift relatively few individuals with non-extreme treatment effects between treatment varieties, extending the insights of De Chaisemartin (2017) to the same setting.

In addition to results on standard methods, I also develop a novel factor-augmented treatment effects (FATE) model that defines treatment effects in terms of a small number (such as

one) of particularly well-behaved instruments, while allowing other instruments to contribute to estimation. The method estimates a convex average of effects of unmodeled component treatments without requiring observation of the component treatments. The key insight motivating the method is that the varieties of treatment that individuals are shifted into by different instruments are outcome invariant, while the effects of unobserved varieties of treatment on outcomes are instrument invariant. This suggests a factor structure for just-identified IV treatment effect estimands wherein IVs work through a lower dimensional number of unobserved treatments to affect a large number of outcomes in logically consistent ways. When instruments are assumed to affect outcomes through a single treatment variety, this model is identical to that of 2SLS with multiple instruments, whereas it is identical to just-identified IV for multiple instruments when the number of treatment varieties is assumed to be equal to the number of instruments.

I use the FATE model to estimate returns to college over the lifecycle using the NLSY79. I find that college attendance associated with local labor market conditions produces a higher wage return profile than college attendance associated with local college proximity. Because local wage variation likely affects college attendance by making the outside option more or less attractive, it is reasonable to think that it does not shift individuals between versions of treatment, suggesting these (high) return estimates are more likely to be valid. It follows that returns to college in the literature that use cost-shifters, which may shift individuals into low effort college because of income effects, may negatively weight high-return types of college and understate true returns to college on average. Comparing the estimated factor-augmented treatment effects to those of just-identified IV shows reductions in standard errors, consistent with the intuition of FATE allowing multiple instruments to work together to estimate treatment effects (as in the extreme case of multiple-IV 2SLS).

Finally, I contribute to the literature on economic applications for factor models. Much of the existing work along these lines uses factor models to extract particularly relevant variation in data with multiple outcomes, where the multiple outcomes often include auxiliary measures such as test scores (Hansen, Heckman, and Mullen, 2004; Cunha, Heckman, and Navarro, 2005; Heckman, Stixrud, and Urzua, 2006; Jiang, 2019) or repeated outcomes of interest in panel data (Bai, 2009). In these cases, the main use of factor models is to extract common sources of variation across outcomes to define otherwise unobserved types of individual-level heterogeneity that vary in their effects on outcomes. These factors act as controls in equations of interest to address omitted variable bias and enable identification of treatment effects of other observed

variables. My factor model leans on this past work in the estimation procedure, but differs substantially in intuition. Rather than extracting particularly meaningful variation in data to identify factors that act as control variables, my factor model extracts particularly meaningful variation in estimated marginal effects to determine their structural interpretation.

The plan of the paper is as follows. Section 2 describes identification of treatment effects for standard IV methods and introduces a net uniform unordered monotonicity condition under which the IV estimand is a positively-weighted average of component treatment effects. Section 3 describes the factor-augmented treatment effects model. Section 4 describes the data for the application on returns to college. Section 5 discusses model estimates from the factor-augmented treatment effects model and their interpretation and relationship to those of multiple instrument 2SLS and just-identified IV. Section 6 concludes.

2   IV Estimands under Unobserved Treatment Heterogeneity

In this section, I will describe conditions under which standard IV methods identify meaningful average effects of treatments with unobserved heterogeneity. In the context of the Rubin (1974) causal model (RCM), unobserved treatment heterogeneity constitutes a violation of the Stable Unit Treatment Value Assumption, which requires there to be only one version of each treatment. I will start by presenting the problem using RCM potential outcome language and notation and going through some assumptions that will be maintained in subsequent sections. Next, I will present sufficient conditions for treatment effect identification in the case where treatments are heterogeneous and their effects are homogeneous, wherein unobserved defier and complier treatments will play similar roles to those of defier and complier individuals in the more familiar case of homogeneous treatments with heterogeneous effects. Afterward, I will allow for both unobserved treatment heterogeneity and treatment effect heterogeneity with a strict monotonicity assumption similar to those of Imbens and Angrist (1994) and Heckman and Pinto (2018). I close the section by relaxing this monotonicity assumption for cases when treatments and their effects are heterogeneous by applying the insights of De Chaisemartin (2017) to allow for compliers (for each version of treatment) to cancel out defiers (for the same version of the treatment) who share their treatment effects.

2.1 Potential Outcomes with Unobserved Treatment Heterogeneity

I work within a version of the Rubin (1974, 1990) causal model similar to that used by Imbens and Angrist (1994), with reference as well to Heckman (1990), Angrist and Imbens (1991), and Angrist, Imbens, and Rubin (1996). The main innovation of this paper concerns an observed composite treatment for individual i, D_i, that is composed of L unobserved mutually exclusive component treatments such that $D_i = \sum_{\ell=1}^{L} D_{i\ell}$, with the vector $\mathbf{D}_i = (D_{i0}, D_{i1}, ..., D_{iL})$ denoting i's treatment status for each version of the treatment, with D_{i0} denoting the untreated state. Because component treatments are binary and mutually exclusive, it follows that D_i is binary as well.

I assume that the outcome Y_i is observed with $Y_i = \sum_{\ell=0}^{L} D_{i\ell} Y_i(\ell)$, where $Y_i(\ell)$ gives the value of $Y_i(\mathbf{D}_i)$ when $D_{i\ell} = 1$ and $D_{im} = 0$ for all $m \neq \ell$. We observe $Y_i(0)$ for individuals that do not select treatment, but not those that do, as is standard in the potential outcomes framework. However, with unobserved treatment heterogeneity, $D_i = 1$ is not sufficient to determine the value of $D_{i\ell}$, other than to bound ℓ at $\ell > 0$. It follows that not only do we not observe $Y_i(\ell)|D_i = 0$ for any i or $\ell > 0$, but we also do not observe $Y_i(\ell)|D_{im} = 1$ for any i for $\ell \neq m$.[1] This muddling of potential outcomes for treated individuals is the central empirical challenge I address in this paper. Not only is the single composite treatment effect not observed for any individuals, it not well-defined. We instead are left with multiple component treatment effects, $Y_i(\ell) - Y_i(0)$, which can be differenced to define unobserved treatment switching effects, $Y_i(\ell) - Y_i(\ell')$ for all $\ell \neq \ell'$.

In the language of the Rubin causal model, unobserved treatment heterogeneity constitutes a violation of the stable unit treatment value assumption (SUTVA). In randomized controlled trials, random assignment to treatment ensures that $\mathbb{E}[Y_i(\ell)|D_{i\ell} = 1] = \mathbb{E}[Y_i(\ell)|D_{im} = 1]$ for all ℓ and m, including the cases where $\ell = 0$ or $m = 0$. Experimental control efforts are intended to ensure that $D_i = D_{i\ell}$ for a given ℓ for all individuals in an observed treatment arm. In other words, the administration of the treatment is explicitly designed to hold equal between treated individuals all variations that are expected to substantively alter the nature of the treatment.[2] The control of the trial ensures identification of treatment effects that are entirely driven by a

---

[1] More precisely, we do observe $Y_i(\ell)$ for individuals who choose component treatment ℓ, but I do not know the ℓ for which we are observing $Y_i(\ell)$.

[2] Even well-designed experiments differ in this respect. Variation in subtle priming from the outward affect or word choices of research staff administering a medical treatment might not complicate treatment effect identification for treatments of physical ailments but may do so for treatments of psychological ailments.

single component treatment, while the randomization ensures that the effects can be estimated via comparisons of treated and untreated individuals.

In observational studies such as the application below on the wage returns to college, treatment assignments are not explicitly randomized, individuals who select treatment are not necessarily representative of the population, and versions of treatment received by individuals are not controlled. To address the lack of randomization of treatment, it is common for researchers to make use of instrumental variables, variables which are assumed to be as-good-as-random that shift individuals into (or out of) treatment without otherwise affecting outcomes. The implications for treatment effect estimation of selected subpopulations of individuals responding to instrumental variables are well-understood and commonly addressed in empirical work. The implications of individuals responding to instrumental variables by selecting particular unmodeled varieties of treatment are less well-understood, but produce similar problems, as I will show. This is especially of interest for estimates of returns to college, considering the large literature detailing differences in effects of college on outcomes depending on variations in treatment such as the subjects of study (Hastings, Neilson, and Zimmerman, 2013; Kirkeboen, Leuven, and Mogstad, 2016) or the quality of the college (Black and Smith, 2006; Dillon and Smith, 2020).

In addition to observing the outcome and composite treatment, I assume that an instrumental variable $Z_i \in \mathcal{Z}$ is observed as well, composite and component treatments $D_i(z)$ and $D_{i\ell}(z)$ representing random variables that give the value of each respective treatment if $Z_i = z$. I assume that (Y_i, D_i, Z_i) is observed for a random sample of the population. I assume that instrumental variables satisfy several conditions, which I will detail.

Condition 1 (Independence). *Let Z be a random variable such that $D_{i\ell}(z)$ is independent of Z_i for all $\ell = 0, 1, ..., L$ and all $z \in \mathcal{Z}$.*

Condition 1 would be sufficient to identify the effect of Z_i on $D_{i\ell}$ if $D_{i\ell}$ were observed. It follows from this condition and the definition of D_i that D_i is independent of Z_i, further implying that the effect of Z_i on the composite treatment, D_i, is identified as well.

In addition to the multiple treatments version of independence just given, I also present a multiple treatments version of exclusion.

Condition 2 (Exclusion). *Let Z be a random variable such that $Y(\ell)$ is independent of Z_i for all $\ell = 0, 1, ..., L$ and all $z \in \mathcal{Z}$.*

The standard composite treatment version of exclusion is likely to be violated in settings with

SUTVA violations, while the multiple treatment version allows will not be for well-chosen instruments.[3] This is readily seen in the regression-based treatment effects model with homogeneous effect λ_ℓ for treatment ℓ,

$$
\begin{aligned}
Y_i &= \lambda_0 + \sum_{\ell=1}^{L} D_{i\ell}(Z_i)\lambda_\ell + \epsilon_i \\
&= \lambda_0 + 0 D_i(Z_i) + (\epsilon_i + \sum_{\ell=1}^{L} D_{i\ell}(Z_i)\lambda_\ell),
\end{aligned}
\tag{1}
$$

where the term $\sum_{\ell=1}^{L} D_{i\ell}(Z_i)\lambda_\ell$ is in the error term in the homogeneous treatment version of the model in the second line. The $Y_i(1)$ outcome is clearly dependent on Z_i via channels other than D_i, unless the instrument only ever induces a particular type of treatment (such as in the case of a controlled experiment, or an intent-to-treat instrument in the same context).

Conditions 1 and 2 imply that the effect of Z_i on Y_i is identified in addition to the effect of Z_i on D_i. The effect of Z_i on Y_i, defined for a comparison of two points (z, z') in the support of $\mathcal{Z}$, is given by

$$
\begin{aligned}
&\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] \\
=&\mathbb{E}[\sum_{\ell=1}^{L} D_{i\ell}(z)Y_i(\ell) + (1 - \sum_{\ell=1}^{L} D_{i\ell}(z))Y_i(0)|Z_i = z] \\
-&\mathbb{E}[\sum_{\ell=1}^{L} D_{i\ell}(z')Y_i(\ell) + (1 - \sum_{\ell=1}^{L} D_{i\ell}(z'))Y_i(0)|Z_i = z'].
\end{aligned}
\tag{2}
$$

Conditions 1 and 2 imply that this is equal to

$$
\begin{aligned}
&\sum_{\ell=1}^{L} \mathbb{E}\Big[(D_{i\ell}(z) - D_{i\ell}(z'))(Y_i(\ell) - Y_i(0))\Big] \\
=&\sum_{\ell=1}^{L} \Pr[D_{i\ell}(z) - D_{i\ell}(z') = 1]\mathbb{E}[Y_i(\ell) - Y_i(0)|D_{i\ell}(z) - D_{i\ell}(z') = 1] \\
-&\sum_{\ell=1}^{L} \Pr[D_{i\ell}(z) - D_{i\ell}(z') = -1]\mathbb{E}[Y_i(\ell) - Y_i(0)|D_{i\ell}(z) - D_{i\ell}(z') = -1].
\end{aligned}
\tag{3}
$$

The IV estimand for the effect of D_i on Y_i for instrument comparison z, z' is obtained by dividing

---

[3]Angrist, Imbens, and Rubin (1996) describe SUTVA as as ruling out treatment spillovers, while modeling outcomes, treatment assignments, and instruments for all individuals as vectors. The arguably simpler presentation in Imbens and Angrist (1994) works, like this paper, at the individual level and makes no mention of SUTVA, implicitly ruling out problematic spillovers and other forms of unobserved treatment heterogeneity with their independence assumption (which includes what I and others call exclusion).

the average effect of Z_i on Y_i by the average effect of Z_i on D_i,

$$\pi_{z,z'} = \frac{\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z']}{P(z) - P(z')}, \tag{4}$$

where $P(z) = \mathbb{E}[D_i|Z_i = z]$ gives the probability of treatment for instrument value $z \in \mathcal{Z}$. There are three potential problems with this estimand even when conditions 1 and 2 are satisfied. In brief, problems arise with this estimand if instruments do not appreciably affect composite treatment status, if instruments shift some individuals into treatments and some out, or if instruments shift individuals into some component treatments and out of others.

First, this estimand is indeterminant or undefined if $P(z) - P(z') = 0$. This is addressed with the usual testable condition that the instrument has an appreciable effect on observed composite treatment status,

Condition 3 (Relevance). *Let Z be a random variable such that $P(z)$ is a nontrivial function of z.*

It follows from condition 3 that the instrument must also have an effect on at least one component treatment, such that the unobserved value $P_\ell(z) = \mathbb{E}[D_{i\ell}|Z_i = z]$ is also a nontrivial function of z for some value of ℓ.

The second and third potential problems with the estimand in equation (4) involve the negative second term. The case in which instruments shift some individuals into a homogeneous treatment and some individuals out, with the treatment effects of those in the minority (defier individuals) being negatively weighted is addressed by the monotonicity condition of Imbens and Angrist (1994), which constrains this term to zero by assuming all individuals respond to the instrument in the same way. The closely-related problem presented by unobserved treatment heterogeneity is that instruments might shift individuals into some versions of treatment and out of others, with the treatment effects of treatments in the minority (defier treatments) being negatively weighted. Multiple sufficient conditions of varying plausibility in applications address this second concern. I will begin with less plausible (but simpler) sufficient conditions and work toward more plausible ones.

2.2 Heterogeneous Treatments with Homogeneous Effects

The simplest ways of addressing the problematic second term of equation (4) make assumptions that imply that $\sum_{\ell=1}^{L} \Pr[D_{i\ell}(z) - D_{i\ell}(z') = -1] = 0$ for all ℓ and all i (assuming without

loss of generality that $P(z) > P(z')$). An example of this approach is in Imbens and Angrist (1994), who implicitly assume $L = 1$, and that instruments only shift individuals into (not out of) treatment. I begin by considering conditions that mirror this approach in the setting of unobserved treatment heterogeneity. First, I consider the case where each unobserved type of treatment has homogeneous treatment effects.

Condition 4 (Homogeneous Treatment Effects). *For all $\ell = 1, 2, ..., L$, define λ_ℓ as a constant that satisfies $\lambda_\ell = Y_i(\ell) - Y_i(0)$.*

Condition 4, combined with 1 and 2, implies that the effect of Z_i on Y_i from equations (2) and (3) can be rewritten as

$$
\begin{aligned}
\mathbb{E}[Y_i | Z_i = z] - \mathbb{E}[Y_i | Z_i = z'] &= \sum_{\ell=1}^{L} (P_\ell(z) - P_\ell(z')) \lambda_\ell \\
&= \sum_{\ell=1}^{L} I(P_\ell(z) \geq P_\ell(z'))(P_\ell(z) - P_\ell(z')) \lambda_\ell \quad (5) \\
&\quad - \sum_{\ell=1}^{L} I(P_\ell(z) \leq P_\ell(z'))(P_\ell(z') - P_\ell(z)) \lambda_\ell,
\end{aligned}
$$

where $I(\cdot)$ is an indicator function that takes a value of unity if its argument is true and zero otherwise.

Even with the strong assumption of homogeneous treatment effects, the IV estimand in equation (4) can be uninformative about treatment effects. To see this, consider the example with high-return college and low-return college as the unobserved versions of treatment, with a randomly assigned low-return-major contingent college scholarship used as an instrument for college attendance. If returns to high-return college are double those of low-return college and this instrument shifts x individuals from non-college into low-return college and x individuals from high-return college into low-return college, the IV estimand in (4) will fail to provide information regarding returns to any type of college. In the model, this setting constitutes $L = 2$ with $\lambda_2 = 2\lambda_1$ and $P_1(z) - P_1(z') = -2(P_2(z) - P_2(z'))$, which produces a value of zero on the right hand side of equation (5).

To avoid some versions of treatment having their effects negatively weighted in the IV estimand, additional assumptions are required even for with treatments with constant effects and instruments that satisfy independence, exclusion, and relevance. A simple sufficient condition

imposes that that instrument never induces changes in unobservable intensive margin treatment status.

Condition 5 (Uniform Treatment Responses)**.** *For all $z, z' \in \mathcal{Z}$, either $P_\ell(z) \geq P_\ell(z')$ for all ℓ, or $P_\ell(z) \leq P_\ell(z')$ for all ℓ.*

Condition 5 does for treatments with heterogeneous effects what the monotonicity condition of Imbens and Angrist (1994) does for individuals with heterogeneous treatment effects.

If the above conditions hold, the estimand in equation (4) has a causal interpretation.

Theorem 1: *If Conditions 1-5 hold, the following weighted average component treatment effect is identified:*

$$\pi_{z,z'} = \sum_{\ell=1}^{L} \frac{P_\ell(z) - P_\ell(z')}{P(z) - P(z')} \lambda_\ell,$$

with $(P_\ell(z) - P_\ell(z'))/(P(z) - P(z')) \geq 0$ for all ℓ.

Proof. Let condition 5 hold with $P_\ell(z) \geq P_\ell(z')$ for all ℓ. It immediately follows that $(P_\ell(z) - P_\ell(z'))/(P(z) - P(z')) \geq 0$ for all ℓ. It also follows that the final term in equation (5) is zero, and

$$\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] = \sum_{\ell=1}^{L}(P_\ell(z) - P_\ell(z'))\lambda_\ell.$$

Diving both sides of (5) by $(P(z) - P(z'))$ shows that the weighted average component treatment effect above can be obtained from moments of the joint distribution of (Y_i, D_i, Z_i). $\qquad\square$

An economic model of the college attendance decision is helpful for understanding the connection between unobserved treatment heterogeneity and treatment effect heterogeneity. For clarity, consider the case of two unobserved component treatments with one observed composite treatment, such that $L = 2$. The discrete choice model is the same as the case where all treatments are observed, such as in Heckman and Pinto (2018) and Mountjoy (2022). Formally, I define individual i's utility from treatment decision $D_{i\ell}(Z_i)$ as $D_{i\ell}^*(Z_i)$, defined as

$$
\begin{aligned}
D_{i0}^*(Z_i) &= 0 \\
D_{i1}^*(Z_i) &= U_{i1} + \mu_1(Z_i) \\
D_{i2}^*(Z_i) &= U_{i2} + \mu_2(Z_i),
\end{aligned}
\tag{6}
$$

where the utility from noncollege is normalized to zero, and individuals choose the treatment

status that gives them the highest utility, such that

$$D_{i0}(Z_i) = I(D_{i1}^*(Z_i) < 0, D_{i2}^*(Z_i) < 0)$$
$$D_{i1}(Z_i) = I(D_{i1}^*(Z_i) \geq D_{i2}^*(Z_i), D_{i1}^*(Z_i) \geq 0) \qquad (7)$$
$$D_{i2}(Z_i) = I(D_{i2}^*(Z_i) \geq D_{i1}^*(Z_i), D_{i2}^*(Z_i) \geq 0).$$

The assumption that all individuals are uniformly affected by instruments ($\mu_i(Z_i) = \mu(Z_i)$ for all i) is unnecessary for the methods in this paper, but permits a helpful graphical representation of the model. Figure 1 represents the college attendance decision model in (7) with $(U_{1i}, U_{2,i})$ drawn from a distribution that implies that each point enclosed by the ellipse contains a common population density. The shares of the population attending each type of college at different values of Z are given by the area inside of the ellipse representing the joint distribution of $(D_1^*(Z), D_2^*(Z))$. The three panels each consider an instrument comparison relative to a common baseline with a relatively low college attendance rate. The instrument comparisons in Panels B and C satisfy condition 5, while the comparison in Panel A does not.

To fix ideas, consider that Panel A shows a control group, individuals for whom $Z_i = z$, and a treated group who are randomly selected to receive a STEM-major-contingent college scholarship, for whom $Z_i = z'$. The STEM-contingent scholarship induces individuals who would not otherwise attend college at all into STEM college ($\ell = 1$), but also induces individuals who would otherwise attend college with a non-STEM major into STEM college. In this case, $P_1(z') >> P_1(z)$ with $P_2(z') < P_2(z)$, violating condition 5. If unobserved treatment D_1 has larger (smaller) treatment effects than treatment D_2, estimating effects of college using the comparison (z, z') will be at risk of overstating (understating) the true returns of any type of college. Panels B and C display individuals in alternative treatment arms for whom $Z_i = z''$ and $Z_i = z'''$ which represent receipt of other randomly-assigned incentives for college that have less-extreme compositional effects on the shares of individuals attending each type of college.

2.3 Heterogeneous Treatments with Heterogeneous Effects

There is a conceptual point to be made that a fair amount of variation in effects of observed composite treatments between individuals may be attributable to variation in versions of treatment chosen, rendering condition 4 less restrictive than it is in the usual homogeneous treatments case. For instance, continuing with the STEM/non-STEM example above, consider the indi-
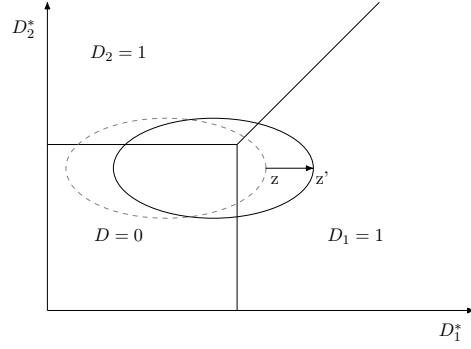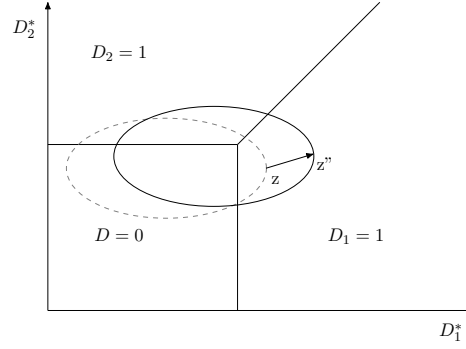
Panel A   Panel B   Panel C
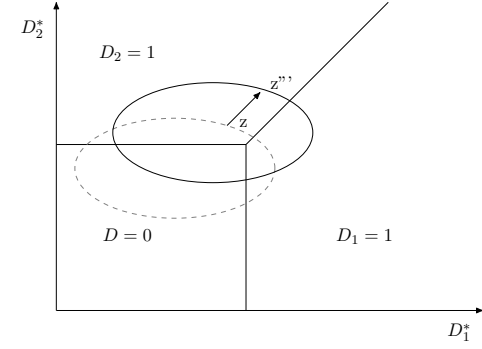
Figure 1: Instrument Comparisons with Two Unobserved Component Treatments

Notes: Graphic representations of instrument comparisons that satisfy independence, exclusion, and relevance in the context of decision model (7). Panel A shows an instrument comparison that satisfies Heckman and Pinto's (2018) unordered monotonicity condition, but violates uniform unordered monotonicity and net uniform unordered monotonicty (there are more type 2 defiers than compliers). Panel B shows an instrument comparison that violates uniform unordered monotonicity, but may satisfy net uniform unordered monotonicity with low treatment effect heterogeneity. Panel C's instrument comparison satisfies uniform unordered monotonicity and net uniform unordered monotonicity.

viduals who are at the margin of non-college and STEM college, who are extremely far from all other margins. Such individuals will generally respond to instruments (if they respond) by attending a particular type of college, so there is no practical distinction for them between treatment heterogeneity and treatment effect heterogeneity — they are, in the running example, STEM-margin individuals at the margin of STEM college. However, individuals who are close to the line $D_1^* = D_2^*$ are at the intensive margin between types of college, such that their returns to college cannot be completely described by attributing to them individual-specific returns; they may choose either type of college and will have different outcomes depending on their choice.[4] It follows that while condition 4 is less restrictive than the corresponding assumption with homogeneous treatments, it nonetheless may fail to capture meaningful variation in treatment effects between populations.

This section replaces conditions 4 and 5 with an alternative condition that allows for treatment effect identification when unobserved component treatments have individual-level treatment effect heterogeneity.

Condition 6 (Uniform Unordered Monotonicity). *For all* $z, z' \in \mathcal{Z}$, *either* $D_{i\ell}(z) \geq D_{i\ell}(z')$ *for all* $i = 1, 2, ..., N$ *and all* $\ell = 1, 2, ... L$, *or* $D_{i\ell}(z) \leq D_{i\ell}(z')$ *for all* $i = 1, 2, ..., N$ *and all* $\ell = 1, 2, ... L$.

This condition strengthens Heckman and Pinto's (2018) unordered monotonicity assumption to require that all treatments respond to instruments uniformly, that is, if an instrument makes one component treatment more likely, it must weakly make all others more likely as well. In related work on observed treatments, Kline and Walters (2016) describe a similar condition that restricts changes in the treatment decision to a single observed component treatment in the presence of multiple observed varieties of treatment, while Hull (2018) describes a condition that restricts changes in treatment decision from one of multiple observed untreated varieties to a single observed treatment variety.

Condition 6 implies that the IV estimand in (4) has a causal interpretation as a weighted average of effects of complier treatments on complier individuals.

Theorem 2: *If Conditions 1-3 and 6 hold, the following component-weighted local aver-*

---

[4]For instance, in Figure 1 many individuals along the top of the $D = 0$ partition when $Z_i = z$ choose treatment 1 when $Z_i = z'$ (Panel A), but they choose treatment 2 when $Z_i = z'''$ (Panel C).

age treatment effect is identified:

$$\pi_{z,z'} = \sum_{\ell=1}^{L} \frac{P_\ell(z) - P_\ell(z')}{P(z) - P(z')} \mathbb{E}[Y_i(\ell) - Y_i(0) | D_{i\ell}(z) \neq D_{i\ell}(z')],$$

with $(P_\ell(z) - P_\ell(z'))/(P(z) - P(z')) \geq 0$ for all ℓ.

Proof. Let Condition 6 hold with $D_{i\ell}(z) \geq D_{i\ell}(z')$ for all i and all ℓ. It immediately follows that $(P_\ell(z) - P_\ell(z'))/(P(z) - P(z')) \geq 0$ for all ℓ. It also follows that the final term in equation (5) is zero, and

$$\mathbb{E}[Y_i | Z_i = z] - \mathbb{E}[Y_i | Z_i = z'] = \sum_{\ell=1}^{L} (P_\ell(z) - P_\ell(z')) \mathbb{E}[Y_i(\ell) - Y_i(0) | D_{i\ell}(z) - D_{i\ell}(z') = 1].$$

Diving both sides of this by $(P(z) - P(z'))$ shows that the weighted average local average component treatment effect above can be obtained from moments of the joint distribution of (Y_i, D_i, Z_i). □

Condition 6 is quite strict. In the context of the decision model in (7), it will be satisfied in only two narrow cases. The first is the case in which the support of $\mathcal{Z}$ contains only two values, (z, z'), where $P(z) = 0$. It follows that any comparison (z, z') trivially satisfies a wide range of monotonicity conditions, including Condition 6.[5] The second is the case in which instruments make all treatments more or less attractive to all individuals by a constant amount. In Figure 1, this is illustrated in Panel C, where the slope of the effect of instruments on preferences for treatments is parallel to the unobserved decision margin, $D2^* = D_1^*$.[6]

2.4   Tolerating Defiance

It is difficult to imagine an instrument, other than treatment assignment in a randomized trial with perfect compliance, that strictly satisfies Condition 6. For instance, exogenous college cost shifters presumably have income effects, which may cause some individuals to switch to less-difficult versions of college, while college distance shifters may cause some individuals to

---

[5]This amounts in practice to using an indicator $I(P(Z_i) = 0)$ as an instrument. The benefits of observing values of Z_i at which no individuals receive treatment is discussed elsewhere for the homogeneous treatment setting, such as by Heckman (1990), Angrist and Imbens (1991), and Imbens and Angrist (1994). Treatment assignment with perfect compliance in the case of a randomized trial falls into this case, so such settings naturally satisfy uniform unordered monotonicity.

[6]Heckmand and Pinto's (2018) unordered monotonicity condition for multiple observed treatments, in the context of model 7, allows for the slope of the effect of instruments to be parallel to any decision margin, one example of which is shown in Panel A of Figure 1. Uniform unordered monotonicity thus allows for $1/(L+1)$ as many admissible slopes for the effects of instruments on treatment utilities.

switch from whatever variety (quality, etc) of college they otherwise would have attended to the nearby college. As in the case of homogeneous treatment monotonicity, it is intuitive that small violations are unlikely to be particularly harmful. De Chaisemartin (2017) formalizes this intuition and provides a weaker alternative to Imbens and Angrists' (1994) monotonicity assumption in the setting with homogeneous treatments. This section formalizes this intuition in the setting with heterogeneous treatments and provides a substantially weaker alternative to Condition 6 that is more likely to be satisfied in empirical applications.

I begin by defining component-treatment-specific versions of the compliers, defiers, always-takers, and never-takers described by Angrist, Imbens, and Rubin (1996) and others. Considering instrument comparison (z, z'), individual i is an ℓ complier with if $D_{i\ell}(z) = 1$ and $D_{i\ell}(z') = 0$, an ℓ defier if $D_{i\ell}(z) = 0$ and $D_{i\ell}(z') = 1$, an ℓ always-taker if $D_{i\ell}(z) = 1$ and $D_{i\ell}(z') = 1$, and an ℓ never-taker if $D_{i\ell}(z) = 0$ and $D_{i\ell}(z') = 0$. Let $C_{i\ell}(z, z') = 1$ if individual i is an ℓ complier ($C_{i\ell}(z, z') = 0$ otherwise), with the frequency of ℓ compliers given by $P_{c\ell}(z, z')$. Let $F_{i\ell}(z, z') = 1$ if i is an ℓ defier ($F_{i\ell}(z, z') = 0$ otherwise), with the frequency of ℓ defiers given by $P_{f\ell}(z, z')$. Allowing for the presence of defiers and assuming that conditions 1-3 hold, the IV estimand in (4) is equivalent to

$$\pi_{z,z'} = \frac{\sum_{\ell=1}^{L} P_{c\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|C_{i\ell}(z, z') = 1] - P_{f\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|F_{i\ell}(z, z') = 1]}{\sum_{\ell=1}^{L} P_{c\ell}(z, z') - P_{f\ell}(z, z')}.$$
(8)

Condition 6 sets $P(F_\ell) = 0$ in this equation, producing a weighted average across treatment varieties of complier treatment effects.

As an alternative to Condition 6, I propose a condition in which compliers exist for each component treatment that are able to cancel out the contributions of defiers to the IV estimand.

Condition 7 (Net Uniform Unordered Monotonicity). *For all $z, z' \in \mathcal{Z}$, there exists a $P_{c\ell}^{f}(z, z')$ share of the population uniquely defined by $C_{i\ell}^{f}(z, z') = 1$ such that $C_{i\ell}^{f}(z, z') = 1 \implies C_{i\ell}(z, z') = 1$, $P_{c\ell}^{f}(z, z') = P_{f\ell}(z, z')$, and $\mathbb{E}[Y_i(\ell) - Y_i(0)|C_{i\ell}^{f}(z, z') = 1] = \mathbb{E}[Y_i(\ell) - Y_i(0)|F_{i\ell}(z, z') = 1]$ for all ℓ.*

Condition 7 is effectively similar to Conditions 4 and 5 holding for many unobserved subgroups of the population, rather than for the population as a whole. It allows for some individuals to respond to instruments by switching out of component treatment ℓ as long as other individuals switch in to component treatment ℓ who happen to have overlapping treatment effects. Situations such as those pictured in Panel B of Figure 1 conceivably satisfy Condition 7 if treatment

effects are nearly homogeneous, while instruments that affect treatment preferences in ways that narrowly deviate from that pictured in Panel C likely satisfy this condition even with extreme treatment effect heterogeneity.

To establish the causal interpretation of the IV estimand under Condition 7, I define treatment ℓ surviving compliers as individuals for whom $C^s_{i\ell}(z, z') = 1$ where $C^s_{i\ell}(z, z') = I(C_{i\ell}(z, z') = 1, C^f_{i\ell}(z, z') \neq 1)$, who are present in the population with frequency $P^s_{c\ell}(z, z') = Pr(C^s_{i\ell}(z, z') = 1)$. The following theorem defines a the IV estimand as a positively-weighted average of a subset of treatments for a subset of individuals.

Theorem 3: *If Conditions 1-3 and 7 hold, the following component-weighted local average treatment effect is identified:*

$$\pi_{z,z'} = \sum_{\ell=1}^{L} \frac{P^s_{c\ell}(z, z')}{P^s_c(z, z')} \mathbb{E}[Y_i(\ell) - Y_i(0)|C^s_{i\ell}(z, z') = 1], \tag{9}$$

with $P^s_{c\ell}(z, z')/P^s_c(z, z') \geq 0$ for all ℓ, where $P^s_c(z, z') = \sum_{\ell=1}^{L} P^s_{c\ell}(z, z')$.

Proof. By the definition of $P^s_c(z, z')$, $P^s_{c\ell}(z, z')/P^s_c(z, z') \geq 0$ if $P^s_{c\ell}(z, z') \geq 0$ for all ℓ, which holds because $P^s_{c\ell}(z, z')$ is a probability bounded by $[0, 1]$. Under conditions 1-3 and 7, the reduced form effect of Z_i on Y_i evaluated at comparison (z, z') given in (2) and (3) can be expanded as

$$\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] = \sum_{\ell=1}^{L} P^s_{c\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|C^s_{i\ell}(z, z') = 1]$$

$$+ P^f_{c\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|C^f_{i\ell}(z, z') = 1] - P_{f\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|F_{i\ell}(z, z') = 1],$$

which reduces to

$$\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] = \sum_{\ell=1}^{L} P^s_{c\ell}(z, z')\mathbb{E}[Y_i(\ell) - Y_i(0)|C^s_{i\ell}(z, z') = 1]$$

because Condition 7 requires that $P^f_{c\ell}(z, z') = P_{f\ell}(z, z')$ and $\mathbb{E}[Y_i(\ell) - Y_i(0)|C^f_{i\ell}(z, z') = 1] = \mathbb{E}[Y_i(\ell) - Y_i(0)|F_{i\ell}(z, z') = 1]$ for all ℓ. Dividing both sides by $P(z) - P(z')$ yields the IV estimand, where $P(z) - P(z') = \sum_{\ell=1}^{L} P_{c\ell}(z, z') - P_{f\ell}(z, z')$ under Conditions 1-3, as shown in equation (8). Condition 7 implies that $\sum_{\ell=1}^{L} P_{c\ell}(z, z') - P_{f\ell}(z, z') = \sum_{\ell=1}^{L} P^s_{c\ell}(z, z') + P^f_{c\ell}(z, z') - P_{f\ell}(z, z')$, with $P^f_{c\ell}(z, z') = P_{f\ell}(z, z')$, such that $P(z) - P(z') = P^s_{c\ell}(z, z')$. It follows that the weighted

average of surviver-compliers' component treatment effects given by theorem 3 is obtained from the joint distribution of (Y_i, D_i, Z_i). $\square$

Several alternative assumptions are possible, which I leave to future study. For instance, using multiple instrument comparisons (z, z') and (z'', z''') allows for the possibility that defiers for one comparison are compliers for the other, which is sufficient for identification if the well-behaved comparison is at least as common in the population. Mogstad, Torgovitsky, and Walters (2021) describe such a condition in the case of multiple instruments with a homogeneous treatment. Additionally, in the case with multiple unobserved treatment varieties, it is possible that an individual who is a defier for one treatment is a complier for another (such individuals are the reason for complications beyond those for homogeneous treatments). If such individuals have identical treatment effects for both versions of treatment, they are effectively the same as always-takers, or, phrased differently, they cancel out their own defier contribution to the estimand with their complier contribution to the estimand.[7] The next section covers a setting when multiple instrumental variables are available to a researcher and only some satisfy the conditions described in this section.

3   Factor-Augmented Treatment Effects

In this section, I describe identification and estimation of wage returns to college using a novel factor-augmented treatment effects model that leverages instruments that violate the uniform treatment responses condition above (and therefore also violate uniform unordered monotonicity) at the cost of a homogeneous treatment effects assumption on the unobserved component treatments. The model imposes a factor structure on unobserved treatments and their effects, allowing for K instruments to work through $L \leq K$ weighted averages of unobserved component treatments to affect J lifecycle wage outcomes.[8] An intuitive interpretation of the model is that it produces data-driven lower-dimensional index variables for instruments, where each instrument's contributions to indexes are determined by its associated treatment effects across multiple outcomes. When $L = 1$, the model's treatment effect estimand for each outcome collapses to the 2SLS estimand. When $L = K$, the model's treatment effect estimand for each

---

[7]Selection on gains suggests some plausibility to this case. Individuals who are at the margin of two different treatment statuses may be marginal specifically because their treatment effects are similar.

[8]Though I consider a single outcome over many time periods in this application, the method described is not restricted to panel data.

outcome for unobserved factor ℓ collapses to the just-identified IV estimand for instrument $k = \ell$ for outcome j while including all other IVs as controls.

I assume that data is generated from the following parametric model of wages (Y_{ij} for time periods $j = 1, 2, ..., J$) and unobserved component treatments ($D_{i\ell}$ for treatment components $\ell = 1, 2, ..., L$),

$$Y_{ij} = X_i \alpha_j + \sum_{\ell=1}^{L} D_{i\ell} \lambda_{\ell j} + \xi_{ij}$$

$$D_{i\ell} = X_i \gamma_{x\ell} + \sum_{k=1}^{K} z_{ik} \gamma_{k\ell} + u_{i\ell} \qquad \forall \, \ell = 1, 2, ..., L. \tag{10}$$

X_i contains an $N \times R$ vector of controls, and each $N \times 1$ instrument $z_{ik} \subset Z_i$ has marginal effect $\gamma_{k\ell}$ on unobserved component treatment ℓ, while $\lambda_{\ell j}$ is the marginal effect of component treatment $D_{i\ell}$ as in the homogeneous treatment effects case in the preceding section. If we observed the component treatments, we could estimate model (10) via 2SLS as long as there are at least as many instruments as component treatments ($K \geq L$).

We do not observe $D_{i\ell}$ for $\ell = 1, 2, ..., L$, but we do observe D_i. It follows that we can estimate the following slightly unconventional IV model,

$$Y_{ij} = X_i \beta_j + \sum_{k=1}^{K} z_{ik} \gamma_k \pi_{kj} + \epsilon_{ij}$$

$$D_i = X_i \gamma_x + \sum_{k=1}^{K} z_{ik} \gamma_k + u_i, \tag{11}$$

where π_{kj} is equivalent to the just-identified IV estimand of instrument k for outcome j when including all other instruments as controls, and γ_k is the marginal effect of z_{ik} on D_i. Substituting the unobserved first stage into the unobserved second stage in model (10) gives an equivalent observed second stage of,

$$Y_{ij} = X_i \left(\alpha_j + \sum_{\ell=1}^{L} \gamma_{x\ell} \lambda_{\ell j} \right) + \sum_{k=1}^{K} z_{ik} \gamma_k \sum_{\ell=1}^{L} \frac{\gamma_{k\ell}}{\gamma_k} \lambda_{\ell j} + \left(\xi_{ij} + \sum_{\ell=1}^{L} u_{i\ell} \lambda_{\ell j} \right)$$

$$= X_i \beta_j + Z_i \gamma \theta \lambda_j + \epsilon_{ij}, \tag{12}$$

where $Z_i = [z_{i1}, z_{i2}, ..., z_{iK}]$, $\gamma = diag([\gamma_1, \gamma_2, ..., \gamma_K])$, $\theta = [\theta_1, \theta_2, ..., \theta_K]'$, with $\theta_k = [\theta_{k1}, \theta_{k2}, ..., \theta_{kL}]$ and $\theta_{k\ell} = \gamma_{k\ell}/\gamma_k$, and $\lambda_j = [\lambda_{1j}, \lambda_{2j}, ..., \lambda_{Lj}]'$. The reduced form coefficient on Z_i is composed

of γ (which is identified from the first stage) and a weighted average of unobserved treatments' effects, $\theta\lambda_j$.

Comparing equation (12) with the second stage in model (11) reveals that the just-identified IV estimand for outcome j using instrument k has a factor structure, $\pi_{kj} = \theta_k\lambda_j$, where instrument k places outcome-invariant weights, θ_k, on instrument-invariant component treatment effects, λ_j. The definition of $\theta_{k\ell}$ and the constraint $D_i = \sum_{\ell=1}^{L} D_{i\ell}$ implies that $\sum_{\ell=1}^{L} \theta_{k\ell} = 1$, so the weights instrument k places on unobserved factors must sum to one. The identification of treatment effects is generally achieved via additional restrictions on θ, some of which result in familiar estimands as we will see in the next subsection.

3.1 Identification of Component Effects

To establish identification conditions for the factor-augmented treatment effects model, I make use of multiple outcomes and arrange each π_{kj} into a $K \times J$ matrix π. The factor structure on instrument-specific treatment effects across many outcomes in (12) implies that

$$\pi = \theta\lambda, \tag{13}$$

where the $K \times L$ matrix of weights instruments place on component treatments, θ, is defined as above and the $L \times J$ matrix of component treatment effects across outcomes is given by $\lambda = [\lambda_1, \lambda_2, ..., \lambda_J]$. A barrier to identification, common to factor models, is that $\pi = \theta\lambda = \theta A A^{-1}\lambda$ for an arbitrary invertible $L \times L$ matrix A, so the factors and their weights are not separately identified even with many instruments and outcomes.

A has L^2 free elements, so we require at least L^2 restrictions to identify the model. Unlike many uses of factor models where the factors are effectively used as control variables, such as Bai (2009), it is necessary for the current method that we preserve the economic interpretation and scale of the treatment effect parameters. This adds some complications to identification, as many restrictions that will formally secure identification will nonetheless obscure the interpretation of model estimands. To emphasize this, I denote by (θ^r, λ^r) the component weights and component treatment effects that satisfy identification restrictions. Even in cases where all instruments are valid and π_j can be interpreted as a weighted average of unobserved treatments' effects (as in Theorem 4), it does not obviously follow that λ_j^r is a weighted average of unobserved treatments' effects for all restrictions that formally secure identification of the model.

I propose the following restrictions for the model (though others are possible). First, I impose the normalization

$$\theta^r = \begin{bmatrix} I_L \\ \theta^r_{L+1} \\ \vdots \\ \theta^r_K \end{bmatrix}, \tag{14}$$

which effectively defines the first L component treatments in terms of the first L instruments. This restriction will ensure interpretability of λ^r_ℓ by tying it closely to the just-identified IV estimand for instrument $k = \ell$. Because $D_i = \sum_{\ell=1}^L D_{i\ell}$ and the weights instruments place on component treatments are defined as $\theta_{k\ell} = \gamma_{k\ell}/\gamma_k$, it follows that $\sum_{\ell=1}^L \theta_{k\ell} = 1$. I therefore impose that

$$\sum_{\ell=1}^L \theta^r_{k\ell} = 1. \tag{15}$$

Together, I describe $L^2 + K - L$ restrictions, which is more than the formal requirement of L^2. These restrictions clearly define the treatment effect estimands λ^r while also imposing economically meaningful restrictions on the weights instruments place on unobserved treatments.

The relationships given in (13), (14), and (15) provide KJ, L^2, and $K - L$ equations, respectively, where the first L restrictions in (15) are redundant given the restrictions in (14). Meanwhile, there are KL parameters in θ^r and JL in λ^r. It follows that identification is achieved if

$$L^2 + KJ + K - L \geq JL + KL,$$

where solving for L when $K > $L yields

$$(K, J + 1) \geq L, \tag{16}$$

while the inequality is satisfied for any J if $K = L$. This expression defines the required number of outcomes and instruments for identification of a model with L unobserved component treatments.

It remains to be shown that the identified parameter λ^r defines a well-behaved treatment

effect. I show that λ^r is a weighted average of unobserved component treatments even when some instruments violate uniform treatment responses.

Theorem 4: *If the model in (10) describes the data-generating process with L unobserved component treatments where L is known and there are $K \geq L$ instruments, Conditions 1-4 hold for all comparisons (z_m, z_{-m}) and (z'_m, z_{-m}) in $\mathcal{Z}$ for all $m = 1, 2, ..., K$, and Condition 5 holds for z_k with $k \leq L$, the following weighted average component treatment effect is identified:*

$$\lambda^r_{kj} = \sum_{\ell=1}^{L} \theta_{k\ell} \lambda_{\ell j},$$

with $\theta_{k\ell} \geq 0$ for all ℓ.

Proof. Let Condition 5 hold for instrument k that satisfies $k \leq L$, with $P_\ell(z_k, z_{-k}) \geq P_\ell(z'_k, z_{-k})$ and $z_k > z'_k$ for all ℓ. If the model in (10) describes the DGP, then it follows that $\gamma_k > 0$ and $\gamma_{k\ell} > 0$ for all ℓ, such that $\theta_{k\ell} \geq 0$ for all ℓ. It also follows that the just-identified treatment effect estimands across all instruments for outcome j are given by $\pi_j = \theta \lambda_j$. The normalization given in (14) is equivalent to defining $\pi_j = \theta^r \lambda^r_j = \theta \theta_L^{-1} \theta_L \lambda_j$, where θ_L contains the first L rows of θ. $\lambda^r_j = \theta_L \lambda_j$ implies that $\lambda^r_{kj} = \sum_{\ell=1}^{L} \theta_{k\ell} \lambda_{\ell j}$ for $k \leq L$, which implies that the normalized component treatment effect defined in terms of instrument k, λ^r_{kj}, is identified. $\square$

The proof above requires that the number of unobserved treatment components be known to the researcher and be fewer than the number of instruments. It also requires homogeneous treatment effects for unobserved treatments' effects. These are sufficient conditions for the result, and perhaps could be weakened. Loosely speaking, if just-identified IV estimands for all instruments and all outcomes can be perfectly explained by constant weights on a lower-dimensional number of treatment effects, the result above will hold. I leave consideration of possible weaker conditions under which the same or similar result holds for large numbers of unobserved treatments and/or heterogeneous effects of the unobserved component treatments.

3.2   Estimation

This section describes estimation of the model. Substituting the factor structure on treatment effects using $\pi_{kj} = \theta^r_{k\ell} \lambda^r_{\ell j}$ in model (11), expanding the model to contain equations for all

outcomes, and rewriting it in matrix notation gives

$$Y_1 = X\beta_1 + Z\gamma\theta^r\lambda_1^r + \epsilon_1$$

$$Y_2 = X\beta_2 + Z\gamma\theta^r\lambda_2^r + \epsilon_2$$

$$\vdots \tag{17}$$

$$Y_J = X\beta_J + Z\gamma\theta^r\lambda_J^r + \epsilon_J$$

$$D = X\gamma_x + Z\gamma + u.$$

It is straightforward to estimate the above system via GMM with moments given by

$$\mathbb{E}[Z^*(Y_1 - X\beta_1 - Z\gamma\theta^r\lambda_1^r)] = 0$$

$$\mathbb{E}[Z^*(Y_2 - X\beta_2 - Z\gamma\theta^r\lambda_2^r)] = 0$$

$$\vdots \tag{18}$$

$$\mathbb{E}[Z^*(Y_J - X\beta_J - Z\gamma\theta^r\lambda_J^r)] = 0$$

$$\mathbb{E}[Z^*(D - X\gamma_x - Z\gamma)] = 0,$$

where $Z^* = [Z, X]'$.

It is helpful to consider two extreme cases to develop intuition for the estimator. When $L = 1$, the constraint in (15) implies that θ is a $K \times 1$ vector of ones, such that the structural second stage in (12) is equivalent to the second stage in (11) under the single treatment effect constraint $\pi_{kj} = \pi_j$, where $\pi_j = \lambda_j$ is the scalar IV-GMM estimand. When $L = K$, the constraint in (15) implies that θ is a $K \times K$ identity matrix, such that the structural second stage in (12) is identical to the second stage in (11) with $\lambda_j = \pi_j = [\pi_{1j}, \pi_{2j}, ..., \pi_{Kj}]'$. The factor-augmented treatment effects model thus nests IV-GMM and just-identified IV as special cases. For $L < K$, the model allows many instruments to identify a lower-dimensional number of effects for each outcome, with the goal of achieving efficiency gains similar to those that may be achieved by 2SLS or IV-GMM without requiring every instrument to satisfy uniform treatment responses.

3.3   Testing the Number of Component Treatments

The previous section describes estimation of the FATE model for L unobserved component treatments, where L is known. This section describes a statistical test for the null hypothesis

that the model is correctly specified, which means instruments satisfy independence, exclusion, and uniform treatment responses, with each of L unobserved treatments having homogeneous effects. If this test is rejected and the researcher is sure that the other conditions are satisfied, it is possible that the instruments work through more than L unobserved component treatments to drive outcomes.

If the above assumptions are satisfied and as $N \to \infty$, the moment conditions in (18) will hold exactly. In small samples, these moment conditions might not exactly hold, but we can test the hypothesis that they hold in the population by replacing them with within-sample estimates:

$$
\begin{aligned}
\sum_{i=1}^{N} Z_i^* (Y_{i1} - X_i \hat{\beta}_1 - Z_i \hat{\gamma} \hat{\theta}^r \hat{\lambda}_1^r) &= 0 \\
\sum_{i=1}^{N} Z_i^* (Y_{i2} - X_i \hat{\beta}_2 - Z_i \hat{\gamma} \hat{\theta}^r \hat{\lambda}_2^r) &= 0 \\
&\vdots \\
\sum_{i=1}^{N} Z_i^* (Y_{iJ} - X_i \hat{\beta}_J - Z_i \hat{\gamma} \hat{\theta}^r \hat{\lambda}_J^r) &= 0 \\
\sum_{i=1}^{N} Z_i^* (D_i - X_i \hat{\gamma}_x + Z_i \hat{\gamma}) &= 0,
\end{aligned}
\tag{19}
$$

The implied test is the Hansen (1982) J-test, which tests the null that the sample estimates of system (18) are satisfied with estimated values. When the assumed value L is $L = 1$, this is similar to the J-test of overidentifying restrictions for IV-GMM. When $L = K$, the test is trivially satisfied because the model is estimated to set the sample moments to zero. When L is between 1 and K, a large J-statistic indicates that an identification assumption is likely to be violated, which includes the possibility that the true number of unobserved treatments is larger than the assumed value of L.

4   Data

I estimate returns to college with the factor-augmented treatment effects model described above using data from the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 follows a group of 12,686 individuals who were between the ages of 14 and 22 in 1979. Respondents were surveyed annually for the first 15 years and biennially afterwards. Survey responses include

information on years of education, earnings, geography of residence, and a wealth of auxiliary information. In addition to meeting the requirements of my model of containing multiple instruments and outcomes, this dataset is attractive because it has been used extensively in the literature on returns to education, which helps to emphasize my methodological contributions by holding fixed some other potential explanations for similarities and differences in results. In particular, I use the same sample of white males that was used by Carneiro, Heckman, and Vytlacil (2011), matched with longitudinal wage data constructed according to the procedure described in Ashworth, Hotz, Maurel, and Ransom (2017).

The composite treatment of interest is college attendance as of age 25. Estimating my model requires multiple outcomes and instruments, as shown in expression (16). The longitudinal nature of the NLSY allows me to set $j = t$ for $t = 26, \ldots, 50$ and rely on wage outcomes over the lifecycle to identify the model.[9]

I use several instruments that are established in the literature on returns to education. They are average wages in the county of residence at age 17, an indicator for the presence of a college in county of residence at age 14, and average tuition of public colleges in the county of residence at age 17.[10] That each of these variables contributes to the college attendance decision is intuitive as well as testable. I discuss the validity of each instrument separately in terms of independence, exclusion, and uniform treatment responses.

Local wages have been used as explanatory variables for educational decisions by Cameron and Heckman (1998), Cameron and Heckman (2001), Cameron and Taber (2004), and Carneiro, Heckman, and Vytlacil (2011). Following Carneiro, Heckman, and Vytlacil (2011), I use the average values of this variable at age 17 at the county level as an instrument for college attendance. Because local labor market conditions may contribute to outcomes through channels other than education, I control for permanent wages in the county of residence at age 17, permanent unemployment in the state of residence at age 17, wages in the county of residence in 1991, and unemployment in the state of residence in 1991 (to control for persistent early life-cycle labor market shocks). It follows that I am effectively using temporal deviations from permanent local wages at age 17 as an instrument. The primary channel through which I expect this instrument to affect college attendance is by making the outside option of directly entering

---

[9]Researchers considering using this method in other applications should rarely be deterred by a lack of outcomes. Even if economically meaningful outcomes are sparse in a given dataset, it is possible to estimate the model using outcomes that lack obvious economic significance but nonetheless contribute to identification.

[10]The unemployment rate in the state of residence at age 17 is another commonly used instrument. I have opted to exclude it because it is extremely weak compared to the other instruments.

the workforce more attractive, without altering individuals' preferences over types of college, suggesting that it is particularly likely to satisfy uniform treatment responses (and uniform unordered monotonicity). It is also possible that this instrument affects college decisions by relaxing credit constraints, allowing individuals who would not have attended college to do so, and allowing individuals who would have attended college to attend more expensive colleges, though this explanation is at odds with the empirical finding of a negative first stage coefficient for this instrument.

The presence of a college in the county of residence at age 14 has been used as an instrument for education by Card (1993); Kane and Rouse (1995); Kling (2001); Currie and Moretti (2003); Cameron and Taber (2004); Carneiro, Heckman, and Vytlacil (2011), and Mogstad, Torgovitsky, and Walters (2021). I expect the presence of colleges in the county of residence to primarily influence education by encouraging students to attend a nearby college through a cost reduction (both pecuniary and psychic) mechanism. If this sort of cost reduction is more important for students inclined toward particular types of college education, such as majors (Altonji, Blom, and Meghir, 2012) or quality (Black and Smith, 2006), then this IV will identify the effects of such types of education, and may violate uniform treatment responses if individuals prefer nearby colleges of an otherwise non-preferred types to distant colleges of more-preferred types. For instance, if college returns are only a function of quality and some individuals induced into college by this instrument have a preference ordering of nearby high-quality college $\succ$ nearby low-quality college $\succ$ distant high-quality college, we might expect this instrument to partially represent the difference between returns to low-quality college and high-quality college (driven by such individuals whose nearby college is low-quality). Independence and exclusion for this instrument are threatened by correlation between college presence and students' ability, as noted by Carneiro and Heckman (2002) and Cameron and Taber (2004). To address this, I control for a measure of ability (the Armed Forces Qualification Test, AFQT), following Carneiro, Heckman, and Vytlacil (2011).

Local tuition costs have been used as an instrument for education by Kane and Rouse (1995), Cameron and Heckman (1998), Cameron and Heckman (2001), and Carneiro, Heckman, and Vytlacil (2011). This is similar to local labor market conditions in that it may be correlated with unobserved ability, such as if expensive colleges produce different externalities for the local community than cheap colleges. I control for mother's years of education, permanent local labor market conditions, and AFQT to address this concern. I view this instrument as a pecuniary

cost-shifter, which produces particular concerns regarding uniform treatment responses. Cost-shifter instruments produce income effects, which I expect to induce exposed individuals to obtain leisure, potentially by putting less effort into their studies. If low-effort college has lower returns than high-effort college and this instrument shifts sufficiently many individuals from high-effort college to low-effort college, it will place negative weights on high-return college effects, yielding a downward-biased estimate. For this reason, I choose this instrument to not identify an unobserved treatment variety, explicitly allowing it to contribute to estimation while taking on negative weights.

I use several variables as controls, which is common in the literature. They are number of siblings, permanent local unemployment in the state of residence at age 17 (average from years 1979-2000), permanent local log wages in the state of residence at age 17 (average from years 1979-2000), a binary indicator for urban residence at age 14, average wages in the county of residence in 1991, average unemployment in the county of residence in 1991, year of birth indicators, mother's years of education, and the Armed Forces Qualifying Test score. Statistics on the variables I use are shown in Table 1.

Table 1: Summary Statistics

	Mean (1)	Standard deviation (2)
Mean log wages, ages 26-30	2.084	0.463
Mean log wages, ages 31-35	2.172	0.533
Mean log wages, ages 36-40	2.201	0.575
Mean log wages, ages 41-45	2.198	0.593
Mean log wages, ages 46-50	2.156	0.618
Attended college	0.495	0.500
Number of siblings	2.927	1.909
Permanent local unemployment at age 17	6.251	0.986
Permanent local log wages at age 17	10.283	0.188
Urban residence at age 14	0.744	0.436
Local unemployment rate in 1991	6.810	1.267
Local log wages in 1991	10.293	0.165
Year of birth	1959.759	2.340
Mother's years of schooling	12.102	2.335
Corrected AFQT score	0.449	0.952
Local log wages at age 17	10.276	0.164
Nearby four-year college at age 14	0.525	0.500
Local tuition at age 17	21.568	7.981
Sample size	1747	

Notes: Means and standard deviations for white males in the National Longitudinal Survey of Youth 1979 sample. AFQT corrected for years of schooling at the timing of test-taking. Permanent log wages and unemployment are calculated as average values from 1973 to 2000 for the county of residence at age 17.

5 Results

I begin by showing IV treatment effects estimated via GMM for just-identified IV excluding each instrument individually and including the others as controls, as well as IV-GMM excluding all IVs from outcome equations estimation. These results are in Table 2, and are shown graphically in Figure 2. I divide all coefficients by four to express returns in annual terms. This table reports F-statistics for each model's first stage, as well as p-values for the Hansen J test of overidentifying restrictions for the model excluding all IVs from outcome equations.

The first stage F-stats for local wages, nearby college, and local tuition when used alone are 18.7, 5.5, and 2.5, respectively, with the F-stat for all of them is 11.1. Two of these fall below the commonly-invoked heuristic threshold of 10, while also falling short of the corresponding statistics calculated in a similar exercise by Mogstad, Torgovitsky, and Walters (2021), likely due to correlation between IVs and my inclusion of a third that they do not use (local tuition). The particularly weak instruments, nearby college and local tuition, have intuitive (positive and negative, respectively) signs on college attendance. Angrist and Kolesár (2021) show that sign-screening reduces bias from weak IVs, so I assume that the instruments are immune to weak IV problems.[11]

The local earnings instrument consistently estimates relatively high returns to college, which grow larger later in the lifecycle. I argued above that this instrument is particularly likely to satisfy net uniform unordered monotonicity, suggesting that these estimates likely place positive weight on returns to different types of college for different types of students. It follows that college may be a better investment than is sometimes suggested in the literature, with many estimates suggesting returns in the proximity of 10% (Card, 2001). It is possible that other instruments are valid as well and this instrument places particularly high weight on high-return versions of college. This could occur if particularly money-motivated individuals, who are inclined toward high-paying versions of college, are particularly sensitive to local variation in earnings when they make their education decisions. In either case, the returns implied by this instrument suggest that extremely high return versions of college exist for some individuals, which could suggest scope for substantial social gains by finding ways to induce more individuals to attend such versions of college.

---

[11]Because I would have used IVs common to the literature (and this dataset) regardless of their first stages, it is inaccurate to say that I sign-screened these. It is likely, however, that prior researchers sign-screened these IVs when introducing them, so the sign-screening argument holds.

The nearby college and local tuition instruments both run some risk of inducing students to shift out of high-effort college and toward low-effort college, as they both can be seen as cost-shifters which likely have income effects. It follows that the estimates for these instruments should be lower than true returns if low-effort college has lower returns than high-effort college, which is consistent with these point estimates being consistently lower throughout the lifecycle than those identified using the local earnings IV. That said, the standard errors on estimates from both of these IVs are quite large, which makes it difficult to draw sharp conclusions.

Excluding all IVs from outcome equations, as in model (4), yields substantial efficiency gains as evidenced by the substantially smaller standard errors. However, it is possible that these estimates inherit some of the bias from net uniform unordered monotonicity violations of the nearby college and local tuition instruments. Notably, the point estimates are almost always smaller when using all IVs to estimate treatment effects relative to only using the local earnings IV, which is consistent with the possibility of these estimates shifting some individuals out of high-return college and into low-return college.
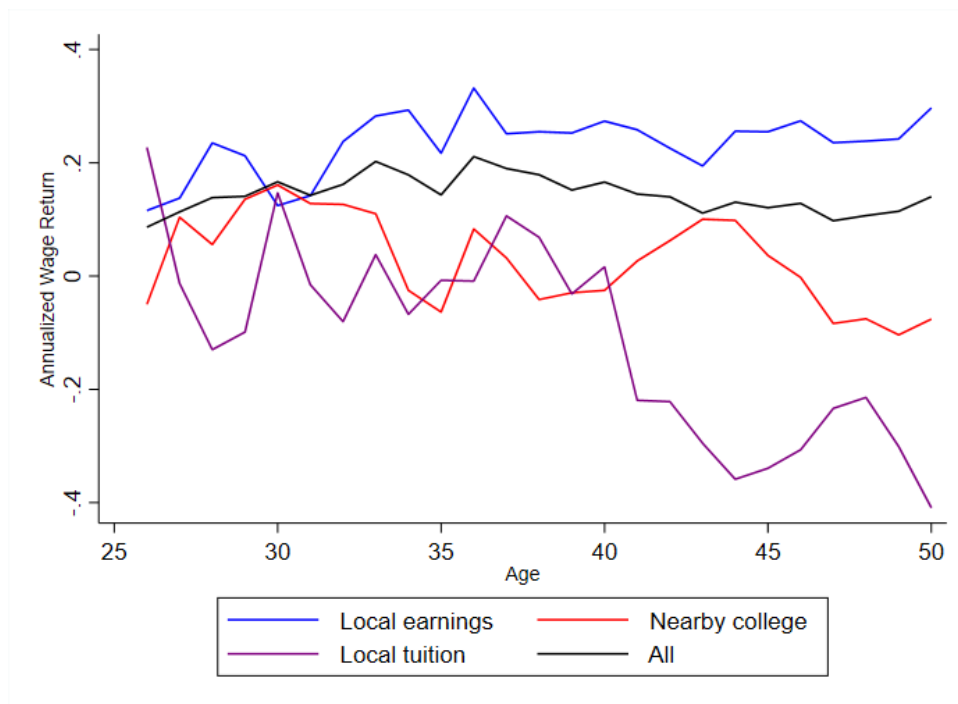
Figure 2: IV-GMM Estimates of Lifecycle Wage Returns to College

Notes: This graph depicts IV-GMM point estimates of the effects of college attendance on lifecycle log wages for each excluded instrument listed, as well as the case where all instruments are excluded. See Table 2 for point estimates and standard errors.

I also estimate the factor-augmented treatment effects model in (17) with results shown in Table 3, defining the first unobserved treatment in terms of the local earnings IV. From

28

Table 2: IV-GMM Estimates of Lifecycle Returns to College, by IV

| | 2nd Stage Excluding: | | | | | | | |
| | Local earnings | | Nearby college | | Local tuition | | All | |
	(1)		(2)		(3)		(4)	
Log wage, age 26	0.116	(0.086)	-0.050	(0.145)	0.227	(0.226)	0.087	(0.052)
Log wage, age 27	0.138	(0.074)	0.104	(0.142)	-0.013	(0.194)	0.113	(0.053)
Log wage, age 28	0.235	(0.087)	0.056	(0.131)	-0.130	(0.216)	0.139	(0.054)
Log wage, age 29	0.213	(0.083)	0.136	(0.139)	-0.099	(0.199)	0.141	(0.055)
Log wage, age 30	0.125	(0.092)	0.161	(0.136)	0.147	(0.210)	0.167	(0.056)
Log wage, age 31	0.142	(0.091)	0.128	(0.133)	-0.015	(0.200)	0.143	(0.054)
Log wage, age 32	0.238	(0.082)	0.127	(0.134)	-0.080	(0.211)	0.162	(0.054)
Log wage, age 33	0.283	(0.094)	0.110	(0.142)	0.038	(0.200)	0.203	(0.060)
Log wage, age 34	0.293	(0.102)	-0.025	(0.144)	-0.067	(0.212)	0.179	(0.066)
Log wage, age 35	0.217	(0.118)	-0.063	(0.151)	-0.007	(0.203)	0.143	(0.069)
Log wage, age 36	0.332	(0.109)	0.083	(0.141)	-0.009	(0.199)	0.211	(0.065)
Log wage, age 37	0.251	(0.099)	0.032	(0.139)	0.106	(0.190)	0.190	(0.062)
Log wage, age 38	0.255	(0.099)	-0.041	(0.152)	0.068	(0.188)	0.179	(0.064)
Log wage, age 39	0.253	(0.091)	-0.029	(0.149)	-0.031	(0.203)	0.152	(0.060)
Log wage, age 40	0.274	(0.094)	-0.025	(0.148)	0.016	(0.188)	0.166	(0.060)
Log wage, age 41	0.259	(0.095)	0.027	(0.149)	-0.219	(0.272)	0.145	(0.059)
Log wage, age 42	0.226	(0.093)	0.063	(0.151)	-0.222	(0.259)	0.140	(0.060)
Log wage, age 43	0.195	(0.087)	0.100	(0.151)	-0.295	(0.299)	0.111	(0.059)
Log wage, age 44	0.256	(0.089)	0.098	(0.145)	-0.359	(0.334)	0.131	(0.058)
Log wage, age 45	0.255	(0.089)	0.037	(0.147)	-0.339	(0.332)	0.121	(0.057)
Log wage, age 46	0.274	(0.106)	-0.002	(0.148)	-0.307	(0.319)	0.128	(0.059)
Log wage, age 47	0.236	(0.102)	-0.084	(0.158)	-0.233	(0.282)	0.098	(0.059)
Log wage, age 48	0.238	(0.108)	-0.075	(0.157)	-0.214	(0.269)	0.107	(0.060)
Log wage, age 49	0.242	(0.094)	-0.104	(0.167)	-0.301	(0.331)	0.115	(0.064)
Log wage, age 50	0.297	(0.108)	-0.076	(0.168)	-0.409	(0.383)	0.140	(0.069)
First Stage F-stat	18.709		5.220		2.472		11.126	
Over-id p-value							0.339	
Sample size	1747		1747		1747		1747	

Notes: Two-step IV generalized method of moments estimates of the effects of college attendance on log wages for white men in the NLSY79 over the lifecycle. Age-specific estimates for models 1-3 obtained excluding a single instrument and including all others as controls along with the other control variables listed in Table 1. Model 4 excludes all instruments from outcome equations. College returns are annualized by dividing the college attendance coefficient by four. Robust standard errors in parentheses.

Table 3: GMM Estimates of FATE Model of Lifecycle Returns to College

	Local earnings component (1)		Nearby college component (2)	
Component effects (λ^r)				
Log wage, age 26	0.076	(0.074)	0.089	(0.055)
Log wage, age 27	0.143	(0.073)	0.105	(0.058)
Log wage, age 28	0.234	(0.085)	0.136	(0.076)
Log wage, age 29	0.214	(0.082)	0.137	(0.068)
Log wage, age 30	0.130	(0.086)	0.134	(0.059)
Log wage, age 31	0.136	(0.085)	0.099	(0.060)
Log wage, age 32	0.227	(0.078)	0.150	(0.067)
Log wage, age 33	0.259	(0.088)	0.197	(0.070)
Log wage, age 34	0.278	(0.096)	0.179	(0.079)
Log wage, age 35	0.173	(0.097)	0.121	(0.079)
Log wage, age 36	0.301	(0.097)	0.209	(0.080)
Log wage, age 37	0.227	(0.089)	0.178	(0.069)
Log wage, age 38	0.234	(0.091)	0.171	(0.073)
Log wage, age 39	0.231	(0.085)	0.137	(0.077)
Log wage, age 40	0.244	(0.087)	0.162	(0.074)
Log wage, age 41	0.244	(0.088)	0.122	(0.082)
Log wage, age 42	0.221	(0.086)	0.114	(0.078)
Log wage, age 43	0.199	(0.082)	0.081	(0.081)
Log wage, age 44	0.245	(0.085)	0.097	(0.092)
Log wage, age 45	0.238	(0.083)	0.091	(0.091)
Log wage, age 46	0.261	(0.100)	0.111	(0.095)
Log wage, age 47	0.218	(0.096)	0.085	(0.090)
Log wage, age 48	0.210	(0.102)	0.083	(0.089)
Log wage, age 49	0.204	(0.085)	0.055	(0.100)
Log wage, age 50	0.259	(0.097)	0.075	(0.114)
Component weights (θ^r)				
Local earnings	1	(.)	0	(.)
Nearby college	0	(.)	1	(.)
Local tuition	-2.498	(2.333)	3.498	(2.333)
Value of criterion function (Q)	0.009			
Over-id p-value	0.904			
Sample size	1747			

Notes: Two-step generalized method of moments estimates of the factor-augmented treatment effect model of effects of college attendance on log wages for white men in the NLSY79. Component-defining instruments have weights normalized to 1 on respective components. College returns are annualized by dividing the college component effects by four. Robust standard errors in parentheses.
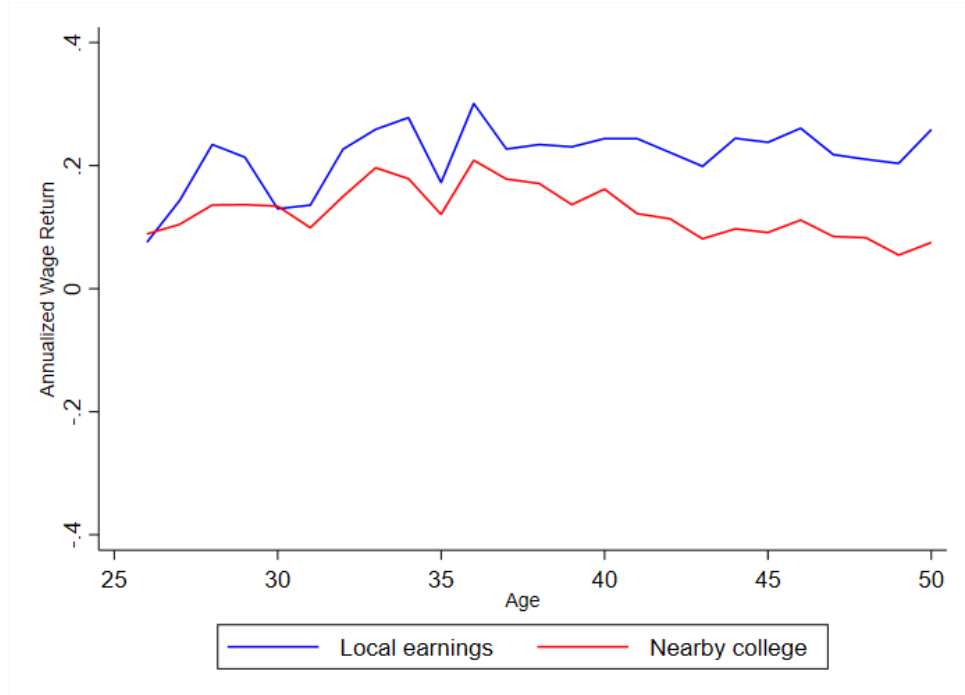
Figure 3: GMM Estimates of Lifecycle Wage Returns to Unobserved College Components

Notes: This graph depicts GMM point estimates of treatment effects from the factor-augmented treatment effects model on lifecycle log wages with components defined in terms of listed instruments. See Table 3 for point estimates and standard errors.

theorem 4, this will provide consistent estimates of a positively-weighted average of unobserved treatments' effects if the local earnings IV satisfies independence, exclusion, relevance, uniform treatment responses, and I properly specify the number of unobserved treatment varieties. I define the second component in terms of the nearby college instrument. If nearby college satisfies the required conditions, these estimates are consistent for a weighted average of effects of unobserved component treatments, whereas the estimates in column (2) are nuisance parameters if local earnings is a valid IV and nearby college is not. Because of the concerns with nearby college listed above, I opt to view the associated estimates as nuisance parameters, but the reader is free to make their own decision.

The F-stat of 18.7 in Table 2 suggests relevance is satisfied for local earnings, while the p-value on the test of overidentifying restrictions of 0.339 in Table 3 means that we fail to reject the null that $L = 2$ component treatments are sufficient to explain the data. The FATE model allows extra instruments to contribute to estimation with negative weights, which we see in the case of local tuition placing weights of -2.5 and 3.5 on the local earnings and nearby college treatment components, respectively. As in the IV-GMM results, the local earnings component treatment has substantially higher point estimates through the lifecycle than the nearby college

component. Interestingly, the local earnings component point estimates visually look (both in the table and in Figure 3) like those of just-identified IV using the local earnings instrument, while the nearby college component effects look like those of IV-GMM using all instruments for estimation of treatment effects.

Perhaps for similar reasons, the standard errors for the (potentially uninterpretable) nearby college component treatment effect estimates are dramatically smaller than those for the nearby college instrument in the standard IV-GMM results in Table 2, while the standard errors for the local earnings component are essentially unchanged (with trivial reductions). Possible intuition for this is that the FATE model was able to rationalize the just-identified IV estimates of the local tuition instrument by altering the nearby college component treatment effects, without needing to alter the local earnings component treatment effects very much relative to the just-identified IV estimates. It follows that the local tuition IV helps to estimate the nearby college component treatment effect parameters without lending much to the local earnings component effect parameter estimation. This is economically reasonable insofar as both nearby college and local tuition may work through a cost-shifter channel, inducing similar types of individuals into similar types of college (with local tuition perhaps doing so more extremely, as evidenced by it placing a weight of 3.5 on the nearby college treatment component).

6   Conclusion

A large literature describes implications of unmodeled treatment effect heterogeneity for estimation of treatment effects. This paper addresses the related problem of unmodeled treatment heterogeneity for treatment effect estimation. It shows the validity of standard methods when instruments shift types of individuals into (and not between) diverse varieties of treatment on net, as well as providing an alternative factor-augmented treatment effects model that allows instruments that shift individuals between types of treatment to contribute to treatment effect estimation. I estimate the FATE model in the context of returns to college using the NLSY79, finding particularly high returns to college that range between 7.6% and 30.1% over the lifecycle for the component treatment associated with the instrumental variable least-likely to shift individuals between treatments (local earnings at age 17).

For empirical applications of instrumental variables methods that make use of single instruments, the most relevant contribution of this paper is the assumption of net uniform unordered

monotonicity, which allows for unmodeled versions of treatment that affect outcomes differently to reside in the residuals of outcome equations (formally violating the exclusion restriction for modeled composite treatments). As in the case of discussions of the monotonicity condition of Imbens and Angrist (1994), I recommend that empirical researchers justify the net uniform unordered monotonicity assumption when possible, or discuss its implications for their estimates when it may not be valid. For instance, an instrument that violates net uniform monotonicity will be only minimally biased if there is minimal heterogeneity in effects between different versions of treatment. Furthermore, treatment effects estimated using a policy-relevant instrument may be informative for changes in that particular policy even if it violates net uniform unordered monotonicity. For instance, the returns to college estimated using local tuition variation may be relevant for describing the likely effects on earnings of increases in subsidization of college, if such counterfactual subsidies move individuals between versions of treatments the same way the instrument does.

For empirical applications involving multiple instruments, the factor-augmented treatment effects model developed here provides an alternative to 2SLS or other standard methods for pooling variation from multiple instruments into a single treatment effect estimate. The FATE method is likely to be particularly useful when a particularly credible instrument, such as randomized assignment to treatment, is relatively weak, while some less credible instruments that may violate net uniform unordered monotonicity are relatively strong.

An attractive option for this is to construct a binary indicator set to one if an instrument takes the value with the lowest propensity of treatment, and use this constructed instrument to define a component treatment in the FATE model while allowing the continuous (or other functional form) versions of the instrument to place negative weights on unobserved versions of treatment. In the extreme case where the treatment propensity is zero at a particular value of the instrument, it is impossible for there to be any violations of net uniform unordered monotonicity, and such violations are likely to still be rare when the treatment propensity is close to zero. Such an exercise can be implemented in applications with a single strong instrument that might not satisfy net uniform unordered monotonicity when used as a single continuous variable, but which retains sufficient strength for treatment effect estimation when split into multiple instruments.

The sufficient conditions given above for the factor-augmented treatment effects model to identify convex averages of unobserved treatments' effects do not allow for individual-level treat-

ment effect heterogeneity. Some heterogeneity in treatment effects between individuals can be attributed to heterogeneity in chosen treatments, but it is plausible that a small number of unobserved treatments (two in the application) is insufficient to fully characterize the relevant heterogeneity. Future theoretical work on the factor-augmented treatment effects estimator to determine its validity in the presence of treatment effect heterogeneity and/or more unobserved component treatments than instruments may be helpful in increasing its usefulness for applications.

In terms of the application to returns to college, I find evidence that effects of college on wages are potentially higher than are commonly found in the literature. Estimating returns as driven by local wages at age 17, which is unlikely to alter students' preferences over different types of college, gives treatment effects from 12.5% to 33.2% (IV-GMM) or 7.6% to 30.1% (FATE), which are generally much higher than the returns around 10% that are commonly found in the literature. Insofar as variation in returns between instruments is due to instruments inducing individuals into different types of treatments, these results suggest substantial potential for gains (in terms of wage increases) from policies that move students into higher-return types of college. Insofar as students' current choices are driven in part by information frictions regarding pecuniary returns to college or students internalize nonpecuniary benefits from types of college more than pecuniary benefits (for instance due to taxes), aggregate welfare gains are likely as well.

The distinction between heterogeneous treatment effects and heterogeneous treatments is central to these policy implications. If some students have high returns and others have low returns, income-maximizing (or similar) policies are those that induce students to attend college if and only if they have positive returns. If, however, different versions of college have different effects (in line with a very large literature on college majors and quality), then optimal policy will induce particular students into well-matched versions of college. The finding, such as in Cunha, Heckman, and Navarro (2005) or Carneiro, Heckman, and Vytlacil (2011), that a large minority of individuals have negative returns to college is potentially due to students failing to choose versions of college that best-improve their human capital. Further work to distinguish between heterogeneous effects of treatments and effects of heterogeneous treatments is essential to inform policies that seek to improve educational allocations.

References

ALTONJI, J. G., E. BLOM, AND C. MEGHIR (2012): "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annu. Rev. Econ.*, 4(1), 185–223.

ANGRIST, J., AND G. IMBENS (1991): "Sources of identifying information in evaluation models,"
.

ANGRIST, J., AND M. KOLESÁR (2021): "One instrument to rule them all: The bias and coverage of just-id iv," Discussion paper, National Bureau of Economic Research.

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(434), 444–455.

ASHWORTH, J., V. J. HOTZ, A. MAUREL, AND T. RANSOM (2017): "Changes across cohorts in wage returns to schooling and early work experiences," .

BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77(4), 1229–1279.

BLACK, D. A., AND J. A. SMITH (2006): "Estimating the returns to college quality with multiple proxies for quality," *Journal of labor Economics*, 24(3), 701–728.

CAMERON, S. V., AND J. J. HECKMAN (1998): "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political economy*, 106(2), 262–333.

——— (2001): "The dynamics of educational attainment for black, hispanic, and white males," *Journal of political Economy*, 109(3), 455–499.

CAMERON, S. V., AND C. TABER (2004): "Estimation of educational borrowing constraints using returns to schooling," *Journal of political Economy*, 112(1), 132–182.

CARD, D. (1993): "Using geographic variation in college proximity to estimate the return to schooling," *NBER working paper*, (w4483).

——— (2001): "Estimating the return to schooling: Progress on some persistent econometric problems," *Econometrica*, 69(5), 1127–1160.

CARNEIRO, P., AND J. J. HECKMAN (2002): "The evidence on credit constraints in post-secondary schooling," *The Economic Journal*, 112(482), 705–734.

CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): "Estimating marginal returns to education," *American Economic Review*, 101(6), 2754–81.

CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): "Separating uncertainty from heterogeneity in life cycle earnings," *oxford Economic papers*, 57(2), 191–261.

CURRIE, J., AND E. MORETTI (2003): "Mother's education and the intergenerational transmission of human capital: Evidence from college openings," *The Quarterly journal of economics*, 118(4), 1495–1532.

DE CHAISEMARTIN, C. (2017): "Tolerating defiance? Local average treatment effects without monotonicity," *Quantitative Economics*, 8(2), 367–396.

DILLON, E. W., AND J. A. SMITH (2020): "The consequences of academic match between students and colleges," *Journal of Human Resources*, 55(3), 767–808.

HANSEN, K. T., J. J. HECKMAN, AND K. J. MULLEN (2004): "The effect of schooling and ability on achievement test scores," *Journal of econometrics*, 121(1-2), 39–98.

HANSEN, L. P. (1982): "Large sample properties of generalized method of moments estimators," *Econometrica: Journal of the econometric society*, pp. 1029–1054.

HASTINGS, J. S., C. A. NEILSON, AND S. D. ZIMMERMAN (2013): "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile," Discussion paper, National Bureau of Economic Research.

HECKMAN, J. (1990): "Varieties of Selection Bias," *The American Economic Review*, 80(2), 313–318.

HECKMAN, J. J., AND R. PINTO (2018): "Unordered monotonicity," *Econometrica*, 86(1), 1–35.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor economics*, 24(3), 411–482.

HULL, P. (2018): "Isolateing: Identifying counterfactual-specific treatment effects with cross-stratum comparisons," *Available at SSRN 2705108.*

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica,* 62(2), 467–475.

JIANG, X. (2019): "Women in STEM: Ability, Preference, and Value," *The Ohio State University, Department of Economics Working Paper.*

KANE, T. J., AND C. E. ROUSE (1995): "Labor-Market Returns to Two-and Four-Year College," *American Economic Review,* 85(3), 600–614.

KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics,* 131(3), 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(4), 1795–1848.

KLING, J. R. (2001): "Interpreting instrumental variables estimates of the returns to schooling," *Journal of Business & Economic Statistics,* 19(3), 358–364.

MOGSTAD, M., A. TORGOVITSKY, AND C. R. WALTERS (2021): "The causal interpretation of two-stage least squares with multiple instrumental variables," *American Economic Review,* 111(11), 3663–98.

MOUNTJOY, J. (2022): "Community colleges and upward mobility," *American Economic Review,* 112(8), 2580–2630.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology,* 66(5), 688.

——— (1990): "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science,* 5(4), 472–480.

SARZOSA, M., AND S. URZÚA (2015): "Bullying among adolescents: the role of cognitive and non-cognitive skills," Discussion paper, National Bureau of Economic Research.

Appendix A: Factor Augmented Treatment Effects Example

In this section, I walk through an intuitive three-step estimation procedure for an extreme case of the model where there are three instruments ($K = 3$), two unobserved treatment varieties ($L = 2$), and J outcomes, where $J \geq 2$. The first step is to estimate just-identified IV treatment effects for each instrument for all outcomes using standard methods, yielding $\hat{\pi}$. The second step infers $(\hat{\lambda}_j, \hat{\theta})$ from $\hat{\pi}_j$ for a single outcome. The third step takes $(\hat{\theta})$ as given (similarly to 2SLS implicitly imposing ex ante that all instruments weight a single common version of treatment), yielding $\hat{\lambda}_{-j}$ for all outcomes other than j.

Considering only outcome j, the restrictions in (13), (14), and (15) set

$$\hat{\pi}_{1,j} = \hat{\lambda}_{1,j}$$

$$\hat{\pi}_{2,j} = \hat{\lambda}_{2,j}$$

$$\hat{\pi}_{3,j} = \hat{\theta}_{3,1}\hat{\lambda}_{1,j} + (1 - \hat{\theta}_{3,1})\hat{\lambda}_{2,j},$$

which is a system of three equations with three unknowns. It follows that the unobserved treatments' effects and their outcome-invariant weights are exactly identified.

The weight parameters in $\hat{\theta}$ estiamted using outcome j can be taken as given for all other outcomes. It follows that estimates of $(\lambda_{1,j'}, \lambda_{2,j'})$ can then be obtained as the two parameters that best fit the system of three equations,

$$\hat{\pi}_{1,j'} = \hat{\lambda}_{1,j'}$$

$$\hat{\pi}_{2,j'} = \hat{\lambda}_{2,j'}$$

$$\hat{\pi}_{3,j'} = \hat{\theta}_{3,1}\hat{\lambda}_{1,j'} + (1 - \hat{\theta}_{3,1})\hat{\lambda}_{2,j'},$$

for any outcome $j' \neq j$. When $N \to \infty$ and there is no model misspecification, this seemingly overidentified system of equations will hold exactly.

In the example given, only a single outcome is used to estimate weights, and they are taken as given for all other outcomes. In the GMM estimator described in Section 3, all outcomes contribute to estimation of weights. Intuitively, in the case where $J \to \infty$, each outcome's contribution to the estimation of weights goes to zero, and the simultaneously-estimated model takes on more of the intuition described in the final step here. Similarly, if a researcher has access to auxiliary outcomes that are not of substantive interest to the research question, it is possible

to construct an estimator that estimates instrument weights on unobserved treatments using the auxiliary treatments only, while taking the weights as given for the outcomes of interest.[12] Such an estimator could be implemented in steps as described here, or could be implemented via imposing a GMM weighting matrix with arbitrarily high weights on auxiliary outcomes, forcing the estimator to choose θ that prioritizes the fit for those particular moments, leaving other moments to take the implied $\hat{\theta}$ as given.

---

[12]The use of auxiliary outcomes is similar to the use of test equations by Cunha, Heckman, and Navarro (2005) and Sarzosa and Urzúa (2015), among others, when estimating models that control for unobserved heterogeneity in ability.