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

Clint Harris†

July 22, 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. 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. I apply the method to lifecycle returns to college, estimating "high-type" college returns that range from 10% to 26% over the life cycle and "low-type" returns that range from 2% to 14%. My findings emphasize 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: Instrumental Variables, Local Average Treatment Effects, Unobserved Heterogeneity, Factor Models, Returns to Education

*I thank Haiqing Zhao, who was a coauthor on a related paper (Estimating Returns to Unobserved Education, 2022). I also thank Xiaoxiao Li, Anna Mikusheva, Jeff Smith, Chris Taber, and Matt Wiswall, as well as seminar participants at the Midwest Economics Association Annual Meeting, the Southern Economic Association Annual Meeting, and the University of Wisconsin for helpful comments. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

†University of Wisconsin, 330 N Orchard Street, Madison, WI 53715, USA; email: clint.harris@wisc.edu

1 Introduction

Many treatment variables used in empirical economics are sums of otherwise unobserved variables. For instance, years of education encompass educational time investments across subjects and levels. Years of work experience encompass time investments across a variety of tasks which develop different skills. 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. In some cases, the estimated composite treatment effect may (nearly) 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 identification strategy 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.

This paper estimates effects of education on wages using a method that is robust to the presence of multiple types of education, but does not require observation of these types of treatment. The advantage of the method is to identify effects of component treatments that affect outcomes differently without necessarily needing to observe the components directly or determine them ex ante. It is built upon on instrumental variable (IV) estimates of treatment effects of composite years of education on multiple outcomes using multiple IVs. It uses a factor model to attribute differing estimated composite treatment effects for a given outcome across IVs to differences between IVs in weights placed on unobserved components of education in a factor model, and differing estimated composite treatment effects for a given IV across outcomes to differences between component treatment effects for the unobserved components of education that the IV places weight on.

This paper contributes to the empirical literature on returns to education, specifically on effects over the lifecycle on wages. We find that college attendance associated with local labor market conditions produces a higher wage return profile than college attendance associated with local college proximity. Our method has the benefit of obtaining these results without establishing ex ante which types of college (major, quality, etc) should be investigated, at the

cost of requiring researcher judgment in determining the types of college responsible for these differences in returns.

We also contribute to the theoretical literature on treatment effect estimation using instrumental variables by introducing an alternative explanation for differences in effects between IVs to the local average treatment effect framework described in by Imbens and Angrist (1994). There is particular conceptual overlap between our method and the marginal treatment effect (MTE) literature (Heckman and Vytlacil, 1999, 2001, 2005, 2007; Carneiro, Heckman, and Vytlacil, 2011; Mogstad, Torgovitsky, and Walters, 2021), as both approaches estimate structural parameters that drive IV estimates through unobserved heterogeneity. The simplest interpretation of results from our method attributes differences between IVs in estimated treatment effects to homogeneous treatment effects of heterogeneous treatments, while MTE attributes them to heterogeneous treatment effects of homogeneous treatments. Furthermore, just as instrumental variables may fail to estimate policy-relevant treatment effects due to arbitrary weighting of parameters in the MTE framework, they may fail to do so in our framework by weighting arbitrary weighting of component treatments.

Finally, we 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. Our 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, our 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 two-stage least squares and introduces a monotonicity condition under which the IV estimand is a positively-weighted average of component treatment effects. Section 3 describes a parametric model well-suited to our empirical application regarding the returns to college, beginning with

the decomposition of standard IV estimates in terms of unobserved component treatment effects. Section 4 describes the data. Section 5 discusses results and their interpretation. Section 6 concludes.

2 Theory

We work with a potential outcomes framework similar to that used by Imbens and Angrist (1994), with reference as well to Rubin (1974, 1990), Heckman (1990), and Angrist and Imbens (1991). The main innovation in our paper is to model an observed composite treatment, D_i, as being composed of L unobserved component treatments $\{D_{i1}, D_{i2}, ...D_{iL}\}$ such that $D_i = \sum_{\ell=1}^{L} D_{i\ell}$. We consider binary component and composite treatments for simplicity of exposition in this section.[1] We define $Y_i(0)$ as the value of the outcome in the absence of treatment, which is observed when the individual selects none of the available component treatments. We next define $Y_i(\ell)$, for $\ell = 1, 2, ..., L$ as the value of the outcome when $D_{i1} = 1$, $D_{i2} = 1$, ..., or $D_{iL} = 1$, respectively.

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 any combination of ℓ and m.[2] This muddling of potential outcomes for treated individuals is the central empirical challenge we 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 (1974) 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

[1] It follows that each individual will either have $D_{i\ell}$=0 for all ℓ or $D_{i\ell} = 0$ for all $\ell \neq m$ and $D_{im} = 1$.

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

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.[3] The control of the trial ensures identification of treatment effects that are entirely driven by a single component treatment, while the randomization ensures that the effects can be estimated via comparisons of treated and untreated individuals.

2.1 Problems and Solutions for Instrumental Variables

It is common for researchers to use instrumental variables to estimate the effects of treatments, especially when randomized controlled trials are infeasible or when random policy variation is assumed to affect outcomes through known channels. We consider a vector of instrumental variables, Z_i, within a subset $\mathcal{Z}$ of $\mathbb{R}^K$, where K gives the number of instruments. We define an instrument as a variable that satisfies the following conditions. First, it is independent of $Y(0)$ and $Y(\ell)$ for all $\ell = 1, ..., L$. Second, it is correlated with the composite treatment D_i, which implies correlation with at least one unobserved component treatment in $\{D_{i\ell}\}_{\ell=1}^{L}$. For any vector value z within $\mathcal{Z}$, we define $D_{i\ell}(z)$ to be the value of the ℓth component treatment for i if the value of Z_i is equal to z. We observe (Y_i, D_i, Z_i) for a random sample of a population of interest. The key distinction in this discussion from that in Imbens and Angrist (1994) is that different values of the instrument not only shift people into and out of the composite treatment, but they also determine which component treatment the individual selects. Formally, we define an instrument according to the following condition:

Condition 1 (Existence of Instruments): *Let Z be a vector of random variables such that (i) for all $z \in \mathcal{Z}$ the $(2L+1)$-tuple $(Y(0), Y(1), ..., Y(L), D_{i1}(z), ..., D_{iL}(z))$ is jointly independent of Z_i, and (ii) $P(z) = \mathbb{E}[D_i | Z_i = z]$ is a nontrivial function of z.*

The difference between this condition and that given by Imbens and Angrist (1994) is the presence of multiple treatments $(D_{i1}(z), ..., D_{iL}(z))$ and potential outcomes $(Y(0), Y(1), ..., Y(L))$. Part (ii) is testable in applications, and further implies that $P_\ell(z) = \mathbb{E}[D_{i\ell} | Z_i = z]$ is a nontrivial function of z for at least some ℓ in $\ell = 1, 2, ..., L$. The effects of the instrument on individual component treatments are not testable because the component treatments are unobserved.

To demonstrate the problems that unobserved treatment effect heterogeneity presents for identification of average treatment effects, we compare expected outcomes at two distinct values

[3] 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.

of the instrument, $\mathbb{E}[Y_i|Z_i = z]$ and $\mathbb{E}[Y_i|Z_i = z']$. Taking differences, we have

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

Condition 1 implies that this is equal to

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

This expression suggests two potential problems for average treatment effect identification. The first, described in detail by Imbens and Angrist (1994), occurs with treatment effect heterogeneity when $\mathbb{E}[Y_i(\ell) - Y_i(0)|D_{i\ell}(z) - D_{i\ell}(z') = 1] \neq \mathbb{E}[Y_i(\ell) - Y_i(0)|D_{i\ell}(z) - D_{i\ell}(z') = -1]$. The additional problem highlighted here occurs with treatment heterogeneity when $\mathbb{E}[Y_i(\ell) - Y_i(0)|D_{i\ell}(z) - D_{i\ell}(z') = a] \neq \mathbb{E}[Y_i(\ell') - Y_i(0)|D_{i\ell'}(z) - D_{i\ell'}(z') = a]$ for $a = -1, 1$.

To more clearly illustrate this new problem, consider the case of homogeneous treatment effects across individuals for each unobserved component treatment. This reduces equation (1) to

$$\begin{aligned}
\sum_{\ell=1}^{L} \mathbb{E}\Big[(D_{i\ell}(z) - D_{i\ell}(z'))(Y_i(\ell) - Y_i(0))\Big] &= \sum_{\ell=1}^{L} (P_\ell(z) - P_\ell(z'))\lambda_\ell \\
&= \sum_{\ell=1}^{L} I(P_\ell(z) \geq P_\ell(z'))(P_\ell(z) - P_\ell(z'))\lambda_\ell \\
&- \sum_{\ell=1}^{L} I(P_\ell(z) \leq P_\ell(z'))(P_\ell(z') - P_\ell(z))\lambda_\ell
\end{aligned} \tag{2}$$

where $P_\ell(z) = \mathbb{E}[D_{i\ell}|Z_i = z]$ gives the probability of component treatment ℓ at instrument value z, λ_ℓ gives the homogeneous treatment effect for component treatment ℓ, and $I(\cdot)$ is an indicator function that takes a value of unity if its argument is true and zero otherwise. Even with this simplification, unobserved treatment heterogeneity leaves room for variation in

identified treatment effects for the composite treatment, D_i, when using different instrumental variables (or different shifts in the value of a single instrument).

To see this, consider the case of two component treatments, where the second component treatment effect is double that of the first, and the change in the value of the instrument shifts equal numbers of individuals from nontreatment and the second component treatment into the first component treatment. Formally, this entails

$$\lambda_2 = 2\lambda_1$$

and

$$P_1(z) - P_1(z') = -2(P_2(z) - P_2(z')).$$

Substituting these expressions into equation (2) yields a difference in outcomes of zero regardless of the strength of the first stage or the magnitudes or signs of the component treatment effects. Amplifying the disparities in component treatments' effects or responses to the instrument further can produce negative IV estimands even when all underlying component treatment effects are positive, and vice versa. Intuitively, if some individuals respond to the instrument by changing their treatment type to or from an extreme version of the treatment, the type they select out of will be weighted negatively and the type they select into will be weighted positively, even though their observed composite treatment status has not changed.[4]

To identify an effect of the composite treatment that is a positively-weighted average of the effects of component treatments, we present the following assumption on instruments in addition to those described in Condition 1.

Condition 2 (Net Monotonicity): *For all $z, z' \in \mathcal{Z}$, either (i) $D_{i\ell}(z) - D_{i\ell}(z') = sgn(P(z) - P(z'))$, (ii) $D_{i\ell}(z) - D_{i\ell}(z') = 0$ for all i and all ℓ, or (iii) $Pr(D_{j\ell}(z) - D_{j\ell}(z') = sgn(P(z) - P(z'))|Y_j(\ell) - Y_j(0) = Y_i(\ell) - Y_i(0)) \geq Pr(D_{j\ell}(z) - D_{j\ell}(z') = -sgn(P(z) - P(z'))|Y_j(\ell) - Y_j(0) = Y_i(\ell) - Y_i(0))$ for all ℓ.*

Cases (i) and (ii) correspond to the monotonicity condition described by Imbens and Angrist (1994), extended to the current setting with multiple unobserved treatments. Kline and Walters (2016) describe a similar condition that restricts changes in the treatment decision to a single

[4]For example, consider a randomized STEM-specific college scholarship. We might expect such a scholarship to induce some individuals to switch from non-college to STEM-college and some to switch from non-STEM-college to STEM-college. If STEM majors have relatively high returns, using this policy as an instrument for college attendance will overstate returns to college.

observed component treatment in the presence of multiple observed component treatments, while Hull (2018) describes a similar condition that restricts changes in treatment decision from one of multiple observed untreated states to a single observed component treatment. The relevant distinction in this paper is that even when (i) and (ii) hold for all individuals, our Condition 2 still permits multiple unobserved types of treatment, wherein identified treatment effects are not only weighted averages of treatment effects for particular groups of individuals, but also for particular component treatments.

To better discuss case (iii), we adapt the language of Angrist, Imbens, and Rubin (1996) to describe groups of individuals with respect to their component treatment statuses for different instrument comparisons. A component ℓ always-taker chooses component treatment ℓ regardless of their value for Z_i. A component ℓ never-taker never chooses component treatment ℓ regardless of their value for Z_i. Always takers for any component ℓ are necessarily never-takers for all other component treatments, as well as the untreated state. Component ℓ compliers at instrument comparison z, z' have $D_{i\ell}(z) - D_{i\ell}(z') = \text{sgn}(P(z) - P(z'))$. Finally, component ℓ defiers at instrument comparison z, z' have $D_{i\ell}(z) - D_{i\ell}(z') = -\text{sgn}(P(z) - P(z'))$.

Case (iii) of Condition 2 essentially allows for defiers for any component as long as there are compliers for that component to cancel them out. When a subset of compliers cancel out defiers, the local treatment effect estimated at the instrument comparison will be driven by the residual subset of compliers that do not cancel out defiers.

To define this set of residual compliers more formally, we define the set of defiers for component treatment ℓ at z, z' as $\mathcal{I}_\ell^{D;z,z'}$ and the set of compliers for component ℓ as $\mathcal{I}_\ell^{C;z,z'}$. We then partition the data such that $i = 1, \ldots, N_{D1} \in \mathcal{I}_1^{D;z,z'}, i = N_{D1} + 1, \ldots, N_{D1} + N_{C1} \in \mathcal{I}_1^{C;z,z'}, \ldots, i = \sum_{\ell=1}^{L-1} N_{D\ell} + N_{C\ell} + 1, \ldots, \sum_{\ell=1}^{L-1} N_{D\ell} + N_{C\ell} + N_{DL} \in \mathcal{I}_L^{D;z,z'}$, $i = \sum_{\ell=1}^{L-1} N_{D\ell} + N_{C\ell} + 1, \ldots, \sum_{\ell=1}^{L} N_{D\ell} + N_{C\ell} \in \mathcal{I}_L^{C;z,z'}$, where $N_{D\ell}$ is the number of defiers for component treatment ℓ and $N_{C\ell}$ is the number of compliers for component treatment ℓ. In words, we order individuals such that defiers for component treatment 1 are first, followed by compliers for component treatment 1, then defiers for component 2, and so on (with always-takers and never-takers for all component treatments coming last). Within the set $\mathcal{I}_\ell^{D;z,z'}$ for each ℓ, we order individuals such that $Y_i(\ell) - Y_i(0) \leq Y_{i+1}(\ell) - Y_{i+1}(0)$ for all $i \in \mathcal{I}_\ell^{D;z,z'}$. Within the set $\mathcal{I}_\ell^{C;z,z'}$, we order individuals such that $\sum_{i=N_{-\ell}+N_{D\ell}+1}^{N_{-\ell}+2N_{D\ell}} |(Y_i(\ell) - Y_i(0)) - (Y_{i-N_{D\ell}}(\ell) - Y_{i-N_{D\ell}}(0))|$ is minimized, where we define $N_{-\ell} = \sum_{m=1}^{\ell-1} N_{Dm} + N_{Cm}$. We define a residual complier for component treatment ℓ as an $i \in \mathcal{I}_\ell^{RC}$ for whom $i > N_{-\ell} + 2N_{D\ell}$. In words, this ordering

pairs each defier for component treatment ℓ with a complier for component treatment ℓ at the same position within the order of compliers, such that compliers that are not paired with a defier have relatively high values of the index i. Condition 2 ensures that the minimum of $\sum_{i=N_{-\ell}+N_{D\ell}+1}^{N_{-\ell}+2N_{D\ell}} |(Y_i(\ell) - Y_i(0)) - (Y_{i-N_{D\ell}}(\ell) - Y_{i-N_{D\ell}}(0))| = 0$ such that each defier-complier pair have equal treatment effects.

Other similar sufficient conditions exist for identification of positively-weighted local average component treatment effects. One is the existence of a value of the instrument z such that $P(z) = 0$ or $P(z) = 1$, such that all comparisons of $z, z' \in \mathcal{Z}$ only include compliers for all component treatments, as discussed in the homogeneous treatment setting by Heckman (1990) and Angrist and Imbens (1991).[5] An intuitively similar condition allows for defiers at some instrument comparisons as long as a sufficient mass of individuals with the same treatment effect are compliers at other instrument comparisons, as described in Mogstad, Torgovitsky, and Walters (2021).[6] We concentrate on our Condition 2 because cases (i) and (ii) are analogous to the commonly-invoked Imbens and Angrist (1994) monotonicity condition, while case (iii) simplifies to the following under component treatment effect homogeneity:

Condition 3 (Homogeneous Treatment Effects Net Monotonicity): *(i)* $Y_i(\ell) - Y_i(0) = \lambda_\ell$ *for all* i *and (ii) 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*.*

This condition preserves the logic of the IA monotonicity condition with the modification that the relevant compliers for instrument comparisons are component treatments rather than individuals.

We can identify an average treatment effect of the composite treatment using instrumental variables by dividing the difference in outcomes at different values of instruments expressed in (1) by the net difference in the share of the population that chooses the composite treatment. Condition 2 ensures that this average treatment effect is a positively weighted average of underlying component treatment effects for a subset of the population of interest. Formally, we provide the following result:

[5]The particular value of restricting individuals to these two cases is one advantage of randomized controlled trials.

[6]We note that the tests described in Mogstad, Torgovitsky, and Walters (2021) to determine whether there are sufficient compliers at well-behaved instrument comparisons to cancel out noncompliers at poorly-behaved instrument comparisons are not immediately applicable to settings with unobserved treatment heterogeneity.

Theorem 1: *If Conditions* 1 *and* 2 *hold, the average treatment effect defined by*

$$\pi_{z,z'} = \sum_{\ell=1}^{L} \theta_\ell \mathbb{E}\Big[Y_i(\ell) - Y_i(0)\Big| i \in \mathcal{I}_\ell^{RC;z,z'}\Big]$$

is identified from the joint distribution of (Y, D, Z) *and* $\theta_\ell \in [0,1]$ *for all* $\ell = 1, 2, ..., L$ *for all* $z, z' \in \mathcal{Z}$ *such that* $\mathbb{E}[Y_i|Z_i = z]$ *and* $\mathbb{E}[Y_i|Z_i = z']$ *are finite and* $P(z) \neq P(z')$*, with* θ_ℓ *given by*

$$\theta_\ell = \frac{P_\ell(z) - P_\ell(z')}{P(z) - P(z')}.$$

Proof. Let Condition 1 be satisfied. It follows that equation (1) holds. Using the sets defined above, this expression is equivalent to

$$\begin{aligned}
&\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] \\
&= \sum_{\ell=1}^{L} \Pr(i \in \mathcal{I}_\ell^{RC;z,z'}) \mathbb{E}[Y_i(\ell) - Y_i(0)| i \in \mathcal{I}_\ell^{RC;z,z'}] \\
&+ \sum_{\ell=1}^{L} \Pr(i \in \mathcal{I}_\ell^{C}, i \notin \mathcal{I}_\ell^{RC;z,z'}) \mathbb{E}[Y_i(\ell) - Y_i(0)| i \in \mathcal{I}_\ell^{C}, i \notin \mathcal{I}_\ell^{RC;z,z'}] \\
&- \sum_{\ell=1}^{L} \Pr(i \in \mathcal{I}_\ell^{D;z,z'}) \mathbb{E}[Y_i(\ell) - Y_i(0)| i \in \mathcal{I}_\ell^{D;z,z'}].
\end{aligned}$$

Let Condition 2 be satisfied as well. It follows from the definitions of the sets $\mathcal{I}_\ell^{D;z,z'}$, $\mathcal{I}_\ell^{C;z,z'}$, and $\mathcal{I}_\ell^{RC;z,z'}$ that the last two lines of the above expression cancel out, yielding

$$\begin{aligned}
&\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] \\
&= \sum_{\ell=1}^{L} \Pr(i \in \mathcal{I}_\ell^{RC;z,z'}) \mathbb{E}[Y_i(\ell) - Y_i(0)| i \in \mathcal{I}_\ell^{RC;z,z'}].
\end{aligned}$$

Given that $D_{i\ell}(z) - D_{i\ell}(z') \neq 0$ iff $i \in \mathcal{I}_\ell^{C;z,z'}$ or $i \in \mathcal{I}_\ell^{D;z,z'}$, it follows that $\Pr(i \in \mathcal{I}_\ell^{RC;z,z'}) = P_\ell(z) - P_\ell(z')$. Making this substitution and multiplying the right-hand side of the equation by $(P(z) - P(z'))/(P(z) - P(z'))$ gives

$$\begin{aligned}
&\mathbb{E}[Y_i|Z_i = z] - \mathbb{E}[Y_i|Z_i = z'] \\
&= (P(z) - P(z')) \sum_{\ell=1}^{L} \frac{P_\ell(z) - P_\ell(z')}{P(z) - P(z')} \mathbb{E}[Y_i(\ell) - Y_i(0)| i \in \mathcal{I}_\ell^{RC;z,z'}].
\end{aligned}$$

Dividing each side of the equation by $(P(z) - P(z'))$ shows that the local component average

treatment effect $\pi_{z,z'}$ is identified in terms of moments of the distribution of (Y, D, Z). □

Figure 1 shows a graphical illustration of the decision setting. The horizontal axes give a latent utility index for the composite treatment, such that the composite treatment is chosen if $V_i \geq 0$. The vertical axes give a relative preference for two component treatments, such that treatment 1 is chosen if $V_{1i} \geq 0$ and $V_i \geq 0$, and treatment 2 is chosen if $V_{11} < 0$ and $V_i \geq 0$. The ovals drawn in each figure contain uniform joint distributions of individuals' preferences for treatments. Each panel shows a different instrument comparison assumed to satisfy Condition 1. The instrument shift in panel A satisfies Condition 2 without any individual defiers for any component treatment. The shift in panel B satisfies Condition 2 under a homogeneous component treatment effects assumption, while the shift in panel C does not.

The assumption that no individuals switch their preferred version of treatment in response to instruments may be strong in many settings, even those in which Imbens and Angrists' monotonicity assumption is uncontroversial. For instance, an instrument that reduces costs for college may produce income effects, which may cause some students to obtain leisure by choosing lower-earnings majors. It may seem similarly strong to assume that there happen to be compliers that exactly cancel out defiers as we describe in case (iii) of Condition 2. However, if treatment composition shifts are small relative to net shifts in treatment, the number of compliers who are available to cancel out defiers will be relatively large. If the defier treatment effect distribution is encompassed by the support of the complier treatment effect distribution, Condition 2 will be satisfied if compliers sufficiently outnumber defiers. The economic content of such an assumption is that instrumental variables cause shifts in observed composite treatments that are large relative to unobserved component treatment shifts, which seems likely to be satisfied in many settings where a strict person-specific component treatment monotonicity condition would be violated.

3 Estimation of Unobserved Treatments' Effects

We apply the insights above to estimate returns to unobserved components of college education. We think that unobserved components of college may manifest in the intensive margin, such as if students alter the composition of classes they choose to take while in college, or if they modulate their study intensity across classes. To capture this, we model the unobserved component treatments as continuous for our application, despite the binary nature of the observed

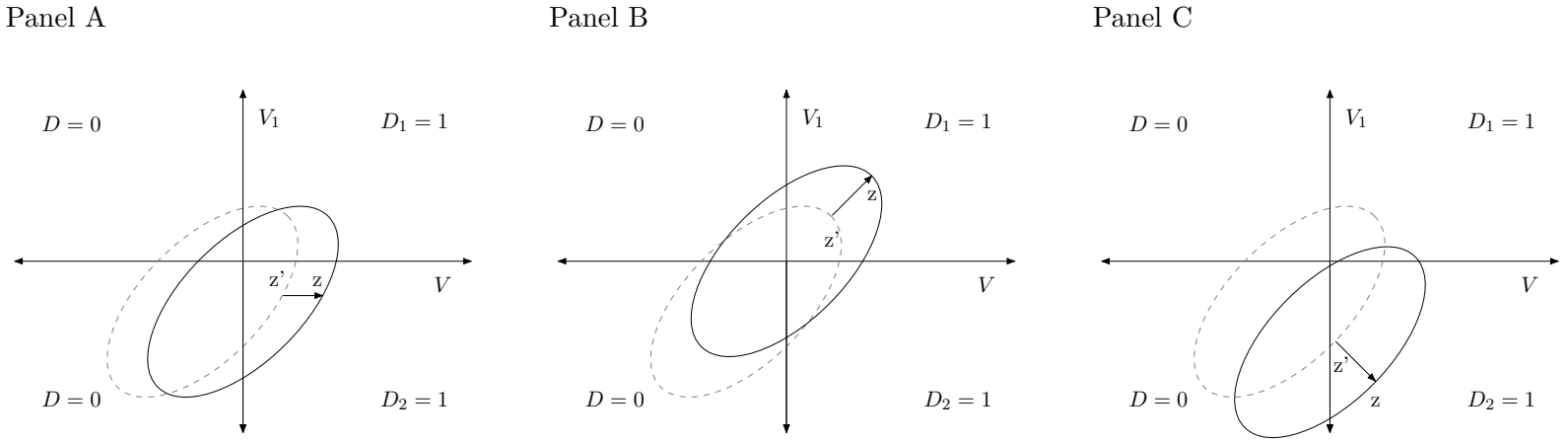

Figure 1: Instrument Comparisons with Two Unobserved Component Treatments

Notes: For composite treatment preference V and relative preference for component treatment 1 V_1, each panel shows a different instrument comparison. Panel A depicts an instrument comparison where individuals' preference for treatment increases, but there is no effect on the composition of treatment (no defiers). Panel B shows a comparison where more individuals choose each type of treatment at z than at z', but there are compositional changes in the individuals who choose each component treatment. If component treatment effects are homogeneous, the situation in Panel B will satisfy our Net Monotonicity condition. Panel C depicts an instrument comparison where the aggregate composition of treatment shifts in response to the treatment. Net monotonicity will only hold in this setting if the idiosyncratic treatment effects of compliers cancel out the treatment effects for the group of people who shift from component treatment 1 to component treatment 2, for instance if component 1 and component 2 have identical treatment effects for defiers.

composite treatment.

For an individual indexed by i among N individuals and an outcome indexed by j among J outcomes, the model describes L endogenous unobserved components of college education, F_i, affecting an outcome, Y_{ij}, as follows

$$Y_{ij} = X_i\beta_j + F_i\lambda_j + \epsilon_{ij}. \tag{3}$$

Outcome-specific marginal effects of F_i $(1 \times L)$ on scalar Y_{ij} are given by λ_j $(L \times 1)$. X_i $(1 \times R)$ contains R observed exogenous control variables with corresponding outcome-specific marginal effects given by β_j $(R \times 1)$ and ϵ_{ij} is an unobserved residual. If we observed the components directly and had access to $K \geq L$ excludable instruments, Z_{1i}, we could estimate the second stage given by equation (3) along with the first stage given by

$$F_i = Z_i\Gamma + U_i \tag{4}$$

via 2SLS, where F_i is determined by instruments $Z_i = [Z_{1i}\ X_i]$, Γ $(R+K \times L)$ gives the marginal associations of the instruments with the components, Γ_k gives the marginal associations of the kth excludeable instrument with the components, and U_i $(1 \times L)$ represents the idiosyncratic portion of the components.[7] The goal of the method is to estimate the the marginal effects, λ_j, despite not observing F_i.

3.1 Decomposing Two-Stage Least Squares

Because we do not observe F_i, we cannot directly estimate (3) or (4). However, we do observe the sum of the latent factors via the composite college attendance variable,

$$D_i \equiv \sum_{\ell=1}^{L} f_{i\ell}, \tag{5}$$

[7] We assume that there are no controls that vary across outcomes. Controls that vary with j can be included into this framework if they are excluded from the first stage, as in Carneiro, Heckman, and Vytlacil (2011). It is possible to include outcome-varying controls if the relative channels, θ, through which instruments affect component treatments are assumed to be outcome-invariant, which will no longer be true by construction with multiple first stages.

where $F_i = [f_{i1}\ f_{i2}\ \dots f_{iL}]$. This presents a tractable alternative; we can estimate the observed model (in matrix notation)

$$\begin{aligned} Y_j &= X_k\beta_{kj} + D\pi_{kj} + \epsilon_{kj} \\ D &= Z\gamma + u, \end{aligned} \tag{6}$$

by 2SLS, wherein we substitute the observed composite variable, D $(N\times 1)$, in for the unobserved components, F $(N \times L)$. To focus on the estimate of π_{kj} associated with a single IV, we also include all but one of the excludable instruments as controls, defining $X_k = [Z_{k'}\ X]$ where X is $N \times R$ and $Z_{k'}$ $(N \times K-1)$ contains all excludable instruments other than a single z_k $(N \times 1)$, such that $Z = [z_k\ X_k]$.[8]

In this setup, β_{kj} $(1 \times (R+K-1))$ gives the marginal effect of X_k on Y_j, π_{kj} gives the scalar "effect" of D on Y_j as determined by z_k, ϵ_{kj} is the idiosyncratic component of the outcome which contains the portion of $F\lambda_j$ that is not captured by $D\pi_{kj}$, γ $((R+K)\times 1)$ gives the marginal associations of the instruments with observed education, γ_k gives the scalar marginal association of the kth excludeable instrument with the observed education, and u $(N \times 1)$ is the idiosyncratic portion of observed education. Note that the instruments included in X_k that are excludable from (3) nonetheless may have nonzero coefficients within β_{kj} because they are correlated with $F\lambda_j$, which is contained within ϵ_{kj} but is not in ϵ_j. The estimated value of π_{kj} (as well as β_{kj} and ϵ_{kj}) varies with the excluded instrument because instruments differ in their correlations with latent components, and because each latent component affects the outcome differently.[9]

[8] Including all but one instrument as controls ensures that the estimate of π_{kj} is only driven by the correlations between the excluded instrument and the latent components. Excluding other instruments from both equations would cause the estimate of π_{kj} to reflect the correlations between other instruments and latent components via correlation between other instruments and z_k, making comparisons between estimates of π_{kj} across instruments difficult. The analogous problem in the context of heterogeneous treatment effects is discussed by Carneiro, Heckman, and Vytlacil (2011) and Mogstad, Torgovitsky, and Walters (2021).

[9] An alternative framing emphasizes that the "true" effect of D on Y_j is 0 conditional on F. It follows that any nonzero estimate for π_{kj} reflects omitted variable "bias" driven by correlation between observed schooling and unobserved components of schooling.

Specifically, defining $\hat{D} = Z(Z'Z)^{-1}Z'D$ and $M_X = I_N - X_k(X_k'X_k)^{-1}X_k'$, we have

$$\begin{aligned}
\hat{\pi}_{kj} &= (\hat{D}'M_X D)^{-1}\hat{D}'M_X Y_j \\
&= (\hat{D}'M_X D)^{-1}\hat{D}'M_X(X\beta_j + F\lambda_j + \epsilon_j) \\
&= (\hat{D}'M_X D)^{-1}\hat{D}'M_X(F\lambda_j + \epsilon_j) \\
&= (\hat{D}'M_X D)^{-1}\hat{D}'M_X F\lambda_j + (\hat{D}'M_X D)^{-1}\hat{D}'M_X\epsilon_j \\
&= \theta_k\lambda_j + \omega_{kj}
\end{aligned} \tag{7}$$

where the first row is the 2SLS estimator for the "effect" of D on Y_j. The second row substitutes in the definition for Y_j from (3). The third row substitutes $(\hat{D}'M_X D)^{-1}\hat{D}'M_X X\beta_j = 0$ by the definition of M_X. The fourth row separates the relationship between z_k and Y_j that is due to its correlation with F from that which is due to its correlation with ϵ_j, which may be nonzero in small samples. Finally, the fifth row extracts $\theta_k = \Gamma_k/\gamma_k = (\hat{D}'M_X\hat{D})^{-1}\hat{D}'M_X F$ and defines a new error, $\omega_{kj} = (\hat{D}'M_X\hat{D})^{-1}\hat{D}'M_X\epsilon_j$. The term ω_{kj} converges to zero as N increases if instruments are valid, so $\hat{\pi}_{kj}$ is a consistent estimate of $\theta_k\lambda_j$.[10]

Note that the weights sum to unity. Defining $\theta_k = [\theta_{k1}\ \theta_{k2}\ldots\theta_{kL}]$, we have

$$\begin{aligned}
\sum_{\ell=1}^{L}\theta_{k\ell} &= (\hat{D}'M_X D)^{-1}\hat{D}'M_X\sum_{\ell=1}^{L} f_\ell \\
&= (\hat{D}'M_X D)^{-1}\hat{D}'M_X D \\
&= 1.
\end{aligned} \tag{8}$$

The first line follows from definitions of $\theta_{k\ell}$ and f_ℓ. The second line follows from the constraint in (5).

3.2 Identification of Component Effects

To establish identification conditions for θ_k and λ_j, we consider model (6) as $N \to \infty$ separately for each instrument and each outcome, such that $\omega_{kj} \to 0$ for all k and j. First, we arrange each π_{kj} into a $K \times J$ matrix Π. Next, we define a $K \times L$ matrix $\Theta = [\theta_1'\ \theta_2'\ ...\ \theta_K']'$, where each row of Θ gives all component weights for a single instrument. We also define the $L \times J$ matrix

[10] In an alternative model where unobserved factors F_j (or controls, X_j) vary with j, θ_{kj} could also vary with j (this could arise, for instance, if j indexes time-periods or sub-populations). To estimate such a model with our method, it would be necessary to assume that $\theta_{kj} = \theta_k\ \forall j$, i.e. the weights instruments place on unobserved components are common across j. This assumption's validity would depend on the application's instruments, outcomes, and composite treatment.

of outcome-specific component treatment effects as $\Lambda = [\lambda_1 \ \lambda_2 \ ... \ \lambda_J]$. It follows that

$$\Pi = \Theta\Lambda, \tag{9}$$

where the reduced form 2SLS estimates of Π are thus decomposed into outcome-invariant component weights and instrument-invariant component effects.

A barrier to identification, common to factor models, is that $\Theta\Lambda = \Theta A A^{-1}\Lambda$ for an arbitrary invertible $L \times L$ matrix A. A has L^2 free elements, so we require L^2 restrictions. To address this, we impose the normalization

$$\Theta = \begin{bmatrix} I_L \\ \theta_{L+1} \\ \vdots \\ \theta_K \end{bmatrix}, \tag{10}$$

where I_L is an $L \times L$ identity matrix. The economic content of this normalization is to define the components in terms of the first L instruments. For instance, if a particular instrument affects college attendance by inducing students to major in math, assigning a component to that instrument via this normalization will define it as the "math" type of college. Meanwhile, the associated λ_j will give the effect of attending college with a math major on outcome j.

The relationships given in (8), (9), and (10) give K, KJ, and $L^2 - L$ restrictions, respectively, where the first L restrictions in (8) are redundant given the restrictions in (10). Meanwhile, there are KL parameters in Θ and JL in Λ. It follows that identification requires

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

where solving for L yields

$$L \leq (K, J + 1). \tag{11}$$

This expression defines the maximum number of factors that can be identified with a given number of instruments and outcomes. Considering the simplest interesting case in which $L = 2$, this requires at least two instruments and at least one outcome. In the just-identified case, this defines $\Pi = \Lambda$ and $\Theta = I_2$. The model in this simple case is largely semantic, but still formalizes

the argument that differences between π_{j1} and π_{j2} may be driven by differences in the latent components through which the two instruments affect the observed composite treatment.

To understand the intuition for the case with more instruments and outcomes than needed, consider the case of $L = 2$, $K = 3$, and $J = 2$. For outcome one, the three instruments will each produce a $\hat{\pi}_{k,1}$. The first two IV estimates for outcome one will define components, while $\hat{\pi}_{3,1}$ must be either a weighted average of the first two components, or it is driven by some third component. With only one outcome, a Θ that rationalizes $\hat{\pi}_{3,1}$ is exactly identified as a weighted average of the first two components. However, if the Θ that rationalizes $\pi_{3,1}$ also happens to rationalize $\pi_{3,2}$, then we are able to describe the variation across 3 instruments and 2 outcomes with only 2 unobserved components. Conversely, if the value of Θ that rationalizes estimated effects across instruments in outcome one fails to do so in outcome two, then we may reject that two components can explain the data, and conclude that $L > 2$ (or that treatment effect heterogeneity is responsible for the variation between IV estimates).

3.3 Estimation

The discussion in section 3.1 considers a single outcome and a single instrument, but identification requires that we estimate the model using multiple outcomes and instruments as discussed in section 3.2. To do this, we leverage the analytical form of the 2SLS estimand in (7) to set $\pi_{kj} = \theta_k \lambda_j$. It follows that we can substitute $\theta_k \lambda_j$ in for π_{kj}, producing equivalent outcome equations for each excluded instrument and each outcome of the form

$$Y_j = X_k \beta_{kj} + D\theta_k \lambda_j + \epsilon_{kj}. \tag{12}$$

We next leverage the assumption that $\mathbb{E}[\epsilon'_{kj} Z] = 0$, which allows us to construct moment conditions of the form

$$\mathbb{E}[(Y_j - X_k \beta_{kj} - D\theta_k \lambda_j) Z] = 0. \tag{13}$$

We then estimate the model by generalized method of moments. The estimation procedure amounts to the standard instrumental variables generalized method of moments estimator with an alternative factor-structure parameterization on treatment effects.

4 Data

Our primary data source is 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 our 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 us to emphasize our methodological contributions by holding fixed some other potential explanations for similarities and differences in results. In particular, we 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 our model requires multiple outcomes and instruments, as shown in expression (11). The longitudinal nature of the NLSY allows us to set $j = t$ for $t = 25, \ldots, 40$ and rely on wage outcomes over the lifecycle to identify the model.[11]

We use several instruments that are established in the literature on returns to education. They are an indicator for the presence of a college in county of residence at age 14, average tuition of public colleges in the county of residence at age 17, and average wages in the county of residence at age 17.[12] That each of these variables contributes to the college attendance decision is intuitive as well as testable. We discuss instruments separately in terms of excludability and intuition for the type of college they identify.

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). We expect the presence of colleges in the county of residence to primarily influence education by encouraging students to attend college through a cost reduction mechanism. If cost reduction is more important for students inclined toward particular

[11] 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.

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

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. The validity of this instrument is threatened by correlation between college presence and students' ability, as noted by Carneiro and Heckman (2002) and Cameron and Taber (2004). To address this, we 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. We control for mother's years of education, permanent local labor market conditions, and AFQT to address this concern. We expect this instrument to affect educational attainment through similar channels as the presence of a college in the county of residence. By defining an unobserved component of education in terms of one of these instruments, we can empirically check if our model describes the other instrument as placing a high weight on this component, which is an intuitive check for the validity of our model.[13]

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), we use the average values of these at age 17 at the county and state level, respectively, as instruments for college attendance. Because local labor market conditions may contribute to outcomes through channels other than education, we include permanent values of these as controls in outcome equations as well as selection equations, while including these values at age 17 only in selection equations. It follows that we effectively are using temporal deviations from average in local labor market conditions at age 17 as instruments. Good labor market conditions could resolve credit constraints, driving price sensitive students to attend colleges, which may be relatively cheap or low quality, or they could induce students to drop out of high school or not attend college due to improving outside options.

We use several variables as controls, which is common in the literature. They are number

[13] A possible explanation for these instruments driving education through different channels is that tuition costs drive the type of education that pecuniary-cost-sensitive students are inclined toward, while the presence of a nearby college drives the type of education that nonpecuniary-cost-sensitive students (for instance, those with high psychic costs of distance from family) are inclined toward.

of siblings, permanent local unemployment, permanent local log wages, a binary indicator for urban residence at age 14, average wages in county of residence in 1991, average unemployment in 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 we use are shown in Table 1.

Table 1: Summary Statistics

	Mean (1)	Standard deviation (2)
Mean log wages, ages 26-30	2.083	0.463
Mean log wages, ages 31-35	2.146	0.521
Mean log wages, ages 36-40	2.133	0.526
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
Local tuition at age 17	21.568	7.981
Nearby four-year college at age 14	0.525	0.500
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

We begin by showing results for two stage least squares estimates for each year, excluding each instrument and including the others as controls. These results are in Table 2, and shown graphically in Figure 2. We note that our results are broadly in line with the literature, especially directing the reader to wage returns around age 30, as this is approximately the age targeted by other papers on returns to college that look at a single snapshot in the lifecycle. We also observe substantial variation both between instruments and across time, with returns as driven by local earnings yielding the most extreme results. This is suggestive that the type of education sought by individuals who are sensitive to local earnings produces high wage returns, which is reasonable considering these individuals have revealed themselves in some sense to be particularly money-

motivated.

We also show the estimated effects from the unobserved treatment heterogeneity model of two unobserved components of college on earnings over a sample of the lifecycle below. We define the first component in terms of local earnings at age 17 and the other in terms of local tuition at age 17. The unobserved hneterogeneity results naturally inherit similarities from the 2SLS results, but allow for instruments that may not satisfy Net Monotonicity to nonetheless contribute to identification of treatment effects. A main advantage of including the additional instrument that may not positively weight all component treatments is that it contributes to efficiency in the estimation, making the unobserved treatment heterogeneity estimates substantially more precise than the just-identified IV.

Table 2: Just-Identified 2SLS Estimates of Lifecycle Returns to College, by IV

	Local earnings		Nearby college		Local tuition	
	(1)		(2)		(3)	
Log wage, age 26	0.114	(0.076)	0.005	(0.102)	0.144	(0.110)
Log wage, age 27	0.128	(0.066)	0.092	(0.102)	0.040	(0.106)
Log wage, age 28	0.210	(0.076)	0.051	(0.095)	-0.016	(0.106)
Log wage, age 29	0.194	(0.073)	0.114	(0.099)	0.014	(0.099)
Log wage, age 30	0.124	(0.081)	0.144	(0.098)	0.129	(0.113)
Log wage, age 31	0.119	(0.078)	0.107	(0.096)	0.014	(0.108)
Log wage, age 32	0.173	(0.068)	0.098	(0.094)	-0.024	(0.108)
Log wage, age 33	0.202	(0.075)	0.096	(0.098)	0.057	(0.105)
Log wage, age 34	0.195	(0.079)	0.029	(0.100)	-0.046	(0.110)
Log wage, age 35	0.132	(0.098)	0.005	(0.101)	-0.024	(0.112)
Log wage, age 36	0.235	(0.082)	0.080	(0.089)	0.041	(0.098)
Log wage, age 37	0.179	(0.076)	0.083	(0.087)	0.064	(0.094)
Log wage, age 38	0.163	(0.078)	0.056	(0.089)	0.056	(0.096)
Log wage, age 39	0.157	(0.077)	0.052	(0.089)	0.062	(0.096)
Log wage, age 40	0.153	(0.077)	0.050	(0.088)	0.067	(0.095)
First Stage F-stat for Excluded IV	22.700		9.496		7.834	
Sample size	1747		1747		1747	

Notes: Two-stage least squares estimates of the effects of college attendance on log wages for white men in the NLSY79. Age-specific estimates obtained excluding a single instrument and including all others as controls along with the other control variables listed in Table 1. College returns are annualized by dividing the college attendance coefficient by four. Robust standard errors in parentheses.

There are three main ways to interpret results such as those we have described from our unobserved heterogeneity model. The first, most ambitious way, is to take a strong stance on the types of college education that are induced by each instrument, such that we can strictly define component effects identified by each IV as in Kline and Walters (2016). If we were comfortable saying, for instance, that local earnings cause students to major in math and local tuition costs cause students to major in art, we would interpret Figure 3 as showing the returns to math and

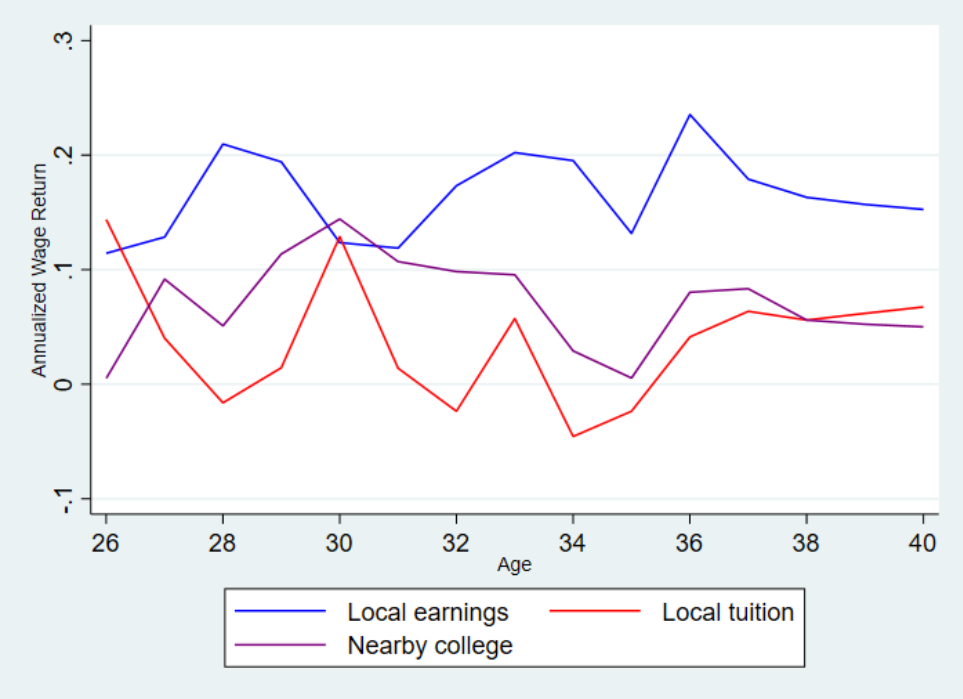

Figure 2: 2SLS Estimates of Lifecycle Wage Returns to College

Notes: This graph depicts just identified 2SLS point estimates of the effect of college attendance on lifecycle log wages for each excluded instrument listed. See Table 2 for point estimates and standard errors.

Figure 3: GMM Estimates of Lifecycle Wage Returns to Unobserved College Components

Notes: This graph depicts GMM point estimates of the effects of unobserved college components on lifecycle log wages for components defined in terms of listed instruments. See Table 3 for point estimates and standard errors.

Table 3: GMM Estimates of Lifecycle Returns to Unobserved College Components

	Local earnings component (1)		Nearby college component (2)	
Component effects (Λ)				
Log wage, age 26	0.100	(0.076)	0.088	(0.063)
Log wage, age 27	0.141	(0.071)	0.078	(0.064)
Log wage, age 28	0.236	(0.082)	0.064	(0.069)
Log wage, age 29	0.217	(0.079)	0.100	(0.065)
Log wage, age 30	0.124	(0.081)	0.140	(0.067)
Log wage, age 31	0.135	(0.080)	0.076	(0.064)
Log wage, age 32	0.201	(0.073)	0.070	(0.065)
Log wage, age 33	0.218	(0.080)	0.105	(0.066)
Log wage, age 34	0.223	(0.083)	0.038	(0.067)
Log wage, age 35	0.148	(0.085)	0.022	(0.070)
Log wage, age 36	0.256	(0.085)	0.095	(0.063)
Log wage, age 37	0.191	(0.076)	0.097	(0.059)
Log wage, age 38	0.173	(0.073)	0.080	(0.058)
Log wage, age 39	0.165	(0.064)	0.080	(0.055)
Log wage, age 40	0.159	(0.071)	0.081	(0.057)
Component weights (Θ)				
Local earnings	1.000	(.)	0.000	(.)
Nearby college	0.000	(.)	1.000	(.)
Local tuition	-1.202	(1.075)	2.202	(1.075)
Value of criterion function (Q)	0.016			
Sample size	1747			

Notes: Two-step generalized method of moments estimates of the effects of unobserved types 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.

art over the lifecycle. This seems quite extreme given the instruments we have on hand for this application, but may be appropriate for other treatments or with other instruments.

A weaker approach to interpretation is to consider more broadly the channels through which instruments might work, so as to get a sense of the likely drivers of treatment effect heterogeneity. For instance, we note that individuals induced to attend college by local earnings must have some (at least indirect) information about and interest in wage differentials across career choices. Such individuals may find earnings differentials between different education choices particularly salient, and may be relative well-informed about them. Students for whom college distance looms large may prioritize nonpecuniary benefits of college over pecuniary returns, as pecuniary transit costs are relatively small compared to pecuniary returns to college. We could infer, then, that the local earnings instrument identifies the effect of high pecuniary-return college, and the nearby college instrument identifies the effect of high amenity-return college.

A still weaker approach to interpretation of results is to leave clarification on exact channels through which instruments affect outcomes to future work that gathers more data on treatment variations. In this framework, the unobserved heterogeneity estimates give insight into the scope and scale of treatment heterogeneity. This would serve as a guide to future research that endeavors to obtain finer data on treatments in an effort to parse out effects of component treatments. Additionally, instruments that are suspected of violating Condition 2 can still be included so as to contribute to efficiency, without taking a strong stance on the channels through which each instrument affects outcomes. With this approach to interpretation, the assumption that some instruments assumed to satisfy Condition 2 work only through particular component treatments is a normalization that fails to pin down a single economically relevant treatment, but does prevent more suspicious instruments from contaminating results.

6 Conclusion

Determining individual level heterogeneity in treatment effects is a major focus in both econometric theory and applied research. This paper draws attention to the similar importance of unobserved heterogeneity in treatments. At a high level, this paper contributes to discussions of both instrumental variables and reduced form results in empirical applications by offering a potential explanation for differences between studies and identification strategies - that different estimators identify effects of different unobserved component treatments. In addition

to explaining differences between empirical results, unobserved treatment heterogeneity is also relevant for policy, as some component treatments may have substantially larger/smaller effects than others.

Additionally, this paper describes a treatment-level net monotonicity condition on instrumental variables that is sufficient for the interpretation of standard IV estimates as positively-weighted averages of unobserved treatment effects. It also provides a method for researchers to use to estimate effects of unobserved treatments when there are multiple outcomes and valid instruments. This method is helpful for investigating heterogeneity in treatments for its own sake, as well as permitting instruments that are exogenous and independent, while violating monotonicity conditions, to contribute to estimation.

Finally, we would be remiss if we did not end with a discussion of similarities and differences between unobserved treatment effect heterogeneity, especially in the the context of the local average treatment effect framework, and unobserved treatment heterogeneity. Broadly, any trait of an individual that affects their returns may manifest through unobserved choices about their preferred version of the treatment, while any variation between individuals in unobserved treatment choices may be perfectly determined by unobserved preexisting characteristics. The differences in policy implications, however, are substantial. In the specific case of returns to college, the possibility that even a portion of the negative returns for a large minority of the population (Cunha, Heckman, and Navarro, 2005) could be driven by unobserved treatment heterogeneity opens the door to potentially massive policy implications for targeted incentives on high-return types of college for these individuals. We leave the ambitious task of separate identification of treatment effect heterogeneity and treatment heterogeneity to future work.

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. 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).

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.

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.

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

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.

Heckman, J. J., and E. Vytlacil (2001): "Policy-relevant treatment effects," *American Economic Review*, 91(2), 107–111.

——— (2005): "Structural equations, treatment effects, and econometric policy evaluation 1," *Econometrica*, 73(3), 669–738.

Heckman, J. J., and E. J. Vytlacil (1999): "Local instrumental variables and latent variable models for identifying and bounding treatment effects," *Proceedings of the national Academy of Sciences*, 96(8), 4730–4734.

——— (2007): "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation," *Handbook of econometrics*, 6, 4779–4874.

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.

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.

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.