NBER WORKING PAPER SERIES

THE ECONOMICS OF EXCLUSION RESTRICTIONS IN IV MODELS

Damon Jones

Working Paper 21391
http://www.nber.org/papers/w21391

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2015

I thank Dan Black, Alex Gelber, Jeff Grogger, Patrick Kline, Jens Ludwig, Aprajit Mahajan, Ofer Malamud, Bruce Meyer, Matthew Notowidigdo and Jim Sallee for helpful comments. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2015 by Damon Jones. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Economics of Exclusion Restrictions in IV Models
Damon Jones
NBER Working Paper No. 21391
July 2015
JEL No. C1,C26,C36

ABSTRACT

We explore a key underlying assumption, the exclusion restriction, commonly used in interpreting IV estimates in the presence of heterogenous treatment effects as a local average treatment effect (LATE). We show through a series of simple examples that in some commonly featured cases that this assumption is likely to be violated among inframarginal agents, i.e. the always- and never-takers. This violation of the exclusion restriction will generally confound the LATE interpretation of the associated IV results. We discuss potential adjustments to IV estimates in the presence of this bias.

Damon Jones
Harris School of Public Policy
University of Chicago
1155 East 60th Street
Chicago, IL 60637
and NBER
damonjones@uchicago.edu

The Economics of Exclusion Restrictions in IV Models

Damon Jones*
University of Chicago
and NBER

June 2015

Abstract

We explore a key underlying assumption, the exclusion restriction, commonly used in interpreting IV estimates in the presence of heterogenous treatment effects as a local average treatment effect (LATE). We show through a series of simple examples that in some commonly featured cases that this assumption is likely to be violated among inframarginal agents, i.e. the always- and never-takers. This violation of the exclusion restriction will generally confound the LATE interpretation of the associated IV results. We discuss potential adjustments to IV estimates in the presence of this bias.

JEL Classification: C26, I26, J01

The method of instrumental variables (IV) is a wildly popular approach for estimating causal effects (see Angrist and Pischke, 2008, for a review). A common interpretation of these IV estimates is that they recover a local average treatment effect (LATE) among a subpopulation of compliers—i.e. those whose treatment status is altered by the instrumental variable (Imbens and Angrist, 1994). We will explore one of the key assumptions, the exclusion restriction, which underlies this interpretation of IV estimates. In particular, we show that restating the standard potential outcome framework as a simple economic model of utility maximization highlights settings where (1) the exclusion restriction is commonly assumed but where (2) the exclusion restriction is likely to be violated. In these cases, the interpretation of standard IV estimates as identifying a local average treatment effect (LATE) among compliers may no longer hold. In particular, estimation of the LATE may be biased by changes in outcomes among inframarginal agents, i.e. the so-called "always-takers" and "never-takers."

*University of Chicago, Harris School of Public Policy and NBER, damonjones@uchicago.edu. I thank Dan Black, Alex Gelber, Jeff Grogger, Patrick Kline, Jens Ludwig, Aprajit Mahajan, Ofer Malamud, Bruce Meyer, Matthew Notowidigdo and Jim Sallee for helpful comments.

To demonstrate the intuition behind this result, consider the estimation of the effect of higher education on subsequent earnings. Suppose we have an instrument—a randomly assigned subsidy for college tuition. We may use this instrument in an IV framework to recover the LATE among those students who will only attend college if they have received this educational subsidy, i.e. the compliers. In the standard LATE framework, there are two other groups: the always-takers, who will attend college with or without the subsidy, and the never-takers, who will not attend college, even with the subsidy. The exclusion restriction assumption implies that since the tuition subsidy does not alter college-going among the two latter groups, it has no it has no effect on their subsequent earnings. It follows that differences in subsequent earnings between those who have received the subsidy and those who have not are driven solely by the group of compliers.

Now reconsider the case of the always-taker. In the absence of the subsidy, she will attend college and will have to pay full tuition. If she is randomly given a tuition subsidy, she will still attend college, but will not have to pay as much to cover the costs. Although her college attendance may not be altered, i.e. there is no direct substitution effect on educational attainment, she may yet experience an income effect, as the college subsidy relaxes the overall budget constraint of inframarginal college-goers. In general, this income effect may cause her to change her subsequent labor supply and, thus, earnings. IV estimates which rely on variation in subsequent earnings between recipients and non-recipients of the tuition subsidy will now pick up differences among the always-takers.

In general, we show below that instruments, such as price subsidies or legal penalties, have the potential to alter outcomes among agents who are inframarginal with respect to treatment status. In what follows, we flesh out this potential bias more formally, in the context of the preceding and other examples, and more generally. We additionally discuss methods of adjusting for this potential bias using either external studies or augmented experimental designs.

This is not the first paper to explore potential violations of the exclusion restriction assumption. For example, Heckman (1997) identifies potential violations to the exclusion restriction that stem from selection into treatment status based on individual, unobserved components of heterogeneous treatment effects. The type of violations discussed here concern affects of an instrument on inframarginal agents, i.e. members of the "aways-takers" and "never-takers" subpopulations. This is discussed in Angrist, Imbens and Rubin (1996), who mention that with respect to the exclusion restriction, "the researcher must consider, at the unit level, the effect of changing the value of the instrument while holding the value of the treatment fixed" (pg. 449). In this paper, we look more formally into cases where such an effect may exist and, thus, alter our interpretation of IV estimates.

Another relevant study is that of Rosenzweig and Wolpin (2000), who discuss the use of the gender composition of earlier born children as an instrument for subsequent fertility in studies of fertility and labor supply (Angrist and Evans, 1998, see e.g.). Rosenzweig and Wolpin (2000) argue that economies of scale in having earlier children of the same sex—via, for example, hand me down

clothes—generate income effects that may directly affect subsequent labor supply. This is similar to one of our examples below, in Section 3.1, where income effects among always-takers are likewise a potential threat to the exclusion restriction. In our case, we explicitly outline how such phenomenon may bias estimation of the LATE, highlight the role played by inframarginal agents and suggest and method for adjusting for this bias.

The rest of the paper proceeds as follows. Section 1 briefly reviews the LATE framework as presented in Imbens and Angrist (1994) and Section 2 recasts the setup in the context of a model of utility maximization, following Pinto (2015). In Section 3 we identify in general the types of violations of the exclusion restriction that may occur and provided specific examples of these violations. We then discuss methods of adjusting for the potential bias in Section 4 and conclude in Section 5.

1 Heterogenous Treatment Effect Framework

We begin by briefly summarizing the standard, potential outcomes framework for heterogenous treatment effects. We are interested in the causal effect of some treatment status, D, on an outcome, y. We also observe an instrument, Z, which affects, ether directly or indirectly, treatment status and the outcome. Each individual i has a potential outcome $y_i(D_i, Z_i)$, which depends on the value of the instrument and treatment. Individual i also has a potential treatment status, $D_i(Z_i)$, which depends on the value of the instrument. We restrict attention to the case where both the instrument and the treatment variables are binary — i.e. $D_i \in \{0,1\}$ and $Z_i \in \{0,1\}$. Following Imbens and Angrist (1994), four assumptions are typically made regarding the variables of interest:

Assumption 1 (LATE Assumptions)

1.1. Independence: $\{y_i(D_i(1),1), y_i(D_i(0),0), D_i(1), D_i(0)\} \perp\!\!\!\perp Z_i$*;*

1.2. Exclusion: $y_i(d,1) = y_i(d,0) \equiv y_i(d)$ *for* $d = 0,1$*;*

1.3. First stage: $E[D_i(1) - D_i(0)] \neq 0$*;*

1.4. Monotonicity: $D_i(1) - D_i(0) \geq 0 \forall i$*.*

Given the assumptions above, define the treatment effect of D_i on y_i for individual i as $y_i(1) - y_i(0)$. Imbens and Angrist (1994) show that the four assumptions imply that a local average treatment effect (LATE) is identified for a subgroup:

$$\frac{\mathbb{E}[y_i|Z_i=1] - \mathbb{E}[y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1] - \mathbb{E}[D_i|Z_i=0]} = \mathbb{E}[y_i(1) - y_i(0)|D_i(1) > D_i(0)] \tag{1}$$

That is, the IV estimate of the effect of D_i on y_i, using the instrument Z_i identifies the causal effect of D_i on y_i for the subgroup of individuals, the *compliers*, for whom the instrument alters treatment status — i.e. those for whom $D_i(1) > D_i(0)$.

2 An Economic Model of Heterogeneous Treatment Effects

We now restate the standard heterogeneous treatment effects framework as model of utility maximization, using a version of the framework presented by Pinto (2015). The model in its most general form nests the potential outcome framework above and will prove useful in exploring the plausibility of Assumption 1.2 above, the exclusion restriction. The model proceeds in two stages.[1] Let the outcome of interest, $y_i(D_i, Z_i) \equiv f(x_i(D_i, Z_i))$ be a function of inputs, $x_i(D_i, Z_i)$. In turn, the inputs are defined as the solution to a second-stage utility maximization problem:

$$x_i(D_i, Z_i) = \underset{x \in B(D_i, Z_i, \boldsymbol{\varepsilon}_i)}{\operatorname{argmax}} \; u(x; D_i, Z_i, \boldsymbol{\varepsilon}_i)$$

where D_i, and Z_i are defined as before, $\boldsymbol{\varepsilon}_i$ is a vector of potentially unobservable covariates for individual i, and $B(D_i, Z_i, \boldsymbol{\varepsilon}_i)$ is the choice set (or budget set) from which x is chosen. The choice set is potentially constrained by treatment status, D_i, the value of the instrument, Z_i, and elements of the vector $\boldsymbol{\varepsilon}_i$. Note that this maximization takes the treatment status, D_i, as given. We separately define $D_i(Z_i)$ as the result of a first stage maximization exercise:

$$D_i(Z_i) = \underset{D \in G(Z_i, \boldsymbol{\varepsilon}_i)}{\operatorname{argmax}} \; v_i(D, Z_i)$$

where $G(Z_i, \boldsymbol{\varepsilon}_i)$ is the choice set from which D_i is chosen, potentially constrained by the instrument, Z_i, and unobservables, $\boldsymbol{\varepsilon}_i$, and $v_i(D_i, Z_i)$ is the value function of the second-stage maximization problem:

$$v_i(D_i, Z_i) \equiv \max_{x \in B(D_i, Z_i, \boldsymbol{\varepsilon}_i)} u(x; D_i, Z_i, \boldsymbol{\varepsilon}_i)$$

Note, the model can accommodate alternative approaches commonly found in the treatment effects literature. For example, the model can capture a Roy framework where treatment status is endogenous and related to the potential outcomes. Alternatively, the model also captures settings where choice is trivial and the relationship between instrument and treatment status is purely mechanical or probabilistic — e.g. the choice set, G, may be a singleton $G(Z_i, \boldsymbol{\varepsilon}_i) = \{\mathbf{1}\{\varepsilon_i + \alpha \cdot Z_i > 0\}\}$.

3 Violations of the Exclusion Restriction

We now turn to the question of when the general model in Section 2 collapses to the standard case in Section 1. In particular, we maintain Assumptions 1.1,

[1] Note, the two-part nature of the optimization problem is solely for illustrative purposes. The solution to the problem is equivalent when treatment status and the inputs are jointly determined.

1.3 and 1.4 and examine when Assumption 1.2, the exclusion restriction, may be violated. First, note that sufficient conditions for the exclusion restriction are:

$$\begin{aligned} B(D_i, 1, \boldsymbol{\varepsilon}_i) &= B(D_i, 0, \boldsymbol{\varepsilon}_i) = B(D_i, \boldsymbol{\varepsilon}_i) \\ u(x_i; D_i, 1, \boldsymbol{\varepsilon}_i) &= u(x_i; D_i, 0, \boldsymbol{\varepsilon}_i) = u(x_i; D_i, \boldsymbol{\varepsilon}_i). \end{aligned}$$

We therefore focus on cases where one of these sufficient conditions do not hold, and examine whether the exclusion restriction continues to hold. To illustrate the key intuition behind this observation, we proceed with two simple and stylized cases of the general model in Section 2. The examples are not intended to provide an exhaustive set of cases where the exclusion restriction may be violated, but rather, to convince the reader that in very common settings where the exclusion restriction is invoked, we may have reason to question the assumption on the basis of first principles. In each case, the agent will solve some modified version of the following, simple labor supply problem:

$$\max_{c,l} u(c, l)$$
$$\text{s.t. } c = wl + I$$

where c is consumption, l is labor, w are wages and I is non-labor income. Let $x_i = (c_i^*(w_i, I_i), l_i^*(w_i, I_i))$ be the optimal level of consumption and labor supply chosen in this setting, generally functions of wages and other income. The outcome of interest will be earnings — i.e. $y_i = w_i l_i^*(w_i, I_i)$. The treatment of interest will be some discrete measure of education, D_i, and the casual relationship of interest will be the return to education on earnings. We will consider two cases based on the instrument used: subsidies for higher education and compulsory school laws.

3.1 Example 1: Subsidies for Higher Education

Let Z_i represent a subsidy that covers a share, $1 - \theta$, of tuition for students and is randomly offered to a subset of students, i.e. those for whom $Z_i = 1$. Treatment status in this case, D_i, is a dummy variable that indicates whether or not a student attends college. Let college tuition be p. Individuals who attend college experience an increase in wages by a factor of ϕ_i. Individuals are also endowed with a source of non-labor income, I_i. Conditional on the instrument, Z_i, and the decision of whether or not to attend college, D_i, individuals solve the following utility maximization problem:

$$\max_{c,l} u(c, l)$$
$$\text{s.t. } c + \theta^{Z_i} p D_i = w_i \phi_i^{D_i} l + I_i$$

Utility maximization yields an optimal level of consumption and labor supply, $c_i^*\left(w_i\phi_i^{D_i}, I_i - \theta^{Z_i} p D_i\right)$ and $l_i^*\left(w_i\phi_i^{D_i}, I_i - \theta^{Z_i} p D_i\right)$, respectively. The decision

to attend college can be modeled by a first-stage decision:

$$\max_{D\in\{0,1\}} v\left(D, Z_i\right) = u\left(c_i^*\left(D, Z_i, \boldsymbol{\varepsilon}_i\right), l_i^*\left(D, Z_i, \boldsymbol{\varepsilon}_i\right)\right)$$

In terms of the notation above in Section 2, we have:

$$\begin{aligned}
\boldsymbol{\varepsilon}_i &= \left(\phi_i, w_i, I_i\right) \\
B\left(D_i, Z_i, \boldsymbol{\varepsilon}_i\right) &= \left\{\left(c, l\right) | c + \theta^{Z_i} p D_i = w_i \phi_i^{D_i} l + I_i\right\} \\
G\left(Z_i, \boldsymbol{\varepsilon}_i\right) &= \{0, 1\} \\
x_i &= \left(c_i, l_i\right) \\
y_i &= w_i \phi_i^{D_i} l_i.
\end{aligned}$$

Given Assumption 1.4, there are three sets of individuals: *compliers*, *always-takers* and *never-takers*. Consider, always-takers, i.e. those for whom $D_i\left(1\right) = D_i\left(0\right) = 1$. These individuals will attend college with or without the subsidy. We can show that the exclusion restriction will generally be violated for these individuals:

$$\begin{aligned}
y_i\left(D_i\left(1\right), 1\right) &= w_i \phi_i l^*\left(w_i \phi_i, I_i - \theta p\right) \\
&\leq w_i \phi_i l_i^*\left(w_i \phi_i, I_i - p\right) \\
&= y_i\left(D_i\left(0\right), 0\right),
\end{aligned}$$

where the second line follows from the assumption that leisure is a normal good. Intuitively, the always-taker is an inframarginal college-going student, for whom a college subsidy generates an income effect of $\left(1 - \theta\right) p$. Except for the case where there are no income effects on labor supply, we may expect the exclusion restriction to be violated for the always-taker. Note as well that this example extends to cases where there is not only a partial offset of costs, but in-kind transfer of a good such as college education, e.g. the case where $\theta = 0$ and we evaluate $x^0 \equiv 1$ when $x = 0$.

3.2 Example 2: Compulsory School Laws

Now let Z_i be an indicator for whether or not and individual's day of birth requires her by law to complete a given year of school, e.g. she must remain in school until her 18th birthday, the last day of school during her senior year is at time t, and her 18th birthday does not occur until some date beyond time t. Let D_i be a dummy variable indicating whether or not a student finishes her senior year. Suppose that if an individual disobeys the law, there is some disutility, proporitional to μ_i, that is experienced. This may represent some fine or penalty, guilt or stigma associated with being caught out of school before age 18. Additionally, let individuals allocate their labor supply, l_i, between working and receiving a market wage or putting forth evasion effort, e_i, to mitigate the disutility associated with disobeying compulsory school laws. Finally, let the additional year of schooling during the senior year increase market wages

by a factor of ϕ_i. Conditional on the bite of the school law, Z_i, and the decision to attend school during the senior year, D_i, individuals solve the following maximization problem:

$$\max_{c,l,e} u(c,l) - Z_i (1 - D_i) \frac{\mu_i}{e}$$
$$\text{s.t. } c = w_i \phi_i^{D_i} (l - e) + I_i$$

The additional term in the objective function captures the fact that the disutility of being caught disobeying compulsory school laws is only relevant for those who are both subject to the laws, $Z_i = 1$, and who are not obeying the law, $D_i = 0$. Utility maximization once again yields optimal consumption, labor supply and evasion, c_i^*, l_i^* and e_i^*, respectively.

In terms of the notation in Section 2, we have:

$$\begin{aligned}
\boldsymbol{\varepsilon}_i &= (\phi_i, w_i, I_i, \mu_i) \\
B(D_i, Z, \boldsymbol{\varepsilon}_i) &= \left\{ (c,l,e) \,|c = w_i \phi_i^{D_i} (l - e) + I_i \right\} \\
G(Z_i, \boldsymbol{\varepsilon}_i) &= \{0, 1\} \\
x_i &= (c_i, l_i, e_i) \\
y_i &= w_i \phi_i^{D_i} (l_i - e_i).
\end{aligned}$$

Now consider the never-takers, for whom $D_i(1) = D_i(0) = 0$. These individuals will choose to drop out of school whether or not the compulsory school laws apply. We can show that the exclusion restriction will now generally be violated for these individuals:

$$\begin{aligned}
y_i(D_i(1), 1) &= w_i \cdot [l_i^*(w_i, I_i - w_i e_i^*) - e_i^*] \\
&\leq w_i l_i^*(w_i, I_i) \\
&= y_i(D_i(0), 0),
\end{aligned}$$

where in the first line, we use the fact that the disutility of disobeying the law enters into utility in an additively separable fashion, which allows us to represent labor supply as a function of the wage and residual income. In the second line, we can show that earnings are lower whenever $e_i^* > 0$ if we assume that consumption is a normal good. In this case, we can view the never takers as an inframarginal group, with respect to dropping out of high school. Since she will do so either way, when she is additionally subject to the compulsory school law, the never taker bears the additional burden of disobeying the law.[2] In this case, we have a fine or penalty that is used as an instrument for treatment status. In such instances, those never takers who incur the penalty may potentially violate the exclusion restriction.

[2]Note, we have chosen a specific way in which effort to avoid geing caught disobeying affects the time available to earn income. However, we could have more generally modeled this effort as a nonseparable argument of utility that potentially affects the marginal rate of substitution between labor supply and consumption.

4 Adjusting for the Potential Bias

4.1 Adjustment via Calibration

The general bias involved with these violations of the exclusion restriction may vary considerably. However, in some cases, we may be able to at least sign the bias and/or potentially quantify it. We can then adjust traditional IV estimates for this bias. We demonstrate this in the case of subsidies for higher education. In this setting, we assume that outcomes for the never-takers, i.e. those for whom $D_i(1) = D_i(0) = 0$ are not affected by the instrument Z_i. In other words, the effect of the subsidy on the choice set of those who do not go to college is irrelevant. We show in Appendix A that, in this case, the Wald estimator of the LATE can be expressed as follows:

$$\frac{\mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \beta_{LATE} + \frac{\pi_{AT}}{\pi_C}\eta_{AT}, \tag{2}$$

where:

$$\begin{aligned}
\beta_{LATE} &= \mathbb{E}[y_i(1,1) - y_i(0,0)\,|D_i(1) > D_i(0)] \\
\pi_{AT} &= \Pr(D_i(1) = D_i(0) = 1) \\
\pi_C &= \Pr(D_i(1) > D_i(0)) \\
\eta_{AT} &= \mathbb{E}[y_i(1,1) - y_i(1,0)\,|D_i(1) = D_i(0) = 1]
\end{aligned}$$

The parameter of interest, β_{LATE}, is the local average treatment effect among compliers. The Wald estimator is biased by a term that is a function of π_{AT}, the share of always-takers, π_C, the share of compliers and η_{AT}, the effect of the instrument on the outcomes among always-takers. We may not be able to directly estimate η_{AT}; however, we can, in some cases, perform a back-of-the-envelope calculation of this parameter. Returning to the example in Section 3.1, the difference in earnings among always-takers is driven by an income effect, i.e. $\eta_{AT} = \partial y/\partial I \times (1-\theta)\,p$. If leisure is a normal good, we might expect the sign of the bias in (2) to be negative. However, if relief from the cost of tuition allows for great human capital investment or more effective job search, the effect may very well be positive.

Suppose that we use $\ln(y)$ as our outcome variable, so that the parameter β_{LATE} may be interpreted as the percent increase in earnings as a result of attending college. We can draw on existing estimates from Rothstein and Rouse (2011), who study the impact of a phase-out of loans at a college university. They find that replacing loans with grants for students who qualify for financial aid reduces earnings by \$2,657 annually, from a base of \$51,025 —i.e. $\eta_{AT} \approx -0.05$. Suppose we assume that tuition subsidies of this type increase college attendance by about 50 percent (see e.g. Dynarski, 2003). This implies that $\pi_{AT}/\pi_C = 2$. Finally, assume that we have estimated an IV estimate of $\beta_{IV} = 0.22$ (see e.g. Zimmerman, 2014). Then, using Equation 2 above, the Wald estimator will be biased downward by 0.1. After adjusting for the bias, we have

a LATE estimate of $\beta_{LATE} = \beta_{IV} - \frac{\pi_{AT}}{\pi_C}\eta_{AT} = 0.32$. The estimate is now 45 percent higher.

Thus, adjusting for the bias from the violation of the exclusion restriction may very well have a significant effect on one's estimates. All things equal, this bias is increasing in the share of always-takers, π_{AT}, i.e. the baseline level of college attendance. The bias is decreasing in the share of compliers, π_C — or put another way, decreasing in the size of the first-stage impact of the subsidy on college-going. The impact of the instrument is larger when tuition subsidies are significant relative to income, e.g. in developing countries. However, the type of subsidy that produces a relatively large π_C will also tend to generate a larger η_{AT} in magnitude, which works in the opposite direction and exacerbates the bias.

4.2 Adjustment via Experimental Design

In the case where the instrument Z_i is varied experimentally, the use of a second instrument may aid in estimating any potential bias associated with a violation of the exclusion restriction. We again demonstrate this in the case of the subsidies for higher education. Suppose that the tuition subsidy Z_i randomly assigned to a set of treatment students. Furthermore, suppose that the researcher is able to introduce a new, orthogonal cash instrument, T_i, which is a binary variable that indicates whether a student is given a cash transfer of $(1-\theta)\,p$ dollars. By virtue of random assignment, we assume that T_i satisfies the independence assumption, i.e. Assumption 1.1. In addition, we will need to assume that treatment participation, D_i, is not affected directly by this instrument, T_i, in a manner similar to the original exclusion restriction, i.e. Assumption 1.2. We now define the treatment, $D_i(Z_i, T_i)$, and potential outcomes, $y_i(D_i, Z_i, T_i)$, as functions of both instruments, Z_i and T_i. Formally, we assume the following:[3]

Assumption 2 (Cash Instrument Assumptions)

2.1. Independence: $\{y_i(D_i(1,1),1,1), y_i(D_i(0,0),0,0), y_i(D_i(1,0),1,0), y_i(D_i(0,1),0,1), D_i(1,1), D_i(0,0), D_i(1,0), D_i(0,1)\} \perp\!\!\!\perp (Z_i, T_i)$;

2.2. Exclusion: $D_i(z,1) = D_i(z,0) \equiv D_i(z)$ *for* $z = 0, 1$.

The variation in this additional treatment, T_i, can be used to identify the kind of income effects that may be causing a violation of the exclusion restriction. To see this, note that in the context of the model in Section 3.1, the budget set is now:

$$B(D_i, Z_i, T_i, \boldsymbol{\varepsilon}_i) = \left\{(c,l)\,|c + \theta^{Z_i} p D_i = w_i \phi_i^{D_i} l + I_i + (1-\theta)\,p T_i\right\}$$

[3]Note, Assumption 2.2 has a testable implication — namely, $\mathbb{E}[D_i|T_i = 1] = \mathbb{E}[D_i|T_i = 0]$. We can also make a slightly weaker assumption, i.e. $D_i(0,1) = D_i(0,0) = D_i(0)$, which carries an analogous implication — i.e. $\mathbb{E}[D_i|Z_i = 0, T_i = 1] = \mathbb{E}[D_i|Z_i = 0, T_i = 0]$.

It follows that among always-takers, i.e. those for whom $D_i(1) = D_i(0) = 1$, we have the following:

$$\begin{aligned} y_i(D_i(0), 0, 1) &= w_i\phi_i l^*(w_i\phi_i, I_i - p + (1-\theta)pT_i) \\ &= w_i\phi_i l^*(w_i\phi_i, I_i - \theta p) \\ &= y_i(D_i(1), 1, 0) \end{aligned} \tag{3}$$

Thus, we may estimate the effect of the instrument among always-takers, η_{AT}, using the following:

$$\begin{aligned} &\mathbb{E}[y_i|Z_i = 0, T_i = 1, D_i = 1] - \mathbb{E}[y_i|Z_i = 0, T_i = 0, D_i = 1] \\ &= \mathbb{E}[y_i(D_i(0), 0, 1)|Z_i = 0, T_i = 1, D_i = 1] - \mathbb{E}[y_i(D_i(0), 0, 0)|Z_i = 0, T_i = 0, D_i = 1] \\ &= \mathbb{E}[y_i(D_i(0), 0, 1)|D_i(0) = 1] - \mathbb{E}[y_i(D_i(0), 0, 0)|D_i(0) = 1] \\ &= \mathbb{E}[y_i(D_i(0), 0, 1)|D_i(1) = D_i(0) = 1] - \mathbb{E}[y_i(D_i(0), 0, 0)|D_i(1) = D_i(0) = 1] \\ &= \mathbb{E}[y_i(D_i(1), 1, 0)|D_i(1) = D_i(0) = 1] - \mathbb{E}[y_i(D_i(0), 0, 0)|D_i(1) = D_i(0) = 1] \\ &= \eta_{AT}, \end{aligned}$$

where in the third line we make use of independence and the monotonicity assumption, i.e. Assumptions 1.1 and 1.4, and in the fourth line, we make use of equation (3). In this way, augmenting an experimental design with an additional instrument that aids in identifying the income effects may allow one to adjust for the bias.

5 Conclusion

We have shown in two very specific examples that the exclusion restriction commonly assumed in applied settings may be plausibly violated. In particular, we have shown a case where the instrument used involves a subsidy that encourages adoption of a treatment, e.g. pursuing college education. In this setting, inframarginal members of the treatment group, i.e. always-takers, will tend to violate the exclusion restriction. Subsidies for this group relax the budget constraint through income effects associated with the instrument. In another case where a penalty discourages not adopting a treatment, e.g. dropping out of high school, we demonstrated that inframarginal members of the control group, i.e. never-takers, are likely to violate the exclusion restriction. Penalties incurred among this group effectively tighten the budget constraint and/or alter outcomes via direct effects on utility and marginal rates of substitution.

It should be noted that although behavioral and psychological effects of instruments can always be constructed by allowing the value of the instrument to directly enter into the utility function, the result here relies on much more traditional economic incentives — income effects associated with subsidies or costly risks and evasive measures associated with breaking the law. Furthermore, the specific nature of these examples is intended to facilitate straightforward illustration of the key argument. The underlying intuition is readily extended to a

much broader class of settings, many of which are commonly considered in the applied literature.

The powerful results that allow us to apply a well-defined casual interpretation to IV estimates as LATEs in part hinge on the implications of the exclusion restriction. Variation in the instrument only alters treatment status for the set of compliers. Since the exclusion restriction implies that the outcome of interest can only be affected by the instrument via its impact on treatment status, the variation in outcomes associated with the instrument can therefore be associated with compliers. Put another way, the numerator in equation (1) is only driven by differences in outcomes among compliers. However, as is demonstrated above, one must take care to note that the appropriateness of this result depends on whether the instrument in question is likely to alter outcome for inframarginal treatment group and control group members as well, even though treatment status is constant for these individuals (Angrist, Imbens and Rubin, 1996). In some cases, additional information regarding income effects, either taken from external studies or built into an experimental design, may help to adjust IV estimates for this potential bias.

References

Angrist, J D, and J S Pischke. 2008. "Mostly harmless econometrics: An empiricist's companion."

Angrist, Joshua D, and William N Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review*, 88(3): 450–477.

Angrist, Joshua D, Guido W Imbens, and Donald B Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444–455.

Dynarski, Susan M. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." *American Economic Review*, 93(1): 279–288.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *The Journal of Human Resources*, 32(3): 441.

Imbens, Guido W, and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*, 62(3): 467–475.

Pinto, Rodrigo. 2015. "Selection Bias in a Controlled Experiment: The Case of Moving to Opportunity." Mimeo.

Rosenzweig, Mark R, and Kenneth I Wolpin. 2000. "Natural "Natural Experiments" in Economics." *Journal of Economic Literature*, 38(4): 827–874.

Rothstein, Jesse, and Cecilia Elena Rouse. 2011. "Constrained after college: Student loans and early-career occupational choices." *Journal of Public Economics*, 95(1–2): 149–163.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics*, 32(4): 711–754.

Appendix A: Derivation of Bias Formula

We maintain Assumptions 1.1, 1.3 and 1.4 and evaluate the following Wald estimator:

$$\frac{\mathbb{E}\left[y_i|Z_i=1\right]-\mathbb{E}\left[y_i|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]}$$

We can express the first term in the numerator as follows:

$$\begin{aligned}
\mathbb{E}\left[y_i|Z_i=1\right] &= \Pr\left(D_i(1)=D_i(0)=1|\,Z_i=1\right)\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)=D_i(0)=1,Z_i=1\right] \\
&\quad+\Pr\left(D_i(1)>D_i(0)|\,Z_i=1\right)\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)>D_i(0),Z_i=1\right] \\
&\quad+\Pr\left(D_i(1)<D_i(0)|\,Z_i=1\right)\cdot\mathbb{E}\left[y_i(0,1)\,|D_i(1)<D_i(0),Z_i=1\right] \\
&\quad+\Pr\left(D_i(1)=D_i(0)=0|\,Z_i=1\right)\cdot\mathbb{E}\left[y_i(0,1)\,|D_i(1)=D_i(0)=0,Z_i=1\right] \\
&= \Pr\left(D_i(1)=D_i(0)=1\right)\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)=D_i(0)=1\right] \\
&\quad+\Pr\left(D_i(1)>D_i(0)\right)\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)>D_i(0)\right] \\
&\quad+\Pr\left(D_i(1)<D_i(0)\right)\cdot\mathbb{E}\left[y_i(0,1)\,|D_i(1)<D_i(0)\right] \\
&\quad+\Pr\left(D_i(1)=D_i(0)=0\right)\cdot\mathbb{E}\left[y_i(0,1)\,|D_i(1)=D_i(0)=0\right] \\
&= \pi_{AT}\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)=D_i(0)=1\right] \\
&\quad+\pi_C\cdot\mathbb{E}\left[y_i(1,1)\,|D_i(1)>D_i(0)\right] \\
&\quad+\pi_{NT}\cdot\mathbb{E}\left[y_i(0,1)\,|D_i(1)=D_i(0)=0\right],
\end{aligned}$$

where in the second step, we made use of Assumption 1.1, in the third step we used the fact that Assumption 1.4 implies $\Pr\left(D_i(1)<D_i(0)\right)=0$, and the π parameters are defined as follows:

$$\begin{aligned}
\pi_{AT} &= \Pr\left(D_i(1)=D_i(0)=1\right) \\
\pi_C &= \Pr\left(D_i(1)>D_i(0)\right) \\
\pi_{NT} &= \Pr\left(D_i(1)=D_i(0)=0\right).
\end{aligned}$$

Using very similar steps, we can show that:

$$\begin{aligned}
\mathbb{E}\left[y_i|Z_i=0\right] &= \pi_{AT}\cdot\mathbb{E}\left[y_i(1,0)\,|D_i(1)=D_i(0)=1\right] \\
&\quad+\pi_C\cdot\mathbb{E}\left[y_i(0,0)\,|D_i(1)>D_i(0)\right] \\
&\quad+\pi_{NT}\cdot\mathbb{E}\left[y_i(0,0)\,|D_i(1)=D_i(0)=0\right]
\end{aligned}$$

Again, we assume in the case of a school subsidy that the instrument, Z_i, will not have an effect on the outcomes of never takers, i.e.:

$$\mathbb{E}\left[y_i(0,0)\,|D_i(1)=D_i(0)=0\right]=\mathbb{E}\left[y_i(0,1)\,|D_i(1)=D_i(0)=0\right]$$

Put together, we have the following expression for the numerator of the Wald estimator:

$$\begin{aligned}\mathbb{E}\left[y_i|Z_i=1\right]-\mathbb{E}\left[y_i|Z_i=0\right] &= \pi_{AT}\cdot\mathbb{E}\left[y_i\left(1,1\right)-y_i\left(1,0\right)|D_i\left(1\right)=D_i\left(0\right)=1\right] \\ &\quad +\pi_C\cdot\mathbb{E}\left[y_i\left(1,1\right)-y_i\left(0,0\right)|D_i\left(1\right)>D_i\left(0\right)\right]\end{aligned} \tag{A.1}$$

Turning to the first term in the denominator, we can show the following:

$$\begin{aligned}\mathbb{E}\left[D_i|Z_i=1\right] &= \Pr\left(D_i\left(1\right)=D_i\left(0\right)=1|\,Z_i=1\right)\cdot\Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)=D_i\left(0\right)=1,Z_i=1\right) \\ &\quad +\Pr\left(D_i\left(1\right)>D_i\left(0\right)|\,Z_i=1\right)\cdot\Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)>D_i\left(0\right),Z_i=1\right) \\ &\quad +\Pr\left(D_i\left(1\right)<D_i\left(0\right)|\,Z_i=1\right)\cdot\Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)<D_i\left(0\right),Z_i=1\right) \\ &\quad +\Pr\left(D_i\left(1\right)=D_i\left(0\right)=0|\,Z_i=1\right)\cdot\Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)=D_i\left(0\right)=0,Z_i=1\right) \\ &= \Pr\left(D_i\left(1\right)=D_i\left(0\right)=1\right)+\Pr\left(D_i\left(1\right)>D_i\left(0\right)\right) \\ &= \pi_{AT}+\pi_C,\end{aligned}$$

where in the second step, we once again made use of Assumption 1.1 and also used the fact that

$$\begin{aligned}\Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)=D_i\left(0\right)=1,Z_i=1\right) &= 1, \\ \Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)>D_i\left(0\right),Z_i=1\right) &= 1, \\ \Pr\left(D_i\left(1\right)=1|\,D_i\left(1\right)=D_i\left(0\right)=0,Z_i=1\right) &= 0\end{aligned}$$

and, by Assumption 1.4,

$$\Pr\left(D_i\left(1\right)<D_i\left(0\right)|\,Z_i=1\right)=0.$$

Likewise, we can show that:

$$\mathbb{E}\left[D_i|Z_i=0\right]=\pi_{AT}$$

Thus, the denominator of the Wald is:

$$\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]=\pi_C \tag{A.2}$$

Combining Equations A.1 and A.2, we have:

$$\begin{aligned}\frac{\mathbb{E}\left[y_i|Z_i=1\right]-\mathbb{E}\left[y_i|Z_i=0\right]}{\mathbb{E}\left[D_i|Z_i=1\right]-\mathbb{E}\left[D_i|Z_i=0\right]} &= \frac{\pi_C}{\pi_C}\mathbb{E}\left[y_i\left(1,1\right)-y_i\left(0,0\right)|D_i\left(1\right)>D_i\left(0\right)\right] \\ &\quad +\frac{\pi_{AT}}{\pi_C}\mathbb{E}\left[y_i\left(1,1\right)-y_i\left(1,0\right)|D_i\left(1\right)=D_i\left(0\right)=1\right] \\ &= \beta_{LATE}+\frac{\pi_{AT}}{\pi_C}\eta_{AT}.\end{aligned}$$