DISCUSSION PAPER SERIES

IZA

IZA DP No. 3980

Comparing IV with Structural Models: What Simple IV Can and Cannot Identify

James J. Heckman
Sergio Urzúa

January 2009

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Comparing IV with Structural Models: What Simple IV Can and Cannot Identify

James J. Heckman
University of Chicago, University College Dublin,
Cowles Foundation, Yale University,
American Bar Foundation and IZA

Sergio Urzúa
Northwestern University

Discussion Paper No. 3980
January 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3980
January 2009

ABSTRACT

Comparing IV with Structural Models: What Simple IV Can and Cannot Identify*

This paper compares the economic questions addressed by instrumental variables estimators with those addressed by structural approaches. We discuss Marschak's Maxim: estimators should be selected on the basis of their ability to answer well-posed economic problems with minimal assumptions. A key identifying assumption that allows structural methods to be more informative than IV can be tested with data and does not have to be imposed.

JEL Classification: C31

Keywords: instrumental variables, structural approaches, Marschak's Maxim

Corresponding author:

James J. Heckman
Department of Economics
University of Chicago
1126 E. 59th Street
Chicago, IL 60637
USA
E-mail: jjh@uchicago.edu

* We thank Donna Gilleskie, Daniel Schmierer, Greg Veramendi and an anonymous referee for helpful comments. This project was supported by NSF grant SES-0241858, NIH grant R01-HD043411, a grant from the Geary Institute, University College Dublin, and a grant from the American Bar Foundation.

1 Introduction

The primary question regarding the choice of an empirical approach to analyzing economic data should be *What economic question does the analyst seek to answer?* Explicit economic models make it easier to formulate and answer economic questions. Advocates of atheoretical approaches to analyzing economic data appeal to randomization as an ideal and invoke IV (or matching or regression discontinuity designs) as a surrogate for randomization. However, even perfectly executed randomizations do not answer all questions of economic interest. There are important examples where structural models produce more information about preferences than experiments.[1]

A valid instrument is not guaranteed to identify parameters of economic interest when responses to choices vary among individuals, and these variations influence choices taken.[2] Different valid instruments answer different questions.[3] The sign of the IV estimator can be different from that of the true causal effect.[4]

No one trained in economics can doubt the value of credible, explicit economic models in interpreting economic data. They are designed to answer a variety of well-posed economic questions and to be invariant to classes of policy interventions.[5] The problem with this approach is that after 60 years of experience with fitting structural models on a variety of data sources, empirical economists have come to appreciate the practical difficulty in identifying, and precisely estimating, the full array of structural parameters that answer the large variety of policy questions contemplated by the Cowles Commission economists — the fathers of structural estimation.[6]

Proponents of IV are less ambitious in the range of questions they seek to answer. The method often gains precision by asking narrower questions. The problem that plagues the IV approach is that the questions it answers are usually *defined* as probability limits of estimators and not by well-formulated economic problems. Unspecified "effects" replace clearly defined economic parameters

[1] See Heckman (1992, 2008) and Heckman and Vytlacil (2007a,b).

[2] See Heckman and Vytlacil (2007b) for a comparison of what different approaches identify. Matching rules out selection on unobservables. Regression discontinuity estimators identify, at best, local effects.

[3] Building on this point, Heckman, Schmierer, and Urzua (2008) develop and apply nonparametric tests for the presence of heterogenous responses to treatment on which agents make choices.

[4] Heckman, Urzua, and Vytlacil (2008). This is true even under the "monotonicity" conditions of Imbens and Angrist (1994).

[5] See the essays in Koopmans (1950). See also Heckman and Vytlacil (2007a).

[6] Pencavel (1986) summarizes structural estimates of taxes on labor supply and reports absurd estimates, including one due to Jerry Hausman, which implied a negative marginal propensity to consume goods out of income. These and other estimates, reported in the literature some 20 years ago, fueled the flight of many empirical economists from structural models.

as the objects of empirical interest.

As noted by Marschak (1953), there is a middle ground. *Marschak's Maxim* emphasizes that one should solve well-posed economic problems with minimal assumptions. Marschak noted that for many problems of policy analysis, it is not necessary to identify fully specified structural models with parameters that are invariant to classes of policy modifications — the goal of structural analysis as conceived by the Cowles pioneers and successor generations of structural economists.[7] All that is required to conduct many policy analyses or to answer many well-posed economic questions are *combinations* of the structural parameters that are often much easier to identify than the individual parameters themselves.[8]

Heckman and Vytlacil (1999, 2005) bridge the structural and IV literatures. They develop an economically interpretable parameter — the marginal treatment effect (MTE) — which is a marginal willingness to pay for the benefit of treatment for persons at well-defined margins of choice. It is invariant to a class of policy modifications. They show how different instrumental variables weight the MTE differently. These weights need not be positive for all values of the argument of the MTE.[9] For classes of well-posed economic questions, it is possible, under the conditions given in Heckman and Vytlacil (2005), to fashion functions of instruments that answer well-posed economic questions. This approach is unusual in the standard IV literature, which traditionally defines the parameter of interest to be an "effect" identified by an instrument.

Many economists follow Imbens and Angrist (1994) and interpret IV as identifying a weighted average of the gains to persons induced to change their choice (or state) by a change in the instrument. Imbens and Angrist work with a two choice model. Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b) extend this analysis to an ordered choice model and to general unordered choice models.[10] This paper develops the unordered case further and gives a

[7] Heckman and Vytlacil (2007a) define structural models precisely following the seminal definition of Hurwicz (1962). These discussions formalize ideas in Marschak (1953).

[8] See the discussion in Heckman (2008) and Heckman and Vytlacil (2007b).

[9] MTE was introduced into the literature in the context of a selection model by Björklund and Moffitt (1987). The Local Average Treatment Effect (LATE) (Imbens and Angrist, 1994) is a discretized version of the MTE. The weights for special cases were derived by Yitzhaki (1989) and applied by Imbens and Angrist (1994). Those weights are always non-negative. Heckman and Vytlacil (1999, 2005, 2007b) and Heckman, Urzua, and Vytlacil (2006) generalize the Imbens-Angrist-Yitzhaki analysis to the case of multiple instruments without restrictions and show how IV weights can be non-positive over certain intervals, but that they must integrate to one.

[10] Angrist and Imbens (1995) propose an ordered choice version of their 1994 paper. As shown by Heckman, Urzua, and Vytlacil (2006), their proposed extension has unsatisfactory features which can be removed by a careful reformulation of the IV method applied to the ordered choice model. See also Heckman and Vytlacil (2007b).

precise characterization of the generalization of the MTE that is appropriate for this case.

The original Imbens-Angrist intuition applies, but in general unordered choice models, agents attracted into a state by a change in an instrument come from many origin states, so there are many margins of choice. Structural models can identify the gains arising from these separate margins. This is a difficult task for IV without invoking structural assumptions. Structural models can also identify the fraction of persons induced into a state coming from each origin state. IV alone cannot.

For some economic questions, these are unimportant distinctions. For others, they are crucial. For specificity, consider an analysis of the GED program. The GED is a test by which high school dropouts can exam certify to be the equivalents of ordinary high school graduates. Heckman, LaFontaine, and Rodríguez (2008) show that the presence of a GED program induces some persons to drop out of high school. It also induces some persons who would remain dropouts to exam certify. Within this context we ask: *What are the wage benefits for those induced to take the GED from the dropout state? For those induced to drop out of high school? What proportion of persons induced to take a GED come from each of the other states?*

IV cannot answer these questions except under structural assumptions. It *can* identify the mean gross gain to the GED for those induced to take it, compared to the next best alternative. This is a weighted average of the effects from each possible origin state that the structural approach can separately identify. In the IV approach, when there are multiple origin states, the weights on the individual effects cannot be estimated without using structural methods. As shown in Heckman and Vytlacil (2007b), IV needs to be supplemented with explicit choice theory to answer many interesting questions, including questions of economic welfare regarding introduction of policies as well as distributional questions such as the percentage of persons harmed by a policy.

This paper demonstrates these points. We first establish a precise framework for discussing IV, and relating it to economic models.

2 The Choice Model and Assumptions

Following Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b), consider the following model with multiple choices and associated multiple outcome states. Let $\mathcal{J}$ denote the agent's choice set, where $\mathcal{J}$ contains a finite number of elements. For example, $\mathcal{J}$ enumerates possible schooling states (e.g., GED, high school dropout, high school graduate). The value to the agent of choosing $j \in \mathcal{J}$ is

$$R_j(Z_j) = \vartheta_j(Z_j) - V_j, \tag{2.1}$$

where Z_j are the agent's observed characteristics that affect the utility from choosing j, and V_j is the unobserved shock to the agent's utility from choice j. We sometimes write R_j for $R_j(Z_j)$ to simplify notation. Let Z denote the random vector containing all unique elements of $\{Z_j\}_{j \in \mathcal{J}}$. We write $R_j(Z)$ for $R_j(Z_j)$, leaving implicit the condition that $R_j(\cdot)$ only depends on the elements of Z that are contained in Z_j. Let D_j be a variable indicating whether the agent would choose j if confronted with choice set $\mathcal{J}$:[11]

$$D_j = \begin{cases} 1 & \text{if } R_j \geq R_k \quad \forall\, k \in \mathcal{J} \\ 0 & \text{otherwise.} \end{cases}$$

Array the D_j into a vector D. Let Y be the outcome that would be observed if the agent faced choice set $\mathcal{J}$, defined as

$$Y = \sum_{j \in \mathcal{J}} D_j Y_j,$$

where Y_j is a potential outcome observed only if option j is chosen. Y_j is determined by

$$Y_j = \mu_j(X_j, U_j),$$

where X_j is a vector of the agent's observed characteristics and U_j is an unobserved random vector. Let X denote the random vector containing all unique elements of $\{X_j\}_{j \in \mathcal{J}}$. (Z, X, D, Y) is assumed to be observed by the analyst.[12]

[11] Below, we invoke conditions so that ties, $R_j = R_k$ for $j \neq k$, occur with probability zero.

[12] Depending on the choice model, Z may or may not include the X. For example, in a Roy model of schooling under perfect certainty (e.g. Willis and Rosen, 1979), X would be among the Z. In models of schooling under uncertainty (e.g. Cunha, Heckman, and Navarro, 2005, Cunha and Heckman, 2007 and Urzua, 2008) innovations in X unknown at the time schooling decisions are made would not be in Z. The key condition on Z is given in Assumption (A-2) below.

Define $R_{\mathcal{J}}$ as the maximum obtainable value given choice set $\mathcal{J}$:

$$\begin{aligned} R_{\mathcal{J}} &= \max_{j\in\mathcal{J}}\{R_j\} \\ &= \textstyle\sum_{j\in\mathcal{J}} D_j R_j. \end{aligned} \tag{2.2}$$

This is the traditional representation of the decision process that if choice j is optimal, choice j is better than the "next best" option:

$$D_j = 1 \Longleftrightarrow R_j \geq R_{\mathcal{J}\setminus j}.$$

Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b) show that this simple, well-known, representation is the key intuition for understanding how instrumental variables estimate the effect of a given choice versus the "next best" alternative. IV is a weighted average of the effects for people induced into a choice from different margins. Analogous to the definition of $R_{\mathcal{J}}$, we define $R_{\mathcal{J}}(z)$ to be the maximum obtainable value given choice set $\mathcal{J}$ when instruments are fixed at $Z = z$,

$$R_{\mathcal{J}}(z) = \max_{j\in\mathcal{J}}\{R_j(z)\}.$$

Following the analysis in Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b), we assume:

(A-1) The distribution of $(\{V_j\}_{j\in\mathcal{J}})$ is continuous.[13]

(A-2) $\{(V_j, U_j)\}_{j\in\mathcal{J}}$ is independent of Z conditional on X.

(A-3) $E \mid Y_j| < \infty$ for all $j \in \mathcal{J}$.

(A-4) $\Pr(D_j = 1 \mid X) > 0$ for all $j \in \mathcal{J}$.

In addition, we assume an exclusion restriction that requires some additional notation.[14] Let $Z^{[-l]}$ denote all elements of Z except for the lth component. We assume

[13] Absolutely continuous with respect to Lebesgue measure on $\prod_{j\in\mathcal{J}}\Re$.

[14] We work here with exclusion restrictions in part for ease of exposition. By adapting the analysis of Cameron and Heckman (1998) and Heckman and Navarro (2007), one can modify our analysis for the case of no exclusion restrictions if Z contains a sufficient number of continuous variables and there is sufficient variation in the ϑ_k function across k.

(A-5) For each $j \in \mathcal{J}$, their exists at least one element of Z, say $Z^{[l]}$, such that the distribution of $\vartheta_j(Z_j)$ conditional on $(X, Z^{[-l]})$ is continuous.

With these assumptions, one can generalize the analysis of Heckman and Vytlacil (1999, 2001, 2005) to the unordered case. Assumptions (A-1) and (A-2) imply that $R_j \neq R_k$ (with probability 1) for $j \neq k$, so that $\text{argmax}_{j\in\mathcal{J}}\{R_j\}$ is unique (with probability 1). Assumption (A-2) assures the existence of an instrument. Assumption (A-3) is required for mean treatment parameters to be well defined. It also allows one to integrate to the limit and to produce well-defined means. Assumption (A-4) requires that at least some individuals participate in each choice for all X. Assumption (A-5) imposes the requirement that one be able to independently vary the index for the given value function. It imposes a type of exclusion restriction, that for any $j \in \mathcal{J}$, Z contains an element such that (i) it is contained in Z_j; (ii) it is not contained in any Z_k for $k \neq j$, and (iii) $\vartheta_j(\cdot)$ is a nontrivial function of that element conditional on all other regressors.[15]

In a series of papers, Heckman and Vytlacil (1999, 2001, 2005, 2007b), develop the method of local instrumental variables (LIV) to estimate the marginal treatment effect (MTE) for the case of binary choices. We now define and interpret the MTE and LIV in the case of general unordered choices.

3 Interpreting Local Instrumental Variables in the Unordered Case

We define local instrumental variables (LIV) using a variable that shifts people toward (or against) choice j by operating only on $R_j(Z_j)$. LIV identifies an average marginal return to j vs. the next best alternative across persons.[16] However, without further assumptions, LIV will not decompose the average marginal return into its component parts corresponding to the effects for persons induced into j from each of the possible origin states.

To see this, consider a three outcome case, $\mathcal{J} = \{1, 2, 3\}$. For concreteness, we pursue the education example previously stated and let 1 be GED, 2 be high school dropout, and 3 be high school graduate. Our results are more general but the three outcome case is easy to exposit.

In this section, we assume that Z_1, Z_2, Z_3 are disjoint sets of regressors so $Z = (Z_1, Z_2, Z_3)$

[15]See Heckman and Vytlacil (2007b) for additional discussion.

[16]See Heckman, Urzua, and Vytlacil (2006) and Heckman and Vytlacil (2007b).

but they are not necessarily statistically independent. We can easily relax this assumption but it simplifies the notation. We condition on X and keep it implicit throughout the analysis of this paper.[17] In this notation,

$$\begin{aligned} E(Y \mid Z) &= E\left[\left.\sum_{j=1}^{3} Y_j D_j \right| Z\right] \\ &= E(Y_1 D_1 \mid Z) + E(Y_2 D_2 \mid Z) + E(Y_3 D_3 \mid Z). \end{aligned} \tag{3.1}$$

$E(Y|Z)$ and its components can be estimated from data on (Y, Z). IV is based on (3.1). From (2.2), choices are generated by the following inequalities:

$$\begin{aligned} D_1 &= \mathbf{1}(R_1 \geq R_2, R_1 \geq R_3) \\ D_2 &= \mathbf{1}(R_2 \geq R_1, R_2 \geq R_3) \\ D_3 &= \mathbf{1}(R_3 \geq R_1, R_3 \geq R_2). \end{aligned}$$

We define the marginal change in Y with respect to Z_1. IV methods are based on such types of variation. The local instrumental variable estimator using Z_1 as an instrument is the sample analogue of

$$\left.\frac{\frac{\partial E(Y|Z)}{\partial Z_1}}{\frac{\partial \Pr(D_1=1|Z)}{\partial Z_1}}\right|_{Z=z} = \mathrm{LIV}(z),$$

where LIV is a function of z. In the case of three choices, there are two margins from which persons can be attracted into or out of choice 1 by Z_1.[18]

From local variations in Z_1, one can recover the following combinations of parameters from the

[17] See Heckman and Vytlacil (2007b) for a more general analysis.

[18] Recall that Z_1 only affects the utility associated with choice 1.

data on $Y_1 D_1$:

$$\begin{aligned}
&\frac{\partial E\left(Y_1 D_1 \mid Z=z\right)}{\partial Z_1} \\
&\quad= \frac{\partial}{\partial Z_1} \int \int_{-\infty}^{\vartheta_1(Z_1)-\vartheta_2(Z_2)} \int_{-\infty}^{\vartheta_1(Z_1)-\vartheta_3(Z_3)} y_1 f_{Y_1, V_1-V_2, V_1-V_3}\left(y_1, v_1-v_2, v_1-v_3\right) d\left(v_1-v_3\right) d\left(v_1-v_2\right) dy_1 \Big|_{Z=z} \\
&\quad= \left.\frac{\partial \vartheta_1\left(Z_1\right)}{\partial Z_1}\right|_{Z_1=z_1} \left[\begin{array}{c} \int y_1 \int_{-\infty}^{\vartheta_1(z_1)-\vartheta_3(z_3)} f_{Y_1, V_1-V_2, V_1-V_3}\left(y_1, \vartheta_1\left(z_1\right)-\vartheta_2\left(z_2\right), v_1-v_3\right) d\left(v_1-v_3\right) dy_1 \\ +\int y_1 \int_{-\infty}^{\vartheta_1(z_1)-\vartheta_2(z_2)} f_{Y_1, V_1-V_2, V_1-V_3}\left(y_1, v_1-v_2, \vartheta_1\left(z_1\right)-\vartheta_3\left(z_3\right)\right) d\left(v_1-v_2\right) dy_1 \end{array}\right].
\end{aligned} \tag{3.2}$$

By similar reasoning, we can recover the following combination of parameters from the data on $Y_2 D_2$:

$$\begin{aligned}
&\frac{\partial E\left(Y_2 D_2 \mid Z=z\right)}{\partial Z_1} \\
&\quad= \frac{\partial}{\partial Z_1} \int y_2 \int_{-\infty}^{\vartheta_2(Z_2)-\vartheta_1(Z_1)} \int_{-\infty}^{\vartheta_2(Z_2)-\vartheta_3(Z_3)} f_{Y_2, V_2-V_1, V_2-V_3}(y_2, v_2-v_1, v_2-v_3)\, d\left(v_2-v_3\right)\, d\left(v_2-v_1\right) dy_2 \Big|_{Z=z} \\
&\quad= \left.\frac{-\partial \vartheta_1\left(Z_1\right)}{\partial Z_1}\right|_{Z_1=z_1} \left[\int y_2 \int_{-\infty}^{\vartheta_2(z_2)-\vartheta_3(z_3)} f_{Y_2, V_2-V_1, V_2-V_3}(y_2, \vartheta_2(z_2)-\vartheta_3(z_1), v_2-v_3) d\left(v_2-v_3\right) dy_2\right].
\end{aligned} \tag{3.3}$$

From data on $Y_3 D_3$, we obtain the following combination of parameters:

$$\begin{aligned}
&\frac{\partial E\left(Y_3 D_3 \mid Z_1=z\right)}{\partial Z_1} \\
&\quad= \left.\frac{-\partial \vartheta_1\left(Z_1\right)}{\partial Z_1}\right|_{Z_1=z_1} \int y_3 \int_{-\infty}^{\vartheta_3(z_3)-\vartheta_2(z_2)} f_{Y_3, V_3-V_1, V_3-V_2}\left(y_3, \vartheta_3\left(z_3\right)-\vartheta_1\left(z_1\right), v_3-v_2\right)\, d\left(v_3-v_2\right)\, dy_3\, .
\end{aligned} \tag{3.4}$$

Agents induced into 1 come from 2 and 3. There are two margins:

$$\left(R_1=R_2\right) \text{ and } \left(R_1 \geq R_3\right) \qquad \text{(margin of indifference between 1 and 2),}$$

and

$$\left(R_1=R_3\right) \text{ and } \left(R_1 \geq R_2\right) \qquad \text{(margin of indifference between 1 and 3).}$$

Unaided, IV does not enable analysts to identify the returns at each of the different margins. Instead, it identifies a weighted average of returns. It does not identify the density of persons at the various margins, i.e., the proportion of people induced into (or out of) 1 from each possible alternative state by a change in the instrument.

Collecting terms and rewriting in more easily interpretable components, which generalize the MTE developed for a two choice model to a multiple choice unordered model:[19]

$$\left.\frac{\left(\frac{\partial E(Y|Z)}{\partial Z_1}\right)}{\left(\frac{\partial \vartheta_1}{\partial Z_1}\right)}\right|_{Z=z} =$$
$$\left[\begin{array}{l} \overbrace{\left[E\left(Y_1 - Y_2 \mid R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3)\right)\right]}^{\substack{\text{Generalization of MTE for persons indifferent} \\ \text{between 1 and 2, where choice 3 is dominated}}} \Pr\left(R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3)\right) \\ + \underbrace{\left[E\left(Y_1 - Y_3 \mid R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2)\right)\right]}_{\substack{\text{Generalization of MTE for persons indifferent} \\ \text{between 1 and 3, where choice 2 is dominated}}} \Pr\left(R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2)\right) \end{array}\right].$$

This is a weighted return to alternative 1 for persons coming from two separate margins: alternative 1 *versus* alternative 2, and alternative 1 *versus* alternative 3, i.e., the return to people induced into 1 from their next best choice. The weights are the proportion of people induced into 1 from each margin. This *combination* of parameters can be identified from IV. The components of the sum cannot be identified by IV without further assumptions. Note that it is possible that a group at one margin gains while a group at another margin loses. IV only estimates a net effect, which might be zero.

Notice that from representation (2.1) and the assumption that the Z_j $(j \epsilon \mathcal{J})$ are distinct, *pairwise monotonicity*, an extension of the monotonicity assumption invoked by Imbens and Angrist (1994) for the binary choice case, is satisfied.[20] In the context of a model with multiple choices, pairwise monotonicity means the same pattern of flow between any two states is experienced by everyone. Thus, as Z_j increases, there is a flow from i to j but not from j to i (or vice versa). From (2.1), changing Z_1 induces all persons to move in the same direction (*i.e.* from 1 to 2 or 2 to 1 but not both, and from 1 to 3 or 3 to 1 but not both). Pairwise monotonicity does not rule out the

[19] Heckman, Urzua, and Vytlacil (2006) generalize the MTE to an ordered choice model. See also Heckman and Vytlacil (2007b).

[20] This is defined as "uniformity" in Heckman, Urzua, and Vytlacil (2006).

possibility that a change in an instrument causes people to move in the direction from j to i but to move away from the direction from k to i for $j \neq k$, and $j, k \neq i$.

By the chain rule, the derivative of $\Pr(D_1 = 1 \mid Z)$ is:

$$\frac{\partial \Pr(D_1 = 1 \mid Z = z)}{\partial Z_1} = \left.\frac{\partial \vartheta_1}{\partial Z_1}\right|_{Z_1 = z_1} \left[\begin{array}{c} \Pr(R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3)) \\ + \Pr(R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2)) \end{array} \right].$$

We can define LIV in terms of the preceding ingredients as

$$\mathrm{LIV}(z) = \left.\frac{\left(\frac{\partial E(Y|Z)}{\partial Z_1}\right)}{\left(\frac{\partial \Pr(D_1 = 1|Z)}{\partial Z_1}\right)}\right|_{Z=z} = \left[\begin{array}{c} E(Y_1 - Y_2 \mid R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3))\,\omega_{12} \\ + E(Y_1 - Y_3 \mid R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2))\,\omega_{13} \end{array} \right]. \tag{3.5}$$

The *combination* of terms can be identified by LIV from the data on (Y, D, Z).

The IV weights are:

$$\omega_{12} = \frac{\Pr(R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3))}{\left[\begin{array}{c} \Pr(R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3)) \\ + \Pr(R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2)) \end{array} \right]} \tag{3.6}$$

$$\tag{3.7}$$

and

$$\omega_{13} = \frac{\Pr(R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2))}{\left[\begin{array}{c} \Pr(R_1(z_1) = R_2(z_2), R_1(z_1) \geq R_3(z_3)) \\ + \Pr(R_1(z_1) = R_3(z_3), R_1(z_1) \geq R_2(z_2)) \end{array} \right]}. \tag{3.8}$$

The weights can be identified from a structural discrete choice analysis.[21] They cannot be identified by an unaided instrumental variable analysis. Thus it is not possible to identify the component

[21]Conditions for nonparametric identification of the multinomial discrete choice model are presented in Matzkin (1993, 1994). Conditions for nonparametric identification of the full choice model with outcomes are given in Heckman and Vytlacil (2007a, Appendix B). Conditions for identification of general dynamic discrete choice models are presented in Abbring and Heckman (2007).

parts of (3.3) by LIV alone, i.e., one cannot separately identify the generalized MTEs:

$$E\left(Y_1 - Y_2 \mid R_1\left(z_1\right) = R_2\left(z_2\right), R_1\left(z_1\right) \geq R_3\left(z_3\right)\right)$$

and

$$E\left(Y_1 - Y_3 \mid R_1\left(z_1\right) = R_3\left(z_3\right), R_1\left(z_1\right) \geq R_2\left(z_2\right)\right),$$

unless one invokes "identification at infinity" arguments.[22]

Using a structural model, one can estimate the components of (3.5) and determine the flow into (or out of) state 1 from all sources. We illustrate this point in Section 5. First we consider what standard IV estimates.

4 What does standard IV estimate?

To see what standard IV estimates, consider the following linear-in-schooling model of earnings that receives much attention in the literature in labor economics.[23] Let Y denote log earnings and write S as years of schooling. The model writes

$$Y = \alpha + \beta S + U \tag{4.1}$$

where

$$S = \sum_{j=1}^{3} j D_j, \tag{4.2}$$

and Y is defined as in Section 2. It is interpreted in this section as an approximation to the general model presented in Section 2. S is assumed to be correlated with U, and β is a random variable that may be statistically dependent on S. The model of Section 2 does *not*, in general, imply (4.1).

[22]See Heckman and Vytlacil (2007b) who show how to vary Z_3 or Z_2 to effectively shut down one margin of choice. Specifically, for any fixed $Z_1 = z_1$, if $\lim_{Z_2 \to \tilde{Z}_2} R_2(Z_2) \to -\infty$ and $\lim_{Z_3 \to \tilde{Z}_3} R_3(Z_3) \to -\infty$ where $\tilde{Z}_2$ and $\tilde{Z}_3$ represent limit sets, then we can identify, respectively, the gains at the $3 \to 1$ margin in the limit set, and the gains in the $2 \to 1$ margin in the limit set. These assumptions require that one can vary Z_2 and Z_3 to shut down one or the other margin of choice. Under these assumptions and some additional mild regularity assumptions, the structural approach can identify distributions of $(Y_1 - Y_2)$ and $(Y_2 - Y_3)$ as we demonstrate in the example in Section 5 of this paper.

[23]We keep conditioning on X implicit.

Indeed, there is much empirical evidence against model (4.1).[24] An analysis of what IV estimates when linearity in S is imposed as an approximation, even though it may be inappropriate, is an interesting exercise because linearity is so often invoked.

Suppose Z_1 is a valid instrument. We now interpret what

$$\Delta_{Z_1}^{\mathrm{IV}} = \frac{\mathrm{Cov}(Z_1, Y)}{\mathrm{Cov}(Z_1, S)} \tag{4.3}$$

estimates. We do this by decomposing $\Delta_{Z_1}^{\mathrm{IV}}$ into components analogous to the decomposition produced by Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b). The Appendix presents the derivation of the following decomposition of IV into our pairwise generalization of MTE for the unordered case:

$$\Delta_{Z_1}^{\mathrm{IV}} = \frac{\mathrm{Cov}(Z_1, Y)}{\mathrm{Cov}(Z_1, S)} = \tag{4.4}$$

$$\frac{\left(\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overbrace{E\left(Y_1 - Y_2 \mid V_2 - V_1 = v_2 - v_1, \vartheta_2(z_2) - \vartheta_3(z_3) \geq V_2 - V_3\right)}^{\text{Generalized MTE } (2 \to 1) \text{ not identified from LIV}} \\ \quad \times \underbrace{\eta_{\vartheta_2(Z_2) - \vartheta_3(Z_3), V_2 - V_1}\left(\vartheta_2(z_2) - \vartheta_3(z_3), v_2 - v_1\right)}_{\substack{\text{weight identified from discrete} \\ \text{choice analysis}}} d(v_2 - v_1)\, d\left(\vartheta_2(z_2) - \vartheta_3(z_3)\right) \end{array}\right) + \left(\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \overbrace{E\left(Y_1 - Y_3 \mid V_3 - V_1 = v_3 - v_1, \vartheta_3(z_3) - \vartheta_2(z_2) \geq V_3 - V_2\right)}^{\text{Generalized MTE} (3 \to 1) \text{ not identified from LIV}} \\ \quad \times \underbrace{\eta_{\vartheta_3(Z_3) - \vartheta_2(Z_2), V_3 - V_1}\left(\vartheta_3(z_3) - \vartheta_2(z_2), v_3 - v_1\right)}_{\substack{\text{weight identified from discrete} \\ \text{choice analysis}}} d(v_3 - v_1)\, d\left(\vartheta_3(z_3) - \vartheta_2(z_2)\right) \end{array}\right)}{\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\left[-\eta_{\vartheta_2(Z_2) - \vartheta_3(Z_3), V_2 - V_1}\left(\vartheta_2(z_2) - \vartheta_3(z_3), v_2 - v_1\right)\right]}_{\substack{\text{weight identified from discrete} \\ \text{choice analysis}}} d\left(\vartheta_2(z_2) - \vartheta_3(z_3)\right) d(v_2 - v_1) \\ +2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\left[-\eta_{\vartheta_3(Z_3) - \vartheta_2(Z_2), V_3 - V_1}\left(\vartheta_3(z_3) - \vartheta_2(z_2), v_3 - v_1\right)\right]}_{\substack{\text{weight identified from discrete} \\ \text{choice analysis}}} d(v_3 - v_1)\, d\left(\vartheta_3(z_3) - \vartheta_2(z_2)\right). \end{array}}$$

IV identifies a weighted average of gains to state 1 compared to the next best alternative which may be 2 or 3. The two terms of the decomposition are defined as generalized MTEs and are weighted averages of the gain of moving from state 2 to state 1 for persons on the margin of

[24] See Heckman, Lochner, and Todd (2006) for discussions of this model and various justifications for it. Heckman, Layne-Farrar, and Todd (1996) present evidence against linearity of the earnings function in terms of years of schooling.

indifference between 1 and 2 and for whom 2 is a better choice than 3 (the first term) and the gain of moving from 3 to 1 for persons on the margin of indifference between 1 and 3 and for whom 3 is a better choice than 2 (the second term).[25]

In the Appendix, we derive the weights on the generalized MTEs and show that they do not sum to 1 even when normalized by the denominator. The mathematical reason for this result is simple. The weights in the numerator do not sum to the weights in the denominator. The second term in the denominator receives twice as much weight as the corresponding term in the numerator. This is a consequence of the definition of S (4.2), which plays no role in the numerator term. Thus, IV applied to the general model produces an arbitrarily weighted sum of generalized MTEs with weights that do not sum to 1, and which, in general, places more weight on the first generalized MTE term than on the second term, compared to the weights placed on the corresponding terms in the denominator.[26] Using IV alone, we cannot decompose (4.4) into its component parts, even though the weights can be identified from discrete choice analysis.[27,28]

[25] Since Z_1 only affects $R(Z_1)$, it has no direct effect on the margin $2 \to 3$.

[26] Thus "2" appears only in the denominator and not in the numerator.

[27] The structural model is nonparametrically identified under the conditions in Appendix B of Heckman and Vytlacil (2007a).

[28] Decomposition (4.4) is not unique. It arises from decomposing Y into

$$Y = D_1 Y_1 + D_2 Y_2 + D_3 Y_3$$

where we solve out $D_1 = 1 - D_2 - D_3$, to obtain

$$Y = Y_1 + D_2(Y_2 - Y_1) + D_3(Y_3 - Y_1).$$

We could also solve out $D_2 = 1 - D_1 - D_3$ to obtain

$$Y = Y_2 + D_1(Y_1 - Y_2) + D_3(Y_3 - Y_2)$$

or $D_3 = 1 - D_1 - D_2$ to obtain

$$Y = Y_3 + D_1(Y_1 - Y_3) + D_2(Y_2 - Y_3).$$

Each decomposition can be used to represent $\Delta^{IV}_{Z_1}$. For each decomposition, the leading terms on the right-hand side, (Y_1, Y_2, Y_3), respectively, are uncorrelated with Z_1 by virtue of (A-2). Corresponding generalized MTEs can be defined for each decomposition. Z_1 affects the lower boundary of the opportunity set in

$$E(Y_2 - Y_3 | R(z_2) \geq R(z_1), R(z_3) \geq R(z_1)).$$

We choose the decomposition reported in the text for its greater interpretability.

4.1 The Mincer Model

The Mincer (1974) model is a specialization of the general model discussed in Section 2 of this paper that justifies the precise functional form of equation (4.1).[29] For this case, the weights in (4.4) in the numerator and denominator are the same. The Mincer model is formulated in terms of log earnings for $Y_1, Y_2,$ and Y_3:

$$\begin{aligned} Y_2 &= \ln(1+g) + Y_1, \\ Y_3 &= \ln(1+g) + Y_2 = 2\ln(1+g) + Y_1, \end{aligned}$$

where g is a growth factor for income that varies in the population. Earnings at each schooling level depend on two parameters: (g, Y_1). In this case, letting $\alpha = \ln(1+g)$,

$$\begin{aligned} \Delta^{\mathrm{IV}}_{Z_1} &= \frac{\operatorname{Cov}(Z_1, Y)}{\operatorname{Cov}(Z_1, S)} \qquad (4.6) \\ &= \left[\left(\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left(\alpha \mid V_2 - V_1 = v_2 - v_1, \vartheta_2(z_2) - \vartheta_3(z_3) \geq V_2 - V_3\right) \\ \quad \times \eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3), V_2-V_1}\left(\vartheta_2(z_2) - \vartheta_3(z_3), v_2 - v_1\right) d\left(v_2 - v_1\right) d\left(\vartheta_2(z_2) - \vartheta_3(z_3)\right) \end{array} \right) \right. \\ &\quad \left. + \left(\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left(\alpha \mid V_3 - V_1 = v_3 - v_1, \vartheta_3(z_3) - \vartheta_2(z_2) \geq V_3 - V_2\right) \\ \quad \times 2\eta_{\vartheta_3(Z_3)-\vartheta_2(Z_2), V_3-V_1}\left(\vartheta_3(z_3) - \vartheta_2(z_2), v_3 - v_1\right) d\left(v_3 - v_1\right) d\left(\vartheta_3(z_3) - \vartheta_2(z_2)\right) \end{array} \right) \right] \\ &\times \left[\begin{array}{l} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3), V_2-V_1}\left(\vartheta_2(z_2) - \vartheta_3(z_3), v_2 - v_1\right) d\left(v_2 - v_1\right) d\left(\vartheta_2(z_2) - \vartheta_3(z_3)\right) \\ \quad +2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \eta_{\vartheta_3(Z_3)-\vartheta_2(Z_2), V_3-V_1}\left(\vartheta_3(z_3) - \vartheta_2(z_2), v_3 - v_1\right) d\left(v_3 - v_1\right) d\left(\vartheta_3(z_3) - \vartheta_2(z_2)\right) \end{array} \right]^{-1}. \end{aligned}$$

In this case, the weights now sum to 1. The weights for the numerator term now are the same as the weights for the denominator term. But again, unaided IV does not identify the component parts of the term bundled in IV — the mean gains at each margin.[30,31]

[29] See Heckman, Lochner, and Todd (2006) for a discussion of the Mincer Model and the powerful body of evidence against it. Card (2001) provides one justification for functional form 4.1.

[30] An anonymous referee has correctly expressed the concern that in the case of an income-maximizing Mincer model under perfect certainty, the general unordered model would not apply. Indeed, the decision problem is not well defined. If $g > r$, the opportunity cost of funds, agents would choose the maximum amount of schooling. If $g < r$, the agent chooses no schooling. If $g = r$, the agent is indifferent to all levels of schooling. Thus, for our analysis to apply to the Mincer earnings equation, we have to assume that choices involve some combination of psychic costs, tuition, uncertainty or the like. The model of Keane and Wolpin (1997) is one of many frameworks that would justify an unordered choice model but could be consistent with a Mincer earnings equation. See Heckman, Lochner, and Todd (2006). We thank the referee for emphasizing this point to us.

[31] Decomposition (4.6) for an ordered choice model is presented in Heckman, Urzua, and Vytlacil (2006) and Heckman and Vytlacil (2007b).

5 An Example

It is instructive to summarize our analysis with an example. Consider a 3 choice model with associated outcomes. This corresponds to the GED, high school dropout and high school graduate example that we have used throughout the paper. Under conditions presented in Heckman and Vytlacil (2007a, Appendix B), the structural model is nonparametrically identified. A key assumption in their proof is the "identification at infinity" assumption previously discussed.[32] This assumes the ability to vary (Z_1, Z_2, Z_3) freely and the existence of limit sets such that fixing any two of (Z_1, Z_2, Z_3), one makes the R_j associated with Z_j arbitrarily small.[33]

Heckman and Vytlacil (2007b) show that if one augments the IV assumptions with the same identification at infinity assumptions used in structural models, one can use IV in the limit to identify the components of (3.5). In the limit sets, one can identify

$$E(Y_1 - Y_2 | R_1(z_1) = R_2(z_2)) \tag{5.1}$$

and

$$E(Y_1 - Y_3 | R_1(z_1) = R_3(z_3)) \tag{5.2}$$

by setting Z_3 and Z_2 respectively to limit set values. Essentially one can use the limit sets to make a three choice model into a two choice model, and the standard results for the two choice model apply.[34] Under these assumptions, and additional mild regularity assumptions, using structural methods, one can identify the distributions of (Y_1, Y_2) and (Y_1, Y_3) so that one can identify *distributions* of treatment effects, $Y_2 - Y_1$ and $Y_3 - Y_1$, in addition to the mean parameters identified by IV.[35] One can also identify the proportion of people induced into 1 from each alternative state using variation in the instrument.

Consider the model with the parameters presented in Table 1. This is a discrete choice model with associated outcome variables. The $Z_j, j = 1, \ldots, 3$, are assumed to be scalar and mutually independent. They are normally distributed so they satisfy large support ("identification at infin-

[32] Alternatively, one can make functional form assumptions about the distribution of the error terms.

[33] See the conditions in footnote 22.

[34] See Heckman and Vytlacil (2007b)

[35] The literature on "quantile treatment effects" uses IV to identify the quantiles of Y_1 and Y_2 separately but not the quantiles of $Y_1 - Y_2$. See Abbring and Heckman (2007).

ity") conditions. Table 2 shows how a change in Z_1, which increases it by .75 standard deviations, shifts people across categories. This corresponds to making GED attainment easier.[36] The estimates reported in Table 2 can be obtained from a structural discrete choice model. The percentage initially in 1 (GED) increases from 33.17% to 38.8%. The percentage in 2 (dropout) decreases from 29.11% to 25.91%. The percentage in 3 (graduating high school) declines from 37.72% to 35.29%.

The IV estimate is -.032. (See the base of Table 3) This is the only number produced by an IV analysis using Z_1 as an instrument that changes within the specified range. The structural analysis in Table 3 shows that the net effect produced by the change in Z_1 is composed of 2 terms. It arises from a gain of .199 for the switchers $2 \to 1$ (dropout to GED) and a loss of .336 ($3 \to 1$) (graduate to GED).

Figure 1 shows what can be identified from the structural model. It plots the distributions of gains for persons going from 2 to 1 and from 3 to 1 as well as the overall distribution of gains to the switchers. Persons switching from 3 to 1 are harmed in gross terms by the policy that changes Z_1, while those who switch from 2 to 1 gain in gross terms. In utility terms, (R_j), people are better off.[37] In terms of gross gains, about 56.8% of the people who switch from 2 to 1 are better off while 39.3% of the people who switch from 3 to 1 are better off. Overall, 49.2% are better off in gross terms even though the IV estimate is slightly negative. If one seeks to understand the distributional effects of the policy associated with a change Z_1, the structural analysis is clearly much more revealing. The IV estimate, which is a mean gross gain aggregating over origin states, does not capture the rich information about choices afforded by a structural analysis. However, it does identify the average gain to the program compared to the next-best alternatives. If that is the object of interest, linear IV is the right tool to use.

6 Summary and Discussion

The choice between using IV or a more structural approach for a particular problem should be made on the basis of Marschak's Maxim: use minimal assumptions to answer well-posed economic questions. Most IV studies do not clearly formulate the economic question being answered by the

[36]Heckman, LaFontaine, and Rodríguez (2008) show that easing GED requirements promotes dropping out of school and causes some dropouts to become GEDs.

[37]This is imposed in a discrete choice model.

IV analysis. The probability limit of the IV estimator is defined to be the object of interest. In the binary outcome case, even if Z is a valid instrument, if Z is a vector, and analysts use only one component of the vector as an instrument, and do not condition on the other components of Z, the weights on the MTE can be negative over certain ranges. The practice of not conditioning on the other instruments is common in the literature.[38] IV can estimate the wrong sign for the true causal effect.[39] Recent analyses show how to improve on this practice and to design functions of standard instrumental variables that answer classes of well-posed economic questions.[40]

We have discussed a model with three or more choices where there is no particular order among the choices. Such examples arise routinely in applied economics. In this case, under conditions specified in this paper, IV estimates a weighted average of the mean gross gain to persons induced into a choice state by a change in the instrument (policy) compared to their next best alternative.[41] It averages the returns to a destination state over all origin states. It does not produce the distribution of gains overall or by each origin state. Again, as in the binary choice case, for vector Z, using one component of Z as an instrument, and not conditioning on the other components can produce negative weights so that the sign of an IV can be opposite to that of the true causal effect which can be identified by a structural analysis.

Structural methods provide a more complete description of the effect of the instrument or the policy associated with the instrument. They identify mean returns as well as distributions of returns for agents coming to a destination state from each margin. They also identify the proportion of people induced into a state from each origin state.

Structural methods come at a cost. Unless distributional assumptions for unobservables are invoked, structural methods require some form of an "identification at infinity" assumption.[42] However, in the general case in which responses to treatment are heterogeneous, IV requires the same assumption if one seeks to identify average treatment effects.[43] An identification at infinity assumption can be checked in any sample so it does not require imposing *a priori* beliefs onto the data. Heckman, Stixrud, and Urzua (2006) present an example of how to test an identification at

[38] See e.g. Card (2001).
[39] See Heckman, Urzua, and Vytlacil (2006).
[40] See Heckman and Vytlacil (2005, 2007b).
[41] See Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b).
[42] See Heckman and Vytlacil (2007a, Appendix B).
[43] See Heckman, Urzua, and Vytlacil (2006, 2008) and Heckman and Vytlacil (2007b).

infinity assumption. See also the discussion in Abbring and Heckman (2007).[44]

Many proponents of IV point to the strong distributional and functional form assumptions required to implement structural methods. They ignore recent progress in econometrics that identifies and empirically implements robust semiparametric and nonparametric approaches to structural analysis.[45] Recent developments respond to arguments against the use of explicit econometric models made by a generation of applied economists that emerged in the 1980s. Those arguments are more properly directed against 1980s versions of structural models that were based on linearity and normality. Structural econometricians in the 21$^{\text{st}}$ century have listened to the critics and have perfected their tools to response to the criticism.

The appeal to standard IV as a preferred estimator is sometimes made on the basis of "simplicity and robustness". Standard IV is certainly simple to compute although problems with weak instruments can make it empirically unstable.[46] Since, in the general case, different instruments identify different parameters, IV is not robust to the choice of instrument.[47] Since the sign of an IV can be different from the true causal effect, IV may even produce a misleading guide to policy or inference, so it is not robust.

The meaning of "simplicity" is highly subjective. How simple is the economic interpretation of IV? Certainly decomposition (4.4) is not simple. The fact that simple IV can estimate wrong signs for true causal effects should give pause to those who claim that it is "robust". The weak instrument literature cautions us against uncritical claims about the sturdiness of IV estimators.

The ability of different statistical estimators to answer questions of economic interest, or to show why they cannot be answered, should drive the choice of empirical techniques for analyzing data. Consider a worst case for structural estimation. Suppose that application of recently developed procedures for testing for structural identification reveal that a structural model is not identified or is only partially identified. Does this conclusion suggest that IV is a better choice for an estimator? That disguising identification problems by a statistical procedure is preferable to an honest discussion of the limits of the data? For underidentified structural models, it is possible to

[44] However, this assumption is not yet routinely checked in many structural analyses.

[45] See e.g. Abbring and Heckman (2007), Carneiro, Hansen, and Heckman (2003), Cunha, Heckman, and Navarro (2007), Cunha, Heckman, and Schennach (2006, revised 2008), and Matzkin (1992, 1993, 2007). These recent developments in robust structural modeling have not yet made their way into widespread use in empirical structural analysis.

[46] See e.g. Stock and Staiger (1997) and the ensuing large literature on weak instruments.

[47] See Heckman, Urzua, and Vytlacil (2006).

conduct sensitivity analyses guided by economic theory to explore the consequences of ignorance about features of the model. With IV, unaided by structural analysis, this type of exercise is not possible. Problems of identification and interpretation are swept under the rug and replaced by "an effect" identified by IV that is often very difficult to interpret as an answer to an interesting economic question.

A Derivation of the Standard IV Estimator

We first study the numerator of $\Delta^{\mathrm{IV}}_{Z_1}$ in the text. Recall that we keep the conditioning on X implicit. Using $\tilde{Z}_1 = Z_1 - \bar{Z}_1$,

$$\operatorname{Cov}(Y, Z_1) = E\left(\tilde{Z}_1 \left(Y_1 D_1 + Y_2 D_2 + Y_3 D_3\right)\right).$$

Using $D_1 = 1 - D_2 - D_3$, we obtain

$$\begin{aligned}\operatorname{Cov}(Y, Z_1) &= E\left(\tilde{Z}_1 \left(Y_1 + \left(Y_2 - Y_1\right) D_2 + \left(Y_3 - Y_1\right) D_3\right)\right) \\ &= E\left(\tilde{Z}_1 Y_1\right) + E\left(\tilde{Z}_1 \left(Y_2 - Y_1\right) D_2\right) + E\left(\tilde{Z}_1 \left(Y_3 - Y_1\right) D_3\right),\end{aligned}$$

where $E\left(\tilde{Z}_1 Y_1\right) = 0$. It is natural to decompose this expression using choice "1" as the base, because Z_1 only shifts $R_1(Z_1)$. The final two terms can be written as

$$\begin{aligned}&\operatorname{Cov}(Y, Z_1) \\ &= E\left(\tilde{Z}_1 \left(Y_2 - Y_1\right) \mathbf{1}\left(R_2(Z_2) \geq R_1(Z_1), R_2(Z_2) \geq R_3(Z_3)\right)\right) + E\left(\tilde{Z}_1 \left(Y_3 - Y_1\right) \mathbf{1}\left(R_3(Z_3) \geq R_1(Z_1), R_3(Z_3) \geq R_2(Z_2)\right)\right) \\ &= E\left[\tilde{Z}_1 \left(Y_2 - Y_1\right) \mathbf{1}\left(\left(\vartheta_2\left(Z_2\right) - \vartheta_1\left(Z_1\right) \geq V_2 - V_1\right), \left(\vartheta_2\left(Z_2\right) - \vartheta_3\left(Z_3\right) \geq V_2 - V_3\right) \right)\right] \\ &\quad + E\left[\tilde{Z}_1 \left(Y_3 - Y_1\right) \mathbf{1}\left(\left(\vartheta_3\left(Z_3\right) - \vartheta_1\left(Z_1\right) \geq V_3 - V_1\right), \left(\vartheta_3\left(Z_3\right) - \vartheta_2\left(Z_2\right) \geq V_3 - V_2\right) \right)\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{z}_1\left(y_2 - y_1\right) \\ &\quad \times \left(\int_{-\infty}^{\vartheta_2(z_2)-\vartheta_1(z_1)} \int_{-\infty}^{\vartheta_2(z_2)-\vartheta_3(z_3)} f_{Y_2-Y_1, V_2-V_1, V_2-V_3}(y_2 - y_1, v_2 - v_1, v_2 - v_3)\, d\left(v_2 - v_3\right)\, d\left(v_2 - v_1\right)\, d\left(y_2 - y_1\right)\right) \\ &\quad \times f_{\tilde{Z}_1, \vartheta_2(Z_2)-\vartheta_1(Z_1), \vartheta_2(Z_2)-\vartheta_3(Z_3)}\left(\tilde{z}_1, \vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right), \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) d\left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) d\left(\vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right)\right) d\tilde{z}_1 \\ &\quad + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{z}_1\left(y_3 - y_1\right) \\ &\quad \times \left(\int_{-\infty}^{\vartheta_3(z_3)-\vartheta_1(z_1)} \int_{-\infty}^{\vartheta_3(z_3)-\vartheta_2(z_2)} f_{Y_3-Y_1, V_3-V_1, V_3-V_2}\left(y_3 - y_1, v_3 - v_1, v_3 - v_2\right) d\left(v_3 - v_2\right)\, d\left(v_3 - v_1\right)\, d\left(y_3 - y_1\right)\right) \\ &\quad \times f_{\tilde{Z}_1, \vartheta_3(Z_3)-\vartheta_1(Z_1), \vartheta_3(Z_3)-\vartheta_2(Z_2)}\left(\tilde{z}_1, \vartheta_3\left(z_3\right) - \vartheta_1\left(z_1\right), \vartheta_3\left(z_3\right) - \vartheta_2\left(z_2\right)\right) d\left(\vartheta_3\left(z_3\right) - \vartheta_2\left(z_2\right)\right) d\left(\vartheta_3\left(z_3\right) - \vartheta_1\left(z_1\right)\right) d\tilde{z}_1.\end{aligned}$$

By Fubini's Theorem, we can simplify the expressions and obtain for the first term:

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left(Y_2 - Y_1 \mid V_2 - V_1 = v_2 - v_1, \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right) \geq V_2 - V_3\right) \\
&\times\left\{\int_{-\infty}^{\infty} \tilde{z}_1 \left[\left(\int_{-\infty}^{\vartheta_2(z_2)-\vartheta_3(z_3)} h_{V_2-V_1,V_2-V_3}\left(v_2 - v_1, v_2 - v_3\right)\, d\left(v_2 - v_3\right)\right)\right.\right. \\
&\times\left.\left.\left(\int_{v_2-v_1}^{\infty} f_{\tilde{Z}_1,\vartheta_2(Z_2)-\vartheta_1(Z_1),\vartheta_2(Z_2)-\vartheta_3(Z_3)}\left(\tilde{z}_1, \vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right), \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) d\left(\vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right)\right)\right)\right] d\tilde{z}_1\right\} \\
&\qquad\qquad\qquad\qquad\times d\left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) d\left(v_2 - v_1\right). \qquad \text{(A.1)}
\end{aligned}$$

$h_{V_2-V_1,V_2-V_3}(.)$ is the joint density of V_2-V_1, V_2-V_3. Define the weighting term in braces in (A.1) as $\eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3),V_2-V_1}(\vartheta_2\left(z_2\right)-\vartheta_3\left(z_3\right), v_2-v_1)$. It is necessary to fix both $\vartheta_2\left(z_2\right)-\vartheta_3\left(z_3\right)$ and v_2-v_1 in forming the weight. This weight can be estimated from a structural discrete choice analysis and the joint distribution of (Z, D_1, D_2, D_3). The terms multiplying the weight are marginal treatment effects generalized to the unordered case. (A.1) cannot be decomposed using IV. An alternative representation of the term in braces, $\eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3),V_2-V_1}(\vartheta_2\left(z_2\right)-\vartheta_3\left(z_3\right), v_2-v_1)$ is

$$\begin{aligned}
&\eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3),V_2-V_1}\left(\vartheta_2\left(z_2\right)-\vartheta_3\left(z_3\right), v_2-v_1\right) = \\
&\quad E\left(Z_1 - E\left(Z_1\right) \mid \vartheta_2\left(Z_2\right) - \vartheta_3\left(Z_3\right) = \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right), \vartheta_2\left(Z_2\right) - \vartheta_1\left(Z_1\right) \geq v_2 - v_1\right) \\
&\qquad\qquad \times \Pr\left(\vartheta_2\left(Z_2\right) - \vartheta_3\left(Z_3\right) = \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right), \vartheta_2\left(Z_2\right) - \vartheta_1\left(Z_1\right) \geq v_2 - v_1\right).
\end{aligned}$$

An analysis parallel to the preceding one shows that the second term can be written as

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E\left(Y_3 - Y_1 \mid V_3 - V_1 = v_3 - v_1, \vartheta_3\left(z_3\right) - \vartheta_2\left(z_2\right) \geq V_3 - V_2\right) \\
&\times\left\{\int_{-\infty}^{\infty} \tilde{z}_1 \left[\left(\int_{-\infty}^{\vartheta_3(z_3)-\vartheta_2(z_2)} h_{V_3-V_1,V_3-V_2}\left(v_3 - v_1, v_3 - v_2\right)\, d\left(v_3 - v_2\right)\right)\right.\right. \\
&\times\left.\left.\left(\int_{v_3-v_1}^{\infty} f_{\tilde{Z}_1,\vartheta_3(Z_3)-\vartheta_1(Z_1),\vartheta_3(Z_3)-\vartheta_2(Z_2)}\left(\tilde{z}_1, \vartheta_3\left(z_3\right) - \vartheta_1\left(z_1\right), \vartheta_3\left(z_3\right) - \vartheta_2\left(z_2\right)\right) d\left(\vartheta_3\left(z_3\right) - \vartheta_1\left(z_1\right)\right)\right)\right] d\tilde{z}_1\right\} \\
&\qquad\qquad\qquad\qquad\times d\left(\vartheta_3\left(z_3\right) - \vartheta_2\left(z_2\right)\right) d\left(v_3 - v_1\right). \qquad \text{(A.2)}
\end{aligned}$$

Define the term in braces in (A.2) as the weight $\eta_{\vartheta_3(Z_3)-\vartheta_2(Z_2),V_3-V_1}(\vartheta_3\left(z_3\right)-\vartheta_2\left(z_2\right), v_3-v_1)$.

To obtain the denominator for the IV, recall that $S = \sum_{j=1}^{3} j\, D_j$. Substitute $D_1 = 1 - D_2 - D_3$,

$$\begin{aligned}\sum_{j=1}^{3} j\, D_j &= (1 - D_2 - D_3) + 2D_2 + 3D_3 \\ &= 1 + D_2 + 2D_3.\end{aligned}$$

Then

$$\begin{aligned}\mathrm{Cov}(S, \tilde{Z}_1) &= E\left(\tilde{Z}_1 D_2\right) + 2E\left(\tilde{Z}_1 D_3\right) \\ &= E\left(\tilde{Z}_1 \left(\mathbf{1}\left(R_2 \geq R_1, R_2 \geq R_3\right)\right)\right) \\ &\quad +2E\left(\tilde{Z}_1 \left(\mathbf{1}\left(R_3 \geq R_1, R_3 \geq R_2\right)\right)\right).\end{aligned} \tag{A.3}$$

Using reasoning similar to that invoked for the analysis of the numerator terms, we obtain expressions for the terms corresponding to the two terms of (A.1) and (A.2). We obtain for the first term of (A.3)

$$\begin{aligned}&\int_{-\infty}^{\infty} \tilde{z}_1 \Bigg[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\vartheta_2(z_2)-\vartheta_1(z_1)} f_{\tilde{Z}_1, \vartheta_2(Z_2)-\vartheta_1(Z_1), \vartheta_2(Z_2)-\vartheta_3(Z_3)} \left(\tilde{z}_1, \vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right), \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) \\ &\quad \times \left(\int_{-\infty}^{\vartheta_2(z_2)-\vartheta_3(z_3)} h_{V_2-V_1, V_2-V_3}\left(v_2 - v_1, v_2 - v_3\right)\, d\left(v_2 - v_3\right) \right) d\left(v_2 - v_1\right) \\ &\quad \times d\left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) d\left(\vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right)\right) \Bigg] d\tilde{z}_1.\end{aligned} \tag{A.4}$$

By Fubini's Theorem, we obtain:

$$\begin{aligned}&\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{z}_1 \Bigg[\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\vartheta_2(z_2)-\vartheta_3(z_3)} h_{V_2-V_1, V_2-V_3}\left(v_2 - v_1, v_2 - v_3\right)\, d\left(v_2 - v_3\right) \right) \\ &\quad \times \left(\int_{v_2-v_1}^{\infty} f_{\tilde{Z}_1, \vartheta_2(Z_2)-\vartheta_1(Z_1), \vartheta_2(Z_2)-\vartheta_3(Z_3)} \left(\tilde{z}_1, \vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right), \vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) \right. \\ &\quad \left. \times d\left(\vartheta_2\left(z_2\right) - \vartheta_1\left(z_1\right)\right) \right) d\left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right) \Bigg] d\left(v_2 - v_1\right)\, d\tilde{z}_1 \qquad \text{(A.5)} \\ = &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \eta_{\vartheta_2(Z_2)-\vartheta_3(Z_3), V_2-V_1} \left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right), v_2 - v_1\right) d\left(v_2 - v_1\right)\, d\left(\vartheta_2\left(z_2\right) - \vartheta_3\left(z_3\right)\right).\end{aligned}$$

By parallel logic, we obtain for the second term in A.3:

$$
\begin{aligned}
& 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\tilde{z}_1\Bigg[\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\vartheta_3(z_3)-\vartheta_2(z_2)} h_{V_3-V_1,V_3-V_2}\left(v_3-v_1,v_3-v_2\right)\,d\left(v_3-v_2\right)\right) \\
& \quad\times\left(\int_{v_3-v_1}^{\infty} f_{\tilde{Z}_1,\vartheta_3(Z_3)-\vartheta_1(Z_1),\vartheta_3(Z_3)-\vartheta_2(Z_2)}\left(\tilde{z}_1,\vartheta_3\left(z_3\right)-\vartheta_1\left(z_1\right),\vartheta_3\left(z_3\right)-\vartheta_2\left(z_2\right)\right)d\left(\vartheta_3\left(z_3\right)-\vartheta_1\left(z_1\right)\right)\right) \\
& \qquad\times d\left(\vartheta_3\left(z_3\right)-\vartheta_2\left(z_2\right)\right)\Bigg]\,d\left(v_3-v_1\right)\,d\tilde{z}_1 \\
= \; & 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\eta_{\vartheta_3(Z_3)-\vartheta_2(Z_2),V_3-V_1}\left(\vartheta_3\left(z_3\right)-\vartheta_2\left(z_2\right),v_3-v_1\right)\,d\left(v_3-v_1\right)\,d\left(\vartheta_3\left(z_3\right)-\vartheta_2\left(z_2\right)\right).
\end{aligned}
$$

These terms can be identified from a structural analysis using the joint distribution of (Z, D_1, D_2, D_3). Collecting results, we obtain decomposition (4.4) in the text if we multiply both the numerator and denominator by -1.

Acknowledgements

We thank Donna Gilleskie, Daniel Schmierer, Greg Veramendi and an anonymous referee for helpful comments. This project was supported by NSF grant SES-0241858, NIH grant R01-HD043411, a grant from the Geary Institute, University College Dublin, and a grant from the American Bar Foundation.

References

Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 5145–5303. Amsterdam: Elsevier.

Angrist, J. D. and G. W. Imbens (1995, June). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Björklund, A. and R. Moffitt (1987, February). The estimation of wage gains and welfare gains in self-selection. *Review of Economics and Statistics 69*(1), 42–49.

Cameron, S. V. and J. J. Heckman (1998, April). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy 106*(2), 262–333.

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

Carneiro, P., K. Hansen, and J. J. Heckman (2003, May). Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review 44*(2), 361–422.

Cunha, F. and J. J. Heckman (2007, October). The evolution of inequality, heterogeneity and uncertainty in labor earnings in the U.S. economy. *National Bureau of Economic Research Working Paper No. 13526.*

Cunha, F., J. J. Heckman, and S. Navarro (2005, April). Separating uncertainty from heterogeneity in life cycle earnings, The 2004 Hicks Lecture. *Oxford Economic Papers 57*(2), 191–261.

Cunha, F., J. J. Heckman, and S. Navarro (2007, November). The identification and economic content of ordered choice models with stochastic cutoffs. *International Economic Review 48*(4), 1273–1309.

Cunha, F., J. J. Heckman, and S. M. Schennach (2006). Meaningful value added equations: Anchoring test scores in adult outcomes. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J. (1992). Randomization and social policy evaluation. In C. Manski and I. Garfinkel (Eds.), *Evaluating Welfare and Training Programs*, pp. 201–230. Cambridge, MA: Harvard University Press.

Heckman, J. J. (2008, April). Econometric causality. *International Statistical Review 76*(1), 1–27.

Heckman, J. J., P. A. LaFontaine, and P. L. Rodríguez (2008). Taking the easy way out: How the GED testing program induces students to drop out. Unpublished manuscript, University of Chicago, Department of Economics.

Heckman, J. J., A. Layne-Farrar, and P. E. Todd (1996, November). Human capital pricing equations with an application to estimating the effect of schooling quality on earnings. *Review of Economics and Statistics 78*(4), 562–610.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings equations and rates of return: The Mincer equation and beyond. In E. A. Hanushek and F. Welch (Eds.), *Handbook of the Economics of Education*, Chapter 7, pp. 307–458. Amsterdam: Elsevier.

Heckman, J. J. and S. Navarro (2007, February). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics 136*(2), 341–396.

Heckman, J. J., D. Schmierer, and S. Urzua (2008). Testing the correlated random coefficient model. Unpublished manuscript, University of Chicago, Department of Economics under revision, *Journal of Econometrics.*

Heckman, J. J., J. Stixrud, and S. Urzua (2006, July). 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., S. Urzua, and E. J. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J., S. Urzua, and E. J. Vytlacil (2008). Instrumental variables in models with multiple outcomes: The general unordered case. *Les Annales d'Economie et de Statistique*. Forthcoming.

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

Heckman, J. J. and E. J. Vytlacil (2001). Local instrumental variables. In C. Hsiao, K. Morimune, and J. L. Powell (Eds.), *Nonlinear Statistical Modeling: Proceedings of the Thirteenth International Symposium in Economic Theory and Econometrics: Essays in Honor of Takeshi Amemiya*, pp. 1–46. New York: Cambridge University Press.

Heckman, J. J. and E. J. Vytlacil (2005, May). Structural equations, treatment effects and econometric policy evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007a). Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4779–4874. Amsterdam: Elsevier.

Heckman, J. J. and E. J. Vytlacil (2007b). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative economic estimators to evaluate social programs and to forecast their effects in new environments. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B, pp. 4875–5144. Amsterdam: Elsevier.

Hurwicz, L. (1962). On the structural form of interdependent systems. In E. Nagel, P. Suppes, and A. Tarski (Eds.), *Logic, Methodology and Philosophy of Science*, pp. 232–239. Stanford University Press.

Imbens, G. W. and J. D. Angrist (1994, March). Identification and estimation of local average treatment effects. *Econometrica 62*(2), 467–475.

Keane, M. P. and K. I. Wolpin (1997, June). The career decisions of young men. *Journal of Political Economy 105*(3), 473–522.

Koopmans, T. C. (1950). *Statistical Inference in Dynamic Economic Models.* Number 10 in Cowles Commission Monograph. New York: John Wiley & Sons.

Marschak, J. (1953). Economic measurements for policy and prediction. In W. Hood and T. Koopmans (Eds.), *Studies in Econometric Method*, pp. 1–26. New York: Wiley.

Matzkin, R. L. (1992, March). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica 60*(2), 239–270.

Matzkin, R. L. (1993, July). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*(1-2), 137–168.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, pp. 2523–58. New York: North-Holland.

Matzkin, R. L. (2007). Nonparametric identification. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics*, Volume 6B. Amsterdam: Elsevier.

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: Columbia University Press for National Bureau of Economic Research.

Mincer, J. (2006). Acceptance Speech Upon Receiving the First IZA Prize November 24th 2002 in New York. In S. Grossbard (Ed.), *Jacob Mincer: A Pioneer of Modern Labor Economics* (1st ed.), Volume XII, pp. 19–20. New York: Springer Science+Business Media, Inc.

Pencavel, J. (1986). Labor supply of men. In O. Ashenfelter and R. Layard (Eds.), *Handbook of Labor Economics*, Volume 1, pp. 3–102. Amsterdam: North-Holland.

Stock, J. H. and D. Staiger (1997). Instrumental variables regression with weak instruments. *Econometrica 65*(3), 557–586.

Urzua, S. (2008). Racial labor market gaps: The role of abilities and schooling choices. Forthcoming, *Journal of Human Resources.*

Willis, R. J. and S. Rosen (1979, October). Education and self-selection. *Journal of Political Economy 87*(5, Part 2), S7–S36.

Yitzhaki, S. (1989). On using linear regression in welfare economics. Working Paper 217, Department of Economics, Hebrew University.

Table 1. Potential Outcomes, Choice Model and Parameterizations

Outcomes	Choice Model
$Y_j = \alpha_j + U_j \quad \text{with} \quad j \in \mathcal{J} = \{1,2,3\}$	$D_j = \begin{cases} 1 & \text{if } R_j \geq R_k \ \forall \ j \in \mathcal{J} \\ 0 & \text{otherwise} \end{cases}$
$Y = \sum_{j \in \mathcal{J}} Y_j D_j$	$R_j = \gamma_j Z_j - V_j \quad \text{with} \quad j \in \mathcal{J}$

Parameterization

$$(U_1, U_2, U_3, V_1, V_2, V_3) \sim N\left(\mathbf{0}, \mathbf{\Sigma}_{UV}\right) \quad , \quad (Z_1, Z_2, Z_3) \sim N\left(\mu_Z, \mathbf{\Sigma}_Z\right)$$

$$\mathbf{\Sigma}_{UV} = \begin{bmatrix} 0.64 & 0.16 & 0.16 & 0.024 & -0.32 & 0.016 \\ 0.16 & 1 & 0.20 & 0.020 & -0.30 & 0.010 \\ 0.16 & 0.20 & 1 & 0.020 & -0.40 & 0.040 \\ 0.024 & 0.020 & 0.020 & 1 & 0.6 & 0100 \\ -0.32 & -0.30 & -0.40 & 0.6 & 1 & 0.2 \\ 0.016 & 0.01 & 0.040 & 0100 & 0.2 & 1 \end{bmatrix}, \ \mu_Z = (1.0, 0.5, 1.5) \text{ and } \mathbf{\Sigma}_Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix} = [\ 0.3 \quad 0.1 \quad 0.7\ ], \quad \begin{bmatrix} \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} = [\ 0.2 \quad 0.3 \quad 0.1\ ]$$

Table 2. Transition Matrix Obtained from the Change in the Instrument Z_1

The Instrument Increases by 0.75 Standard Deviation

		New Value of Instrument ($\widetilde{Z} = Z_1 + 0.75$)			
		$D_1 = 1$	$D_2 = 1$	$D_3 = 1$	*Total*
Original Value of Instrument (Z_1)	$D_1 = 1$	33.17%	0%	0%	*33.17%*
	$D_2 = 1$	3.20%	25.91%	0%	*29.11%*
	$D_3 = 1$	2.43%	0%	35.29%	*37.72%*
	Total	*38.80%*	*25.91%*	*35.29%*	*100%*

Table 3. Marginal Gains Identified from the Change in the Instrument Z_1

The Instrument Increases by 0.75 Standard Deviation

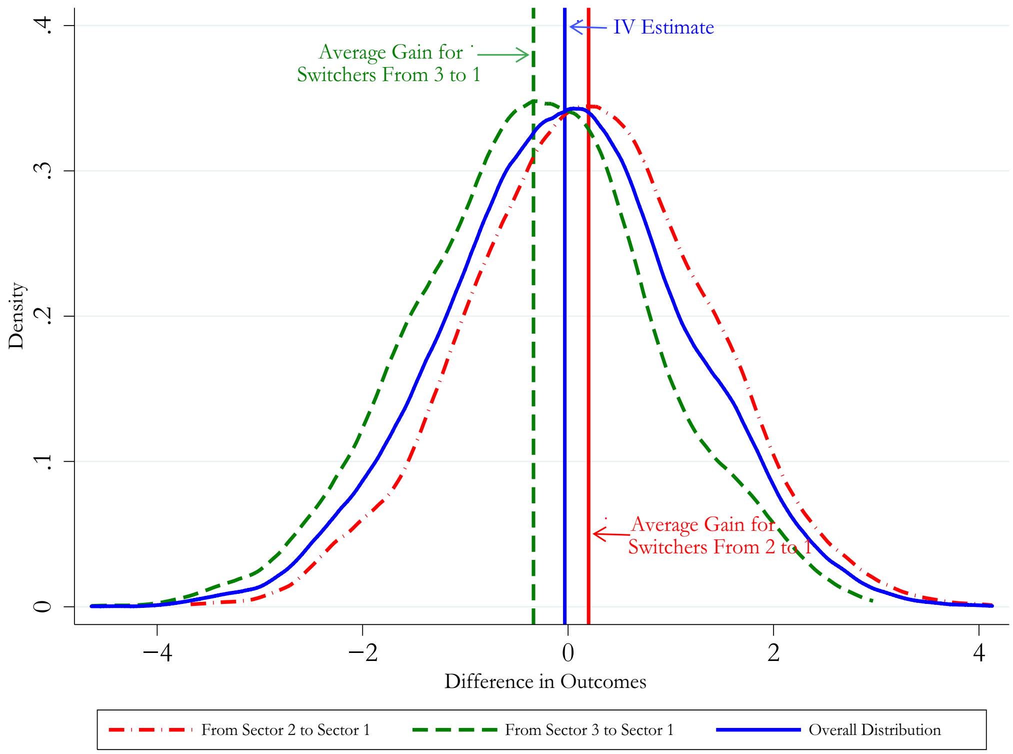
	Gains to Switchers	Fraction of Population Switching
From 2 to 1	0.199	3.20%
From 3 to 1	-0.336	2.43%
Overall (IV estimate)	-0.032	5.63%

IV Estimate:

$$E\left[Y|\widetilde{Z}_1\right] - E\left[Y|Z_1\right] = \frac{3.20}{3.20+2.43} \times 0.199 - \frac{2.43}{3.20+2.43} \times 0.336 = -0.032$$

Figure 1. Distribution of Gains in Outcomes Induced by the Change in the Instrument Z_1

The Instrument Increases by 0.75 Standard Deviation



Fraction of Gross Gainers by Source	
% Gross Gainers from 2 to 1	56.8%
% Gross Gainers from 3 to 1	39.3%
% Gross Gainers from all Sources	49.2%