Identification and Estimation of Local Average Treatment Effects

Author(s): Guido W. Imbens and Joshua D. Angrist

Source: *Econometrica*, Mar., 1994, Vol. 62, No. 2 (Mar., 1994), pp. 467-475

Published by: The Econometric Society

Stable URL: https://www.jstor.org/stable/2951620

Econometrica, Vol. 62, No. 2 (March, 1994), 467–475

IDENTIFICATION AND ESTIMATION OF LOCAL AVERAGE TREATMENT EFFECTS[1]

BY GUIDO W. IMBENS AND JOSHUA D. ANGRIST

1. INTRODUCTION

RANDOM ASSIGNMENT OF TREATMENT and concurrent data collection on treatment and control groups is the norm in medical evaluation research. In contrast, the use of random assignment to evaluate social programs remains controversial. Following criticism of parametric evaluation models (e.g., Lalonde (1986)), econometric research has been geared towards establishing conditions that guarantee nonparametric identification of treatment effects in observational studies, i.e. identification without relying on functional form restrictions or distributional assumptions. The focus has been on identification of average treatment effects in a population of interest, or on the average effect for the subpopulation that is treated. The conditions required to nonparametrically identify these parameters can be restrictive, however, and the derived identification results fragile. In particular, results in Chamberlain (1986), Manski (1990), Heckman (1990), and Angrist and Imbens (1991) require that there be some subpopulation for whom the probability of treatment is zero, at least in the limit.

The purpose of this paper is to show that even when there is no subpopulation available for whom the probability of treatment is zero, we can still identify an average treatment effect of interest under mild restrictions satisfied in a wide range of models and circumstances. We call this a *local average treatment effect* (LATE). Examples of problems where the local average treatment effect is identified include latent index models and evaluations based on natural experiments such as those studied by Angrist (1990) and Angrist and Krueger (1991). LATE is the average treatment effect for individuals whose treatment status is influenced by changing an exogenous regressor that satisfies an exclusion restriction.

2. IDENTIFICATION OF CAUSAL EFFECTS

The framework we use is essentially similar to that outlined by Rubin (1974, 1990), Heckman (1990), and described in our previous paper on identification of treatment effects (Angrist and Imbens (1991)). It defines causal effects in terms of *potential outcomes* or *counterfactuals* rather than in terms of the parameters of a regression model. Let $Y_i(0)$ be the response without the treatment or program for individual i. $Y_i(1)$ is the response with treatment. D_i is an indicator of treatment. We observe D_i and $Y_i = Y_i(D_i) = D_i \cdot Y_i(1) + (1 - D_i) \cdot Y_i(0)$ for a random sample of individuals. The individual treatment effect, or *causal effect*, is $Y_i(1) - Y_i(0)$ but since $Y_i(1)$ and $Y_i(0)$ are never observed for the same individual we are forced to rely on comparisons between different individuals and estimate average treatment effects.

The solution to the identification problem dominating the evaluation of medical treatments is random assignment to treatment and control groups. This guarantees that $E[Y_i(j)|D_i = 0] = E[Y_i(j)|D_i = 1]$ for $j = 0, 1$. In that case an unbiased estimator for the

[1] We are grateful to Gary Chamberlain, Larry Katz, Don Rubin, Geert Ridder, Jim Heckman, Charles Manski, seminar participants at Harvard/MIT, New York University, the Institute for Research on Poverty at the University of Wisconsin, and the University of Chicago, two anonymous referees and a co-editor for comments and suggestions, and to the NSF for financial support under Grants SES9122477 and SES9122627.

average treatment effect, $E[Y_i(1) - Y_i(0)]$, is available in the difference of the treatment/control averages, $\Sigma D_i Y_i / \Sigma D_i - \Sigma(1 - D_i)Y_i / \Sigma(1 - D_i)$.

In the evaluation of social programs, researchers have often relied on instrumental variables strategies to identify treatment effects. We define an instrumental variable, Z_i, to be a variable independent of the responses $Y_i(0)$ and $Y_i(1)$, and correlated with the participation indicator D_i. In order to formalize this, let $\mathcal{Z}$ be the support of Z_i. Define for each $z \in \mathcal{Z}$ a random variable $D_i(z)$.[2] $D_i(z)$ is equal to zero if an individual would not participate if he or she had the instrument Z_i equal to z, and it is equal to one if that individual would participate with $Z = z$. Clearly, we cannot observe the entire set of potential participation indicators $\{D_i(z) | z \in \mathcal{Z}\}$, but we can think about them in the same way we think about $Y_i(0)$ and $Y_i(1)$ even though they are not observed. We observe (Z_i, D_i, Y_i) for a random sample of individuals, where $D_i = D_i(Z_i)$, the participation indicator associated with Z_i, and $Y_i = Y_i(D_i)$, the response variable given the participation status D_i. The formal condition defining an instrument is the following.

CONDITION 1 (Existence of Instruments): *Let Z be a random variable such that (i) for all $w \in \mathcal{Z}$ the triple $(Y_i(0), Y_i(1), D_i(w))$ is jointly independent of Z_i, and, (ii) $P(w) = E[D_i | Z_i = w]$ is a nontrivial function of w.*

Part (ii) of Condition 1 is testable in a given application. Part (i) is similar to an exclusion restriction in a regression model. It is not testable and has to be considered on a case by case basis. Note that random assignment of Z_i does not guarantee that part (i) is satisfied because although random assignment implies that Z_i is independent of $D_i(w)$, it does *not* imply that Z_i is independent of $Y_i(0), Y_i(1)$. In a related paper (Angrist, Imbens, and Rubin (1993)), we discuss conditions similar to this in great detail, and investigate the implications of violations of these conditions.

In econometric program evaluation, linear latent index models are often employed (see, for example, Heckman and Robb (1985), and Heckman and Hotz (1989)). In these models, the participation decision is typically modeled by a latent index

$$D_i^* = \gamma_0 + Z_i \cdot \gamma_1 + \nu_i,$$

with the observed participation indicator, D_i, related to the unobserved latent index, D_i^*, by

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0, \\ 0 & \text{if } D_i^* \leq 0. \end{cases}$$

The response, Y_i, is related to the treatment via the equation

$$Y_i = \beta_0 + D_i \cdot \beta_1 + \varepsilon_i.$$

In this notation the counterfactuals are $Y_i(0) = \beta_0 + \varepsilon_i$, $Y_i(1) = \beta_0 + \beta_1 + \varepsilon_i$, and $D_i(z) = 1\{\gamma_0 + z \cdot \gamma_1 + \nu_i > 0\}$ for $z \in \mathcal{Z}$, where $1\{\cdot\}$ is the indicator function. In this regression framework Condition 1 is satisfied if Z_i is independent of ε_i and ν_i. The advantage of our setup is that it allows us to avoid the functional form and distributional assumptions inherent in these models.

Chamberlain (1986), Heckman (1990), and Angrist and Imbens (1991) have each noted that Condition 1 by itself is not enough to identify *any* average treatment effect.[3] To see

[2] The $D_i(z)$ notation was suggested to us by Gary Chamberlain.

[3] Manski (1990) shows conditions similar to Condition 1 are informative in the sense that they sharpen bounds on population averages of bounded functions of the treatment effect. However, we focus on (point) identification of average treatment effects without restrictions on the range of outcomes.

this, compare the expectation $E[Y_i|Z_i=z]$ for two points of support, z and w, with $P(z)>P(w)$:[4]

$$\begin{aligned}E[Y_i|Z_i=z]-E[Y_i|Z_i=w]&=E[D_i(z)\cdot Y_i(1)+(1-D_i(z))\cdot Y_i(0)|Z=z]\\&\quad-E[D_i(w)\cdot Y_i(1)+(1-D_i(w))\cdot Y_i(0)|Z=w].\end{aligned}$$

Using the independence in Condition 1, this is equal to

$$\begin{aligned}(1)\qquad &E[(D_i(z)-D_i(w))\cdot(Y_i(1)-Y_i(0))]\\&=\Pr[D_i(z)-D_i(w)=1]\cdot E[Y_i(1)-Y_i(0)|D_i(z)-D_i(w)=1]\\&\quad-\Pr[D_i(z)-D_i(w)=-1]\\&\quad\cdot E[Y_i(1)-Y_i(0)|D_i(z)-D_i(w)=-1].\end{aligned}$$

Equation (1) highlights an identification problem arising in the use of IV to estimate average treatment effects. The difference in equation (1) can be zero or even negative despite a strictly positive causal effect of D on Y for all individuals. For example, if

$$\begin{aligned}&\Pr[D_i(z)-D_i(w)=-1]\\&=\Pr[D_i(z)-D_i(w)=1]\cdot\frac{E[Y_i(1)-Y_i(0)|D_i(z)-D_i(w)=1]}{E[Y_i(1)-Y_i(0)|D_i(z)-D_i(w)=-1]},\end{aligned}$$

the difference is zero. Intuitively the problem here is that the treatment effect for those who shift from nonparticipation to participation when Z is switched from z to w can be cancelled out by the treatment effect of those who shift from participation to nonparticipation.

One commonly invoked condition that prevents this is the assumption of a constant treatment effect, $\alpha=Y_i(1)-Y_i(0)$, for all individuals in the population. Then $E[Y_i|Z_i=z]-E[Y_i|Z_i=w]$ is equal to $\alpha\cdot[P(z)-P(w)]$, and α is clearly identified. A second approach is to assume the existence of a value of the instrument, w, such that the probability of participation conditional on that value, $P(w)$, is equal to zero. Then $\Pr[D_i(z)-D_i(w)=-1]=0$, and the difference $E[Y_i|Z_i=z]-E[Y_i|Z_i=w]$ is equal to $P(z)\cdot E[Y_i(1)-Y_i(0)|D_i(z)=1]$. In this case the average treatment effect for the treated is identified. This type of condition is explored in Heckman (1990) and Angrist and Imbens (1991). Below we present a third assumption that solves the identification problem by preventing shifts in participation status in the opposite direction.

CONDITION 2 (Monotonicity): *For all* $z,w\in\mathcal{Z}$, *either* $D_i(z)\geqslant D_i(w)$ *for all* i, *or* $D_i(z)\leqslant D_i(w)$ *for all* i.

This condition ensures that the instrument affects the participation or selection decision in a monotone way. That is, if people are more likely, on average, to participate given $Z=w$ than given $Z=z$, then *anyone* who would participate given $Z=z$ must also participate given $Z=w$. Similar to Condition 1, this condition is fundamentally untestable, and its validity has to be argued in the context of a particular application (see Section 4). Note that in the linear latent index model discussed above, Condition 2 is automatically satisfied.

[4] We assume that these conditional expectations are finite.

Our main result is the following:

THEOREM 1: *If Conditions* 1 *and* 2 *hold, then we can identify the following average treatment effect*:

$$\alpha_{z,w} = E\big[Y_i(1) - Y_i(0) | D_i(z) \neq D_i(w)\big]$$

from the joint distribution of Y, D, *and* Z, *for all* z *and* w *in the support of* Z *such that* $E[Y_i|Z_i = z]$ *and* $E[Y_i|Z_i = w]$ *are finite, and* $P(z) \neq P(w)$.

PROOF: Let Condition 2 be satisfied with $D_i(z) \geqslant D_i(w)$. Then $\Pr[D_i(z) - D_i(w) = -1] = 0$ which implies that the second term in (1) is equal to zero, and

$$E\big[Y_i|Z_i = z\big] - E\big[Y_i|Z_i = w\big]$$
$$= \big(P(z) - P(w)\big) \cdot E\big[Y_i(1) - Y_i(0) | D_i(z) - D_i(w) = 1\big].$$

Dividing both sides by $P(z) - P(w)$ shows that the local average treatment effect can be expressed in terms of moments of the joint distribution of (Y, D, Z). *Q.E.D.*

The local average treatment effect is analogous to a regression coefficient estimated in linear models with individual effects using panel data. In models with fixed effects, the data are only informative about the impact of binary regressors on individuals for whom the value of the regressor changes over the period of observation. Under Theorem 1 the treatment effect identified is an average for those who can be induced to change participation status by a change in the instrument.

3. INSTRUMENTAL VARIABLES ESTIMATION

Theorem 1 implies that local average treatment effects can be estimated by comparing the average of outcome Y and treatment D at two different values of the instrument Z. This is exactly what the instrumental variables approach estimates in the case of a binary instrument. One way to exploit a multi-valued instrument is to estimate the ratio of the covariance of Y and some scalar function $g(Z)$, and the covariance of D and $g(Z)$. If Z is a scalar random variable, then the choice $g(z) = z$ leads to the standard IV estimator. If Z is a vector, $g(z)$ is often an estimate of $P(z)$. To guarantee that the IV estimand, $\mathrm{Cov}(Y, g(Z))/\mathrm{Cov}(D, g(Z))$, is a weighted average of local average treatment effects with nonnegative weights, we impose the following condition on the function $g(z)$:

CONDITION 3: *$g(z)$ is a function from the support of Z to $\Re$ such that*
(i) either for all $z, w \in \mathcal{Z}$, $\Pr(z) \leqslant P(w)$ *implies* $g(z) \leqslant g(w)$, *or, for all* $z, w \in \mathcal{Z}$, $P(z) \leqslant P(w)$ *implies* $g(z) \geqslant g(w)$;
(ii) $\mathrm{Cov}(D, g(Z)) \neq 0$.

There are three important cases where Condition 3(i) is immediately satisfied. The first is the basis of Theorem 1. If Z is binary, it is clear that $P(z)$ is either increasing or decreasing in $g(z)$. The second case is where $g(z) = E[D_i|Z_i = z] = P(z)$. For example, in the linear latent index model, $g(z) = E[D_i|Z_i = z] = \Pr(\gamma_0 + \gamma_1 z + \nu > 0)$. The third case is where Z is a scalar random variable, and both $g(z)$ and $P(z)$ are monotone in z.

The following theorem gives the relation between IV estimators and the local average treatment effects defined in the previous section. To avoid additional notation and smoothness assumptions, and because the examples in Section 4 all involve discrete instruments, it is formulated in terms of instruments with discrete support.

THEOREM 2: *Suppose that Conditions* 1, 2, *and* 3 *are satisfied. Let* Z *be a discrete random variable with support* $\{z_0, z_1, \ldots, z_K\}$, *ordered in such a way that if* $l < m$ *then* $P(z_l) \leq P(z_m)$. *Then, if* $\text{Cov}(D, g(Z)) \neq 0$, *the IV estimator for the effect of* D *on* Y *using* $g(Z)$ *as an instrument estimates*

$$\alpha_g^{IV} = \text{Cov}(Y, g(Z))/\text{Cov}(D, g(Z)) = \sum_{k=1}^{K} \lambda_k \cdot \alpha_{z_k, z_{k-1}},$$

with weights

$$\lambda_k = \frac{(P(z_k) - P(z_{k-1})) \cdot \sum_{l=k}^{K} \pi_l \cdot (g(z_l) - E[g(Z)])}{\sum_{m=1}^{K} (P(z_m) - P(z_{m-1})) \cdot \sum_{l=m}^{K} \pi_l \cdot (g(z_l) - E[g(Z)])}, \tag{2}$$

where $\pi_k = \Pr(Z = z_k)$ *and* $\alpha_{z_k, z_{k-1}}$ *is the local average treament effect* $E[Y_i(1) - Y_i(0) | D_i(z_k) = 1, D_i(z_{k-1}) = 0]$. *The weights* λ_k *are nonnegative and add up to one.*

PROOF: See Appendix.

The second part of this section analyzes the asymptotic distribution of the IV estimator. We consider here the case where $g(z)$ is a known function of z. In the Appendix we derive the asymptotic distribution for the case where $g(\cdot)$ depends on an unknown parameter which is estimated jointly with the average treatment effect α_g^{IV}. That case includes two-stage procedures where in the first stage the conditional expectation of D given Z is estimated.

THEOREM 3: *Let* $(Y_i, D_i, Z_i)_{i=1}^N$ *be* N *independent, identically distributed random variables, and* $g(\cdot)$ *a function from the support of* Z *to* $\Re$ *such that* $\text{Cov}(D, g(Z)) \neq 0$, *and let* $\hat{\alpha}_g^{IV}$ *be the instrumental variables estimator, given by*

$$\hat{\alpha}_g^{IV} = \widehat{\text{Cov}}(Y, g(Z))/\widehat{\text{Cov}}(D, g(Z)) = \frac{\sum_{i=1}^{N} g(Z_i) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^{N} g(Z_i) \cdot (D_i - \bar{D})},$$

where $\bar{Y} = \sum_{i=1}^{N} Y_i/N$ *and* $\bar{D} = \sum_{i=1}^{N} D_i/N$. *Assume that all variances and covariances are finite. As* N *goes to infinity,*

$$\sqrt{N}\left(\hat{\alpha}_g^{IV} - \alpha_g^{IV}\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{E\left[\varepsilon^2 \cdot \{g(Z) - E[g(Z)]\}^2\right]}{\text{Cov}^2(D, g(Z))}\right),$$

with $\varepsilon = Y - E[Y] - \alpha_g^{IV} \cdot (D - E[D])$.

PROOF: See Appendix.

In textbook discussions of instrumental variables estimation often the assumption $E[\varepsilon^2 | Z = z] = \sigma^2$ is made. In that case the variance in Theorem 3 simplifies to the standard IV variance, $\sigma^2 \cdot \text{Var}(g(Z))/\text{Cov}^2(D, g(Z))$.

4. EXAMPLES

In this section we give a number of examples and discuss the applicability of Conditions 1, 2, and 3. The examples exploit the manner in which a particular program or treatment is implemented to create instruments that are exogenous. Evaluations of this type are sometimes referred to as natural experiments, in contrast with the identification achieved in clinical trials where individuals are directly randomized into treatment and control groups.

EXAMPLE 1 (Draft Lottery): Angrist (1990) uses the Vietnam-era draft lottery to estimate the effect of veteran status on earnings. The instrument is the draft lottery number, randomly assigned to date of birth and used to determine priority for military conscription. The average probability of serving in the military falls with lottery number. Condition 1 requires that potential earnings with and without military service be independent of the lottery number. This is a standard IV assumption which would be violated if, for example, lottery numbers are related to earnings through some variable other than veteran status. Condition 2 requires that someone who would serve in the military with lottery number k would also serve in the military with lottery number l less than k, which seems plausible. In the Angrist draft lottery application, $g(z)$ is an estimate of $P(z)$, and Condition 3 is therefore satisfied. The average effect of veteran status identified under Theorem 1 is for men who would have served with a low lottery number, but not with a high lottery number.

EXAMPLE 2 (Administrative Screening):[5] Suppose applicants for a social program are screened by two officials. The two officials are likely to have different admission rates, even if the stated admission criteria are identical. Since the identity of the official is probably immaterial to the response, it seems plausible that Condition 1 is satisfied. The instrument is binary so Condition 3 is trivially satisfied. However, Condition 2 requires that if official A accepts applicants with probability $P(0)$, and official B accepts people with probability $P(1) > P(0)$, official B must accept *any* applicant who would have been accepted by official A. This is unlikely to hold if admission is based on a number of criteria. Therefore, in this example we *cannot* use Theorem 1 to identify a local average treatment effect nonparametrically despite the presence of an instrument satisfying Condition 1.

EXAMPLE 3 (Randomization of Intention to Treat):[6] Let the instrument be an indicator for assignment to treatment group in a randomized trial. The actual treatment indicator, D, may differ from the instrument Z because some individuals may not comply with their assignment. Condition 1 requires that the two counterfactual outcomes, say health status if treated and health status if untreated, are independent of the original assignment. Condition 2 requires that anyone who would take the treatment if assigned to the control group would also take the treatment if assigned to the treatment group. This seems plausible if noncompliance is the result of a decision by patients. The instrument is binary so Condition 3 is satisfied. The treatment effect identified here is the average treatment effect for those who always comply with their assignment.

Dept. of Economics, Harvard University, Cambridge, MA 02138, U.S.A., and NBER
and
Dept. of Economics, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 91905, Israel, and NBER

Manuscript received December, 1991; final revision received May, 1993.

APPENDIX

PROOF OF THEOREM 2: We start with three preliminary observations: First, without loss of generality, we assume that the first version of Condition 3 applies. Given that the points of support are ordered, this implies that if $l < m$, then $g(z_l) \leq g(z_m)$ and that $\text{Cov}(D, g(Z)) > 0$.

[5] This example was suggested to us by Geert Ridder.
[6] A similar example is discussed in Robins (1989).

Second, given the $K+1$ points in the support of Z, we can define $K\times(K+1)/2$ local average treatment effects α_{z_l,z_m}, one for each (unordered) pair of points of support (z_l, z_m). These $K\times(K+1)/2$ local average treatment effects are related in the following way (using the definition of α_{z_m,z_l}):

$$\alpha_{z_m,z_k}=\frac{P(z_l)-P(z_k)}{P(z_m)-P(z_k)}\cdot\alpha_{z_l,z_k}+\frac{P(z_m)-P(z_l)}{P(z_m)-P(z_k)}\cdot\alpha_{z_m,z_l},$$

for all $k\neq l$, $k\neq m$, and $l\neq m$.

Third, the conditional expectation of Y given $Z=z_k$ for $k\geqslant 1$ can be written as

$$\begin{aligned}E[Y|Z=z_k]&=E[Y|Z=z_0]+\alpha_{z_k,z_0}\cdot(P(z_k)-P(z_0))\\&=E[Y|Z=z_0]+\sum_{l=1}^{k}\alpha_{z_l,z_{l-1}}\cdot(P(z_l)-P(z_{l-1})).\end{aligned}$$

The IV procedure estimates

$$\alpha_g^{IV}=\frac{\mathrm{Cov}(Y,g(Z))}{\mathrm{Cov}(D,g(Z))}=\frac{E[Y\cdot(g(Z)-E[g(Z)])]}{E[D\cdot(g(Z)-E[g(Z)])]}. \tag{3}$$

First we analyze the numerator of this expression:

$$\begin{aligned}&E[Y\cdot(g(Z)-E[g(Z)])]\\&=\sum_{l=0}^{K}\pi_l\cdot E[Y|Z=z_l]\cdot(g(z_l)-E[g(Z)])\\&=\sum_{l=0}^{K}\pi_l\cdot E[Y|Z=z_0]\cdot(g(z_l)-E[g(Z)])\\&\quad+\sum_{l=1}^{K}\pi_l\sum_{k=1}^{l}\alpha_{z_k,z_{k-1}}\cdot(P(z_k)-P(z_{k-1}))\cdot(g(z_l)-E[g(Z)])\\&=\sum_{k=1}^{K}\alpha_{z_k,z_{k-1}}\cdot(P(z_k)-P(z_{k-1}))\sum_{l=k}^{K}\pi_l\cdot(g(z_l)-E[g(Z)]),\end{aligned}$$

where the factor multiplying $\alpha_{z_k,z_{k-1}}$ is equal to the numerator of λ_k in (2). A similar calculation shows that the denominator of α_g^{IV} in (3) is equal to the denominator of λ_k in (2).

The weights λ_k clearly add up to one. They are nonnegative because $P(z_k)\geqslant P(z_{k-1})$ and $g(z_k)\geqslant g(z_{k-1})$ which follows from the ordering of the points of support, and this in turn implies that $\sum_{l=k}^{K}\pi_l\cdot(g(z_l)-E[g(Z)])\geqslant 0$ for all k. *Q.E.D.*

PROOF OF THEOREM 3: We give the proof for the case where $g(\cdot)$ is estimated, which will as a special case give the variance for the known $g(\cdot)$ case. Let $g(z,\theta)$ be a known function, with θ an unknown parameter to be estimated in the first stage. Also, let $\psi(Z,D,\theta)$ be the estimating equation for θ. That is, $\hat\theta$ is characterized by $\sum_{n=1}^{N}\psi(Z_n,D_n,\hat\theta)=0$. We assume that there is a unique solution to $E\psi(Z,D,\theta)=0$ and that $E[\partial\psi(Z,D,\theta)/\partial\theta)]$ has a full rank. $\hat\alpha_g^{IV}$ can be written as the second element of the solution to

$$h\left(\gamma,\alpha_g^{IV}\right)=\frac{1}{N}\sum_{n=1}^{N}\begin{pmatrix}\left(Y_n-\gamma-\alpha_g^{IV}\cdot D\right)\\ g\left(Z_n,\hat\theta\right)\cdot\left(Y_n-\gamma-\alpha_g^{IV}\cdot D\right)\end{pmatrix}=0,$$

where γ is a nuisance parameter, equal to $E[Y]-\alpha_g^{IV}\cdot E[D]$. Using standard asymptotic theory, the asymptotic variance of $\sqrt{N}((\hat{\theta}-\theta)',\hat{\gamma}-\gamma,\hat{\alpha}_g^{IV}-\alpha_g^{IV})'$ is equal to $V=\Gamma^{-1}\Delta(\Gamma')^{-1}$, with

$$\Delta=\begin{pmatrix} E[\psi(Z,D,\theta)\cdot\psi(Z,D,\theta)'] & E[\varepsilon\cdot\psi(Z,D,\theta)] & E[g(Z)\cdot\varepsilon\cdot\psi(Z,D,\theta)] \\ E[\varepsilon\cdot\psi(Z,D,\theta)]' & E[\varepsilon^2] & E[g(Z)\cdot\varepsilon] \\ E[g(Z)\cdot\psi(Z,D,\theta)]' & E[g(Z)\cdot\varepsilon] & E[g^2(Z)] \end{pmatrix},$$

$$\Gamma=\begin{pmatrix} E\left[\dfrac{\partial\psi(Z,D,\theta)}{\partial\theta'}\right] & 0 & 0 \\ 0 & -1 & -E[D] \\ E\left[\varepsilon\cdot\dfrac{\partial g}{\partial\theta}\right] & -E[g(Z)] & -E[D\cdot g(Z)] \end{pmatrix},$$

where ε is again equal to $Y-E[Y]-\alpha_g^{IV}\cdot(D-E[D])=Y-\gamma-\alpha_g^{IV}\cdot D$.

Substitution of these matrices in the variance formula $V=\Gamma^{-1}\Delta(\Gamma')^{-1}$ gives the desired variance of $\sqrt{N}(\hat{\alpha}_g^{IV}-\alpha_g^{IV})$ as the bottom right element of V.

If $g(z)$ is a known function, the first column and row of Γ and Δ can be removed, and the asymptotic variance of $\sqrt{N}(\hat{\alpha}_g^{IV}-\alpha_g^{IV})$ is equal to the (2,2) element of

$$\begin{pmatrix} -1 & -E[D] \\ -E[g(Z)] & -E[D\cdot g(Z)] \end{pmatrix}^{-1}\cdot\begin{pmatrix} E[\varepsilon^2] & E[\varepsilon^2\cdot g(z)] \\ E[\varepsilon^2\cdot g(Z)] & E[\varepsilon^2\cdot g^2(Z)] \end{pmatrix}$$

$$\begin{pmatrix} -1 & -E[g(Z)] \\ -E[D] & -E[D\cdot g(Z)] \end{pmatrix}^{-1}.$$

The relevant matrices are invertible because $\mathrm{Cov}(D,g(Z))\neq 0$ and $E[\partial\psi(Z,D,\theta)/\partial\theta]\neq 0$.

The reason that the variance of the IV estimator is affected by the first stage estimation of $g(Z,\theta)$ is that Condition 1 implies only that $E[g(Z,\theta)\cdot\varepsilon]=0$, not necessarily that $E[\varepsilon|Z=z]=0$. The latter is often assumed in textbook discussions of instrumental variables, and making this assumption implies that the variance of the IV estimator is not affected by the first stage estimation because $E[\varepsilon\cdot\partial g/\partial\theta(Z,\theta)]$ is equal to zero. *Q.E.D.*

REFERENCES

ANGRIST, J. D. (1990): "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records," *American Economic Review*, 80, 313–336.

ANGRIST, J. D., AND A. KRUEGER (1991): "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106, 979–1014.

ANGRIST, J. D., AND G. W. IMBENS (1991): "Sources of Identifying Information in Evaluation Models," NBER Technical Working Paper 117.

ANGRIST, J. D., G. W. IMBENS, AND D. RUBIN (1993): "Identification of Causal Effects Using Instrumental Variables," NBER Working Paper.

CHAMBERLAIN, G. (1986): "Asymptotic Efficiency in Semi-parametric Models with Censoring," *Journal of Econometrics*, 32, 189–218.

HECKMAN, J. J. (1990): "Varieties of Selection Bias," *American Economic Review, Papers and Proceedings*, 80, 313–338.

HECKMAN, J. J., AND J. HOTZ (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84, 862–874.

HECKMAN, J. J., AND R. ROBB (1985): "Alternative Methods for Evaluating the Impact of Interventions," in *Longitudinal Analysis of Labor Market Data*, ed. by J. Heckman and B. Singer. New York: Cambridge University Press.

LALONDE, R. J. (1986): "Evaluating the Econometric Evaluations of Training Programs Using Experimental Data," *American Economic Review*, 76, 602–620.

MANSKI, C. F. (1990): "The Selection Problem," University of Wisconsin, Social Systems Research Institute Working Paper 9012.

ROBINS, J. M. (1989): "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in *Health Service Research Methodology: A Focus on AIDS*, ed. by L. Sechrest, H. Freeman, and A. Bailey. Washington, D.C.: NCHSR, U.S. Public Health Service.

RUBIN, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66, 688–701.

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