Submitted to Econometrica

On the Use of Instrumental Variables to Identify the Effect of a Mis-measured, Binary Regressor

Francis J. DiTraglia and Camilo Garcia-Jimeno

July 11, 2015

Submitted to Econometrica

ON THE USE OF INSTRUMENTAL VARIABLES TO IDENTIFY THE EFFECT OF A MIS-MEASURED, BINARY REGRESSOR[1]

Francis J. DiTraglia[a] and Camilo Garcia-Jimeno[a]

Abstract goes here.

Keywords: Instrumental Variables, Measurement Error, Binary Regressor, Endogeneity.

1. INTRODUCTION

Introduction goes here.

2. NOTES ON MAHAJAN (2006)

Mahajan (2006) considers regression models of the form

$$E\left[y - g(x^*, z)\right] = 0 \tag{1}$$

where x^* is an unobserved binary regressor and z is a $d_z \times 1$ vector of control regressors. Rather than x^* we observe a noisy measure x called the "surrogate" and an additional variable v that acts, in essence, as an instrumental variable. Since v does not, strictly speaking, meet the traditional requirements for an instrument, Mahajan refers to it as an "instrument-like variable" or ILV for short. Throughout the paper, Mahajan assumes that v is binary although he claims that the same idea applies to arbitrary discrete variables. The paper considers two main cases: one in which x^* is assumed to be exogenous, and another in which it is not.

[1]We thank…

[a]Dept. of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104, U.S.A.; fditra@econ.upenn.edu; gcamilo@econ.upenn.edu

2.1. *The Case of Exogenous x^**

The first is based on the restriction

$$E\left[y - g(x^*, z) \mid x^*, x, z, v\right] = 0 \tag{2}$$

2.2. *The Case of Endogenous x^**

While the preceding case required x^* to be exogenous, Mahajan claims (page 640) that his identification results can be extended to account for endogeneity provided that one is willing to restrict attending to additively separable models of the form

$$y = g^*(x^*, z) + \varepsilon \tag{3}$$

In this case, the ILV is assumed to satisfy the usual instrumental variables mean independence assumption

$$E\left[\varepsilon | z, v\right] = 0 \tag{4}$$

and Equation 2 is replaced by

$$E\left[y | x^*, x, z, v\right] = E\left[y | x^*, z\right] \tag{5}$$

Unfortunately, Mahajan's proof is incorrect and the model in Equation 3 is unidentified. The mistake stems from a false analogy with the identification proof in the case of exogenous x^*. In A.2 Mahajan argues, correctly, that under 3–5 knowledge of the mis-classification rates is sufficient to identify the model even when x^* is endogenous. He then appeals to Theorem 1 to argue that the mis-classification rates are indeed identified. The proof of Theorem 1, however, depends crucially on the assumption that x^* is exogenous. Without this assumption, the mis-classification rates are unidentified, as we now show For ease of exposition we consider the case without covariates. Equivalently, one can interpret all of the expressions that follow as implicitly conditioned on $z = z_a$ where z_a is a value in the support of z.[1]

Without covariates we can write

$$y = \alpha + \beta x^* + \varepsilon \tag{6}$$

where $\alpha = g^*(0)$ and $\beta = g^*(1) - g^*(0)$ and the mis-classification rates become $\eta_0 = P(x = 1|x^* = 0)$ and $\eta_1 = P(x = 0|x^* = 1)$. Now define

$$m_{jk} = E\left[\varepsilon|x^* = j, v = k\right] \tag{7}$$

3. LEWBEL (2007)

Lewbel shows that under an exogenous (but missclasified) treatment, and an instrument that takes on (at least) three values, the treatment effect is identified. Lewbel's three-valued instrument equivalent to having two binary instruments. Here we show how the logic in Lewbel's argument maps into the two binary instruments case within our framework.

The model is

$$\mathbb{E}[Y|T^*, T] = \alpha + \beta T^*$$

Using iterated expectations over the distribution of T^* given T,

$$\mathbb{E}[Y|T] = \mathbb{E}_{T^*|T}\left[\mathbb{E}[Y|T^*, T]\right] = \mathbb{P}(T^* = 1|T)\mathbb{E}[Y|T^* = 1] + \mathbb{P}(T^* = 0|T)\mathbb{E}[Y|T^* = 0]$$

$$\mathbb{E}[Y|T] = \mathbb{P}(T^* = 1|T)(\alpha + \beta) + \mathbb{P}(T^* = 0|T)\alpha$$

[1]Because the covariates are held fixed throughout the proof of Mahajan's Theorem 1, there is no loss of generality.

$$\mathbb{E}[Y|T] = \alpha + \mathbb{P}(T^* = 1|T)\beta$$

which implies that

$$\mathbb{E}[Y|T=1] - \mathbb{E}[Y|T=0] \equiv \beta^{OLS} = \left[\mathbb{P}(T^*=1|T=1) - \mathbb{P}(T^*=1|T=0)\right]\beta$$

Lewbel defines $M(\alpha_0, \alpha_1, p) = \mathbb{P}(T^* = 1|T = 1) - \mathbb{P}(T^* = 1|T = 0)$, implying that

$$\beta = \frac{\beta^{OLS}}{M(\alpha_0, \alpha_1, p)} \tag{8}$$

Lewbel also points out that $M(\alpha_0, \alpha_1, p)$ can also be expressed as

$$M(\alpha_0, \alpha_1, p) = \frac{1}{1 - \alpha_0 - \alpha_1}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p} - \frac{(1-\alpha_0)\alpha_1}{1-p}\right]$$

which I show below:

Using Bayes rule,

$$M(\alpha_0, \alpha_1, p) = \frac{\mathbb{P}(T=1|T^*=1)p^*}{p} - \frac{\mathbb{P}(T=0|T^*=1)p^*}{1-p}$$

$$M(\alpha_0, \alpha_1, p) = \frac{(1-\alpha_1)p^*}{p} - \frac{\alpha_1 p^*}{1-p}$$

$$= p^* \frac{(1-\alpha_1)(1-p) - \alpha_1 p}{p(1-p)} \tag{9}$$

Recall that previously we have shown that

$$p^* = \frac{p - \alpha_0}{1 - \alpha_0 - \alpha_1}$$

Replacing for p^* in equation (2),

$$M(\alpha_0,\alpha_1,p) = \frac{p-\alpha_0}{1-\alpha_0-\alpha_1}\frac{(1-\alpha_1)(1-p)-\alpha_1 p}{p(1-p)}$$

$$= \frac{1}{1-\alpha_0-\alpha_1}\left[\frac{(p-\alpha_0)(1-\alpha_1)(1-p)-(p-\alpha_0)\alpha_1 p}{p(1-p)}\right]$$

$$= \frac{1}{1-\alpha_0-\alpha_1}\left[\frac{p(1-p)-p(1-p)\alpha_1-(1-p)(1-\alpha_1)\alpha_0-(p-\alpha_0)\alpha_1 p}{p(1-p)}\right]$$

$$= \frac{1}{1-\alpha_0-\alpha_1}\left[1-\frac{(1-p)(1-\alpha_1)\alpha_0+p\alpha_1-p^2\alpha_1+p^2\alpha_1-\alpha_0\alpha_1 p}{p(1-p)}\right]$$

$$= \frac{1}{1-\alpha_0-\alpha_1}\left[1-\frac{(1-p)(1-\alpha_1)\alpha_0}{p(1-p)}-\frac{p\alpha_1(1-\alpha_0)}{p(1-p)}\right]$$

$$= \frac{1}{1-\alpha_0-\alpha_1}\left[1-\frac{(1-\alpha_1)\alpha_0}{p}-\frac{(1-\alpha_0)\alpha_1}{1-p}\right] \tag{10}$$

Now we use iterated expectations over the distribution of T^* given T and Z_k, where Z_k are the instruments $(k = 1, 2)$:

$$\mathbb{E}[Y|T,Z_k] = \mathbb{E}_{T^*|T,Z_k}\left[\mathbb{E}[Y|T^*,T,Z_k]\right] = \mathbb{P}(T^*=1|T,Z_k)\mathbb{E}[Y|T^*=1]+\mathbb{P}(T^*=0|T,Z_k)\mathbb{E}[Y|T^*=0]$$

$$\mathbb{E}[Y|T,Z_k] = \mathbb{P}(T^*=1|T,Z_k)(\alpha+\beta)+\mathbb{P}(T^*=0|T,Z_k)\alpha$$

$$\mathbb{E}[Y|T,Z_k] = \alpha+\mathbb{P}(T^*=1|T,Z_k)\beta$$

which implies that

$$\mathbb{E}[Y|T=1,Z_k]-\mathbb{E}[Y|T=0,Z_k] \equiv \beta^{OLS}_{Z_k} = \left[\mathbb{P}(T^*=1|T=1,Z_k)-\mathbb{P}(T^*=1|T=0,Z_k)\right]\beta$$

Notice that analogous to equation (3),

$$\mathbb{P}(T^* = 1|T = 1, Z_k) - \mathbb{P}(T^* = 1|T = 0, Z_k) \equiv M(\alpha_0, \alpha_1, p_k) = \frac{1}{1-\alpha_0-\alpha_1}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p_k} - \frac{(1 -}{1 -}\right.$$

where $p_k = \mathbb{P}(T = 1|Z_k)$, under the assumption that the missclasiffication probabilities are independent of Z_k which Lewbel assumes.

This implies that if we run an OLS regression on the observations for which $Z_k = 1$ only, then we have that

$$\beta = \frac{\beta^{OLS}_{Z_k=1}}{M(\alpha_0, \alpha_1, p_k)} \tag{11}$$

Since we have two instruments, we have two of these equations. Equations (1) and (4) imply that

$$\frac{\beta^{OLS}}{M(\alpha_0, \alpha_1, p)} = \frac{\beta^{OLS}_{Z_k=1}}{M(\alpha_0, \alpha_1, p_k)}$$

$$\beta^{OLS} M(\alpha_0, \alpha_1, p_k) = \beta^{OLS}_{Z_k=1} M(\alpha_0, \alpha_1, p)$$

$$\beta^{OLS} M(\alpha_0, \alpha_1, p_k) - \beta^{OLS}_{Z_k=1} M(\alpha_0, \alpha_1, p) = 0, \qquad k = 1, 2 \tag{12}$$

(5) are two equations in two unknowns, α_0 and α_1:

$$\beta^{OLS}\frac{1}{1-\alpha_0-\alpha_1}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p_k} - \frac{(1-\alpha_0)\alpha_1}{1-p_k}\right] - \beta^{OLS}_{Z_k=1}\frac{1}{1-\alpha_0-\alpha_1}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p} - \frac{(1-\alpha_0)\alpha}{1-p}\right.$$

$$\beta^{OLS}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p_k} - \frac{(1-\alpha_0)\alpha_1}{1-p_k}\right] - \beta^{OLS}_{Z_k=1}\left[1 - \frac{(1-\alpha_1)\alpha_0}{p} - \frac{(1-\alpha_0)\alpha_1}{1-p}\right] = 0$$

$$(1-\alpha_1)\alpha_0\left[\frac{\beta^{OLS}_{Z_k=1}}{p} - \frac{\beta^{OLS}}{p_k}\right] + (1-\alpha_0)\alpha_1\left[\frac{\beta^{OLS}_{Z_k=1}}{1-p} - \frac{\beta^{OLS}}{1-p_k}\right] = \beta^{OLS}_{Z_k=1} - \beta^{OLS}$$

which we can rewrite as

$$B_0 w_0^k + B_1 w_1^k = w_2^k$$

This is a system of two linear equations in two unknowns, B_0 and B_1. In matrix form,

$$\begin{bmatrix} w_0^1 & w_1^1 \\ w_0^2 & w_1^2 \end{bmatrix} \begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \begin{bmatrix} w_2^1 \\ w_2^2 \end{bmatrix}$$

as long as $w_0^1 w_1^2 - w_0^2 w_1^1 \neq 0$ (which is Assumption A5 in Lewbel (2007)),

$$\begin{bmatrix} B_0 \\ B_1 \end{bmatrix} = \frac{1}{w_0^1 w_1^2 - w_0^2 w_1^1} \begin{bmatrix} w_1^2 w_2^1 - w_1^1 w_2^2 \\ w_0^1 w_2^2 - w_0^2 w_2^1 \end{bmatrix}$$

Finally, given that $B_0 = (1-\alpha_1)\alpha_0$ and $B_1 = (1-\alpha_0)\alpha_1$, we can solve for the missclassification rates:

$$B_1 = \alpha_1 - \alpha_1 \frac{B_0}{1-\alpha_1}$$

$$\alpha_1 = \frac{1}{2}\left[1 - B_0 + B_1 \pm \sqrt{(1-B_0+B_1)^2 - 4B_1}\right]$$

and

$$\alpha_0 = \frac{B_0}{1 - \frac{1}{2}\left[1 - B_0 + B_1 \pm \sqrt{(1-B_0+B_1)^2 - 4B_1}\right]}$$

Once we have (α_0, α_1) we can go back to equation (1) and recover β.

In page 544 Lewbel observes that if hs instrument were binary (if we only had one instrument in our case), identification could be achieved with one additional restriction on the missclasification rates. One such restriction is implied by homoskedasticity on the instrument, which he does not mention.

4. IDENTIFICATION BY HOMOSKEDASTICITY

This section uses our notation rather than Mahajan's. We'll have to decide what notation we want to use in the paper itself but for the moment I'm trying to avoid confusion by talking about Mahajan's proofs using his own notation while keeping our derivations in the same notation we used on the whiteboard. I think that by assuming the instrument takes on three values (as in Lewbell) and imposing our homoskedasticity assumption we'll get identification in the case where T^* is endogenous so I've written out this derivation for arbitrary discrete z.

Now suppose that one is prepared to assume that

$$E[u^2|z] = E[u^2]. \tag{13}$$

When combined with the usual IV assumption, $E[u|z] = 0$, this implies $Var(u|z) = Var(u)$. Whether this assumption is reasonable, naturally, depends on the application. When z is the offer of treatment in a randomized controlled trial, for example, Equation 13 holds automatically as a consequence of the randomization. Similarly, in studies based on a "natural" rather than controlled experiment one typically argues that the instrument is not merely uncorrelated with u but *independent* of it, so that Equation 13 follows.

To see why homoskedasticity with respect to the instrument provides additional identifying information, first express the conditional variance of y as follows

$$Var(y|z) = \beta^2 Var\,(T^*|z) + Var(u|z) + 2\beta Cov(T^*, u|z) \tag{14}$$

Under 13, $Var(u|z)$ does not depend on z. Hence the *difference* of conditional variances evaluated at two values z_a and z_b in the support of z is simply

$$(15)\qquad \Delta Var(y|z_a, z_b) = \beta^2 \Delta Var(T^*|z_a, z_b) + 2\beta \Delta Cov(T^*, u|z_a, z_b)]$$

Where $\Delta Var(y|z_a, z_b) = Var(y|z = z_a) - Var(y|z = z_b)$, and we define $\Delta Var(T^*|z_a, z_b)$ and $\Delta Cov(T^*, u|z_a, z_b)$ analogously.

First we simplify the $\Delta Var(T^*|z_a, z_b)$ term. Since T is conditionally independent of z given T^*,

$$\begin{aligned} P(T = 1|z) &= E_{T^*|z}\left[E\left(T|z, T^*\right)\right] = E_{T^*|z}\left[E\left(T|T^*\right)\right] \\ &= P\left(T^* = 1|z\right)(1-\alpha_1) + \left[1 - P\left(T^* = 1|z\right)\right]\alpha_0 \\ &= \alpha_0 + (1-\alpha_0-\alpha_1)\, P(T^* = 1|z) \end{aligned}$$

Rearranging,

$$(16)\qquad P(T^* = 1|z) = \frac{P(T = 1|z) - \alpha_0}{1 - \alpha_0 - \alpha_1}$$

and accordingly,

$$(17)\qquad Var(T^*|z) = \frac{\left[P\left(T = 1|z\right) - \alpha_0\right]\left[1 - P\left(T = 1|z\right) - \alpha_1\right]}{\left(1 - \alpha_0 - \alpha_1\right)^2}$$

Thus, evaluating Equation 17 at z_a and z_b and simplifying,

$$(18)\qquad \Delta Var(T^*|z_a, z_b) = \frac{\Delta Var(T|z_a, z_b) + (\alpha_0 - \alpha_1)\, \Delta E(T|z_a, z_b)}{\left(1 - \alpha_0 - \alpha_1\right)^2}$$

Turning our attention to $\Delta Cov(T^*, u|z_a, z_b)$ first note that

$$(19)\qquad Cov(T^*, u|z) = E_{T^*|z}\left[E\left(T^* u|z, T^*\right)\right] = P(T^* = 1|z) E\left(u|T^* = 1, z\right)$$

since $E[z|u] = 0$. Combining this with Equation 16 and evaluating at z_a

and z_b gives

$$\Delta Cov(T^*, u|z_a, z_b) = \frac{[E(T|z_a) - \alpha_0]\, m_{1a} - [E(T|z_b) - \alpha_0]\, m_{1b}}{1 - \alpha_0 - \alpha_1} \tag{20}$$

where $m_{1a} = E[u|T^* = 1, z_a]$ and $m_{1b} = E[u|T^* = 1, z_b]$.

Both Equations 18 and 20 involve only observable quantities and the mis-classification rates α_0 and α_1. Equation 15, however, also involves β. Fortunately we can eliminiate this quantity as follows. First, let $\mathcal{W}(z_a, z_b)$ denote the Wald Estimator of β given by

$$\mathcal{W}(z_a, z_b) = \frac{E(y|z_a) - E(y|z_b)}{E(T|z_a) - E(T|z_b)} \tag{21}$$

Since $E(u|z) = 0$,

$$E(y|z_a) - E(y|z_b) = \beta\,[E(T^*|z_a) - E(T^*|z_b)]$$

and by Equation 16,

$$E(T|z_a) - E(T|z_b) = (1 - \alpha_0 - \alpha_1)\,[E(T^*|z_a) - E(T^*|z_b)]$$

thus we find that

$$\beta = (1 - \alpha_0 - \alpha_1)\mathcal{W}(z_a, z_b). \tag{22}$$

Finally, combining Equations 15, 18, 20 and 22 we have

(23)
$$\begin{aligned}\Delta Var(y|z_a, z_b) = \mathcal{W}(z_a, z_b)^2 \{\Delta Var(T|z_a, z_b) + (\alpha_0 - \alpha_1)\Delta E(T|z_a, z_b)\}\\ + 2\mathcal{W}(z_a, z_b)\{[E(T|z_a) - \alpha_0]\, m_{1a} - [E(T|z_b) - \alpha_0]\, m_{1b}\}\end{aligned}$$

an equation relating $\alpha_0, \alpha_1, m_{1a}$ and m_{1b} to various observable quantities.

Equation 23 provides an additional identifying restriction for each unique *pair* of values (z_a, z_b) in the support of z. If z takes on two values it provides one restriction, whereas if z takes on three values if provides two restrictions, and so on. To take a particularly simple example, suppose that z is binary and Mahajan's (2006) assumption that $E[u|z, T^*] = 0$ holds. Then Equation 23 reduces to

$$\Delta Var(y|1,0) = \left[\frac{Cov(z,y)}{Cov(z,T)}\right]^2 \left\{\Delta Var(T|1,0) + (\alpha_0 - \alpha_1)\left[\frac{Cov(z,T)}{Var(z)}\right]\right\}$$

Rearranging, we see that

$$\alpha_0 - \alpha_1 = \Delta Var(y|1,0)\left[\frac{Cov(z,T)Var(z)}{Cov(z,y)^2}\right] - \Delta Var(T|1,0)\left[\frac{Var(z)}{Cov(z,T)}\right]$$

In other words, the homoskedasticity restriction identifies the *difference* between the mis-classification rates. This makes intuitive sense. Provided that the variance of u is unrelated to z the only way that the variance of y can differ across values of z is if some values of z provide *more* information about the distribution of T^* than others. This is only possible if the mis-classification rates differ.

Of course, one need not impose the restriction that $E[u|z, T^*] = 0$ to use the identifying information provided by Equation 23. Indeed, by exploiting homoskedasticity with respect to the instrument we can identify β using weaker conditions than Mahajan (2006) without requiring that z take on three or more values, as in Lewbel (2007). Moreover, when z does take on three or more values we can identify β even when T^* is endogenous.

I'm pretty sure this is true, but we do still need to prove it!

5. CONCLUSION

Conclusion goes here.