Instrumental Variables for Binary Treatments with Heterogeneous Treatment Effects: A Simple Exposition

Alan Manning

February 2004

Abstract
This note provides a simple exposition of what IV can and cannot estimate in a model with a binary treatment variable and heterogeneous treatment effects. It shows how linear IV is a misspecification of functional form and the reason why linear IV estimates for this model will always depend on the instrument used is because of this misspecification. It shows that if one can estimate the correct functional form (non-linear IV) then the treatment effects are independent of the instrument used. However, the data may not be rich enough in practice to be able to identify these treatment effects without strong distributional assumptions. In this case, one will have to settle for estimates of treatment effects that are instrument-dependent.

JEL Classification: C2
Keywords: Instrumental Variables, treatment effects, identification

This paper was produced as part of the Centre's Labour Markets Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
I would like to thank David Card, Maarten Goos, Guido Imbens and Marco Manacorda for their comments on this paper.

Alan Manning is a Programme head at the Centre for Economic Performance and a Professor of Economics, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

ISBN 0 7530 1721 0

Introduction

In the programme evaluation literature where 'treatment' is endogenous, applied economists make widespread use of the method of instrumental variables to 'deal' with the problem of endogeneity bias. Where the effect of treatment on the outcome variable is the same for everyone (the homogeneous treatment effect case) IV provides consistent estimates of this treatment effect and this is independent of the instrument used. But, economists have become more interested in the heterogeneous treatment effect case and the question then arises as to what exactly IV estimates. Although the literature on this subject is very large and very thorough, experience (including personal) tells me that the message is not getting through to many practising applied economists. The purpose of this note is to provide a very simple exposition of the main results in the literature that can be found in the work of Angrist, Heckman and Imbens amongst others[1].

It shows how linear IV is a misspecification of functional form when there is heterogeneity in treatment effects and the reason why linear IV estimates for this model will always depend on the instrument used is because of this misspecification. It shows that if one can estimate the correct functional form (using non-linear IV) then the treatment effects are independent of the instrument used. However, the data may not be rich enough in practice to be able to identify these treatment effects without strong distributional assumptions. In this case, one will have to settle for estimates of treatment effects that are instrument-dependent.

1. The Set-Up

To keep things as simple as possible, we assume that there is an outcome variable Y, a binary treatment variable D and an instrument Z. We omit other (exogenous) covariates in the interests of simplicity though it would be simple to include them. We also restrict discussion to the estimation of the treatment effects: we do not, for example discuss how one would compute standard errors for those estimates though standard techniques can be used for this.

The model for the outcome variable we will use is:

[1] A reader of this literature will be aware that a certain amount of attention is paid to the question of 'who said what when' with identical or similar results being presented in different papers as novel. I am even less of an historian of econometric thought than I am an econometrician and have no expertise in this question of attribution so the references in this paper to where results can be found should not be taken to imply this is the origin of the result referred to.

$$Y = b_0 + b_1 D + U + DU_1 \qquad (1)$$

where U and U_1 are unobserved errors with mean zero. Note that there is no functional form restriction like linearity in doing this because D can only take two values. The homogenous treatment effect case obviously corresponds to the case where $U_1 = 0$.

Once there is heterogeneity in treatment effects, there is a question about what treatment effects are we interested in. In this note we focus on just two:

- the average treatment effect (ATE) which here is just b_1

-the treatment effect on the treated (TT) which here is:

$$TT = b_1 + E\left(U_1 \middle| D = 1\right) \qquad (2)$$

We assume that one cannot simply obtain estimates of (1) by OLS as D is potentially correlated with U and U_1.

We need to specify the model determining D. A fairly general specification (see Vytlacil, 2002) is the following:

$$\begin{aligned} D^* &= E\left(D \middle| Z\right) - V \\ D &= I\left(D^* \geq 0\right) \end{aligned} \qquad (3)$$

where V is uniformly distributed on [0,1]. The familiar probit or logit models (amongst many others) can be written in the form of (3). V is potentially correlated with $\left(U, U_1\right)$. But Z is assumed to be independent of $\left(U, U_1, V\right)$ - this is stronger than the mean independence assumption used in traditional IV but most researchers will be happy to make this stronger assumption in practice.

2. Instrumental Variables

IV can be thought of as a two-step procedure (the most familiar incarnation of which is two-stage least squares). In the first stage one models the expectation of the endogenous variable conditional on the instruments i.e. models $E\left(D \middle| Z\right)$ which, given that D here is binary amounts to modelling $\Pr\left(D = 1 \middle| Z\right)$. There are many ways, parametric and non-parametric, in which one might do this: for our purposes here we will simply assume that we have a consistent estimate of this relationship.

In the second stage, IV estimates the relationship between the outcome variable, Y, and $E(D|Z)$ i.e. one examines the relationship between the expectation of the outcome variable conditional on the instruments, $E(Y|Z)$, and $E(D|Z)$. In traditional IV one models this as a linear relationship but, as we shall see, that is inevitably a misspecification of functional form for the model considered here.

Taking expectations of Y conditional on Z in (1) as the second stage of IV does leads to:

$$\begin{aligned} E(Y|Z) &= b_0 + b_1 E(D|Z) + E(U_1 D|Z) \\ &= b_0 + \left[b_1 + E(U_1|D=1,Z) \right] E(D|Z) \end{aligned} \tag{4}$$

Note that the coefficient on $E(D|Z)$ in the final line is the treatment effect on the treated, conditional on Z: for future use let us denote this by TT(Z) and define it as:

$$TT(Z) = b_1 + E(U_1|D=1,Z) \tag{5}$$

Using (3), the final line in (4) can be written as:

$$E(Y|Z) = b_0 + \left[b_1 + E(U_1|V \le E(D|Z)) \right] E(D|Z) \tag{6}$$

which can be written as:

$$E(Y|Z) = b_0 + \left[b_1 + k(E(D|Z)) \right] E(D|Z) \tag{7}$$

for some function $k(E(D|Z))$. Note that this model is intrinsically non-linear in $E(D|Z)$ if there is any heterogeneity in treatment effects and an endogeneity problem. In fact, the non-linearity of the relationship in (7) is, given the assumptions made, a necessary and sufficient condition for the existence of heterogeneity in treatment effects that are correlated with treatment status. So, a test of non-linearity is a test of this hypothesis (although such a test may have low power if there is little variation in $E(D|Z)$ in the data).

This intrinsic non-linearity helps us to understand the limitations of the familiar linear IV estimator for this model. As the 'true' relationship is non-linear, fitting a linear relationship is a misspecification of functional form for the relationship between $E(Y|Z)$ and $E(D|Z)$. As in any misspecification of this type the actual coefficient one will estimate depends on where the data actually is and this depends on the instrument. This is why linear IV estimates depend on the instrument used.

If one wants to estimate the true model (7) one obviously has to use a non-linear IV estimator. To understand what (7) might look like let us consider the familiar case where we have a probit the model for the determination of D and the errors are jointly normally distributed. Then we have:

$$E(Y|Z) = b_0 + b_1 E(D|Z) + rsf\left(\Phi^{-1}\left(E(D|Z)\right)\right) \qquad (8)$$

and the relationship looks something like the curve plotted in Figure 1. The actual curve one has is dependent on the instrument used. But, note that, independent of the instrument and the distribution of the errors $E(Y|Z) = b_0$ when $E(D|Z) = 0$ and $E(Y|Z) = b_0 + b_1$ when $E(D|Z) = 1$. The latter follows because:

$$k(1) = E(U_1|V \le 1) = E(U_1) = 0 \qquad (9)$$

given the assumption that V is uniformly distributed on the unit interval.

One can use Figure 1 to depict the different treatment effects. For example, the treatment effect on the treated conditional on $Z=z_0$, is given by the slope of the chord connecting the curve at $E(D|Z = z_0)$ to the point on the curve where $E(D|Z) = 0$: this is depicted on Figure 1[2]. And the ATE is given by the slope of the chord connecting the curve at $E(D|Z) = 1$ to the point on the curve where $E(D|Z) = 0$.

The ATE will clearly be independent of the instrument used but TT(Z) is obviously not. But, if one can successfully model the non-linear relationship (7) then the estimated treatment effects will be independent of the instrument used. To see this, let us leave to later the identification of (7) and assume that we do have consistent estimates available.

If one can estimate the model of (7) then one can estimate TT(Z) as given by (5). Obviously it will depend on Z, the instrument which may have no intrinsic interest. We are more likely to be interested in the treatment effect unconditional on Z. To obtain this we need to weight TT(z) by the density function of Z conditional on D=1, $f_{Z|D}(Z|D=1)$. Now we have that:

$$f_{Z|D}(z|D=1) = \frac{E(D|z)\ f_Z(z)}{E(D)} \qquad (10)$$

[2] If one is interested in it, the treatment effect on the untreated is given by the slope of the chord connecting a point on the curve to the other end of the curve.

where $f_Z(z)$ is the marginal density function of Z. The following result holds whether Z has a discrete or continuous distribution but, to keep the notation simple, we will assume it has a continuous distribution. In this case the estimated treatment effect on the treated for the population can be written as:

$$TT = \int TT(z) f_{Z|D}(z|D=1)dz = \frac{\int TT(z)E(D|z)\ f_Z(z)dz}{E(D)}$$
$$= \frac{\int \left[b_1 + E(U_1|D=1,z)\right]E(D|z)\ f_Z(z)dz}{E(D)} = b_1 + \frac{\int E(DU_1|z)\ f_Z(z)dz}{E(D)} \quad (11)$$
$$= b_1 + \frac{E(DU_1)}{E(D)} = b_1 + \frac{E(U_1|D=1)E(D)}{E(D)} = b_1 + E(U_1|D=1)$$

which is the true value and is independent of the instrument used. In practice, one way to compute this is to use the estimates of TT(Z) and then simply average this across the sample members for whom D=1.

But there is a simpler way to estimate TT in this case using a linear IV estimator. The trick is to define a binary instrument (let us denote it by $\tilde{Z}$) that takes the value one if $E(D|Z)>0$ and zero if $E(D|Z)=0$. We can then show that TT is given by:

$$TT = \frac{E(Y|\tilde{Z}=1) - E(Y|\tilde{Z}=0)}{E(D|\tilde{Z}=1) - E(D|\tilde{Z}=0)} \quad (12)$$

i.e. is the linear IV estimate of Y on D using $\tilde{Z}$ as an instrument. Because the proof of this detracts from the main story, the proof is in the Appendix. However, the intuition is very simple. For those with the value of Z such that $E(D|Z)=0$, the average value of Y will be b_0 as Z is a valid instrument. For all other observations the average value of Y will be b_0 plus the fraction receiving treatment times the average treatment effect for this group. The treatment effect on the treated can then be found by taking the difference in the value of Y for the two groups and dividing by the fraction treated in the second group.

But all the discussion in this section assumes that one can successfully identify the relevant treatment effects and, as the next section makes clear, this may not be so easy in practice.

3. Identification

The previous discussion presumed that one could identify all the parameters of (7). Let us now consider whether this is possible or not. Start by considering the case where one takes a non-parametric approach to estimation and does not impose any a priori functional form restrictions on $k\left(E\left(D|Z\right)\right)$. For every value of $E\left(D|Z\right)$ observed in the data one will be able to estimate $E\left(Y|Z\right)$ so will observe that point on the curve in Figure 1. But, as the prior discussion of Figure 1 has made clear to identify TT(Z) one needs to estimate the slope from the origin to $E\left(Y|Z\right)$: for that one needs to observe some value of Z for which $E\left(D|Z\right)=0$ (i.e. to observe the value of Y for some group who will never take the treatment). This result is in Angrsit and Imbens (1991) or Heckman and Vytlacil (2000). If one observes data at the origin then one can identify TT(Z) for every value of Z observed in the population, and one can then use this to estimate TT as in (11) above. Note that one does not need to observe all possible values of $E\left(D|Z\right)$ to identify TT as, in (11), the only values of Z that get any weight are the ones observed in the population and these can be estimated.

The condition for the identification of the ATE can also be readily understood from the earlier discussion of Figure 1. The ATE is the slope of the line connecting the two endpoints in the curve in Figure 1: to estimate this we need some observations on a group for whom $E\left(D|Z\right)=0$ and some on a group for whom $E\left(D|Z\right)=1$.

The condition for identifying TT that we have some observations for whom $E\left(D|Z\right)=0$ is a strong requirement as the data may lie nowhere close to this point. Suppose we are not in this fortunate position. What can de done in this case?

One option is to use a specific assumption about the functional form of $k\left(E\left(D|Z\right)\right)$ and then use the estimates for extrapolation to estimate the value of $E\left(Y|Z\right)$ when $E\left(D|Z\right)=0$. The risk is, of course, that the functional form may be incorrect and the extrapolation lead to large errors, a risk that is particularly large when the data is not close to the point $E\left(D|Z\right)=0$. Any error in the estimation of b_0 will translate one-for-one into an error in the estimation of TT.

If one does not observe any data for which $E(D|Z)=0$ and one is not prepared to use extrapolation, what can be identified in this case? Let us consider the worst-case scenario (that figures prominently in the work of Angrist and Imbens – see Imbens and Angrist (1984), or Angrist (2003)) where we observe only two values of $E(D|Z)$ i.e. the instrument is binary. Suppose the two values are z_0 and z_1 as depicted in Figure 2. Then one only observes two points on the curve in Figure 1 and one can do no more than estimate the slope of the line connecting the two points. If one makes the monotonicity assumption of Angrist and Imbens that when $E(D|Z)$ rises all those who were initially taking the treatment continue to do so plus some extra (the compliers in their terminology) then the increase in $E(Y|Z)$ will be the result of the compliers switching from non-treatment to treatment. As a result the linear IV estimator is the average treatment effect for the compliers. This is the local average treatment effect (LATE) – a more formal proof can be found in Imbens and Angrist (1984). LATE will not be independent of the instrument used, although one can obtain results (shown in Angrist, 2003) that if the instrument is symmetric and the distribution of the errors is also symmetric, then LATE will equal the ATE. This can also be readily understood using Figure 1 as the slope of the true curve in the middle of the distribution is the same as ATE.

The case of a binary instrument is a bit pessimistic (especially as covariates can also often be used as a source of variation in $E(D|Z)$, although typically at the cost of some restriction on functional form for the way the covariates affect Y). So let us consider what can be identified when the instrument takes a number of finite values e.g. we add an extra value of Z, z_2, as illustrated in Figure 2. One can join up each of the points on the curve observed using a straight line and interpret the slopes as the LATE for each change in the instrument. Or, one could begin to fit the curvature of the true relationship. Suppose we take the latter approach and, in addition that Z is continuous over some range. Then one can estimate the true curve over some interval. In particular one can estimate the slope at each point. This slope is what Heckman and Vytlacil (2000) call the marginal treatment effect (MTE). It can be thought of as LATE for small changes in the value of the instrument. Note, that MTE, like LATE will be instrument-dependent.

This discussion can be summarized as follows. If the data contains observations on individuals for whom $E(D|Z)=0$ then one can identify TT and the resulting estimate will be independent of the instrument used. However, one cannot obtain this estimate by linear

IV as the functional form is inevitably non-linear. In addition, if one also has observations on individuals for whom $E(D|Z)=1$ then one can identify ATE: again, this is instrument-independent. In this case IV estimates have what Angrist calls external validity.

But, if the data is not that rich then one can hope to identify TT and ATE if one makes specific assumptions on functional form and extrapolates. But, if one is loathe to do this then one has to settle for estimates of LATE and MTE that are instrument-dependent. In this case, the IV estimates will have what Angrist calls 'internal validity'.

4. Conclusion

This note provides a simple exposition of what IV can and cannot estimate in a model with a binary treatment variable and heterogeneous treatment effects. It shows how linear IV is a misspecification of functional form and the reason why linear IV estimates for this model will always depend on the instrument used is because of this misspecification. It shows that if one can estimate the correct functional form (non-linear IV) then the treatment effects are independent of the instrument used. However, the data may not be rich enough in practice to be able to identify these treatment effects without strong distributional assumptions. In this case, one will have to settle for estimates of treatment effects that are instrument-dependent.

Appendix: Proof of (12)

By construction, for $\tilde{Z}=0$ we have that $E\left(D\left|\tilde{Z}=0\right.\right)=0$ so that:

$$E\left(Y\left|\tilde{Z}=0\right.\right)=b_0 \tag{13}$$

For $\tilde{Z}=1$ we have (from (4)) that:

$$E\left(Y\left|\tilde{Z}=1\right.\right)=b_0+\left[b_1+E\left(U_1\left|D=1,\tilde{Z}=1\right.\right)\right]E\left(D\left|\tilde{Z}=1\right.\right) \tag{14}$$

Now, using (13) and (14) we have that the IV estimator of (12) estimates:

$$\left[b_1+E\left(U_1\left|D=1,\tilde{Z}=1\right.\right)\right] \tag{15}$$

The reason that this is TT is that:

$$\begin{aligned}E\left(U_1\left|D=1\right.\right)&=E\left(U_1\left|D=1,\tilde{Z}=1\right.\right)\Pr\left(\tilde{Z}=1\left|D=1\right.\right)+E\left(U_1\left|D=1,\tilde{Z}=0\right.\right)\Pr\left(\tilde{Z}=0\left|D=1\right.\right)\\&=E\left(U_1\left|D=1,\tilde{Z}=1\right.\right)\end{aligned} \tag{16}$$

as $\Pr\left(\tilde{Z}=1\left|D=1\right.\right)=1$ and $\Pr\left(\tilde{Z}=0\left|D=1\right.\right)=1$ by the construction of $\tilde{Z}$ which takes the value zero if there is zero probability of treatment.

Figure 1

The Relationship between E(Y|Z) and E(D|Z)

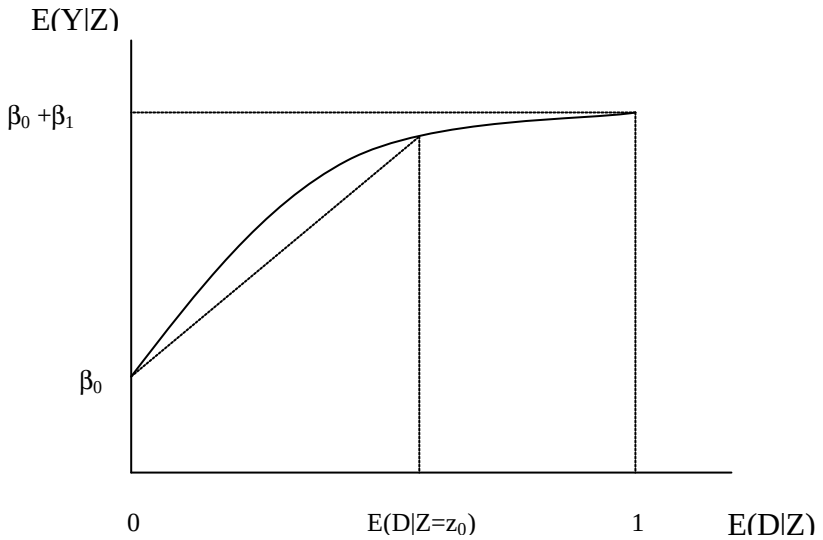
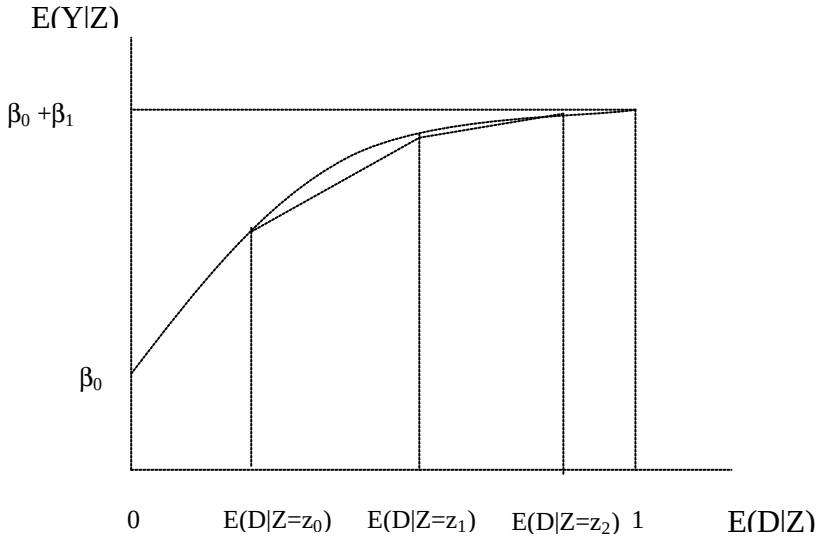


Figure 2
LATE and MTE



References

Angrist, Joshua D. (2003), 'Treatment Effect Heterogeneity in Theory and Practice', Economic Journal.

Angrist, Joshua D. and Imbens, Guido W. (1991), 'Sources of Identifying Information in Evaluation Models', NBER Technical Working Paper No. 117, December.

Imbens, Guido W. and Angrist, Joshua D. (1994), 'Identification and Estimation of Local Average Treatment Effects', Econometrica, 62: pp. 467-475.

Heckman, James J. and Vytlacil, Edward J. (2000), 'Local Instrumental Variables' in C. Hsiao, K. Morimune, and J. Powell, (eds.), *Nonlinear Statistical Inference: Essays in Honor of Takeshi Amemiya*, *Essays in Nonlinear Econometrics*, Cambridge University Press: Cambridge.

Vytlacil, Edward J. (2002), 'Independence, Monotonicity, and Latent Index Models: An Equivalence Result', Econometrica, 70(1): pp. 331-341.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

618	Mari Kangasniemi L. Alan Winters Simon Commander	Is the Medical Brain Drain Beneficial? Evidence from Overseas Doctors in the UK
617	Vicente CuÑat Maria Guadalupe	Executive Compensation and Product Market Competition
616	James Harrigan Anthony J. Venables	Timelines, Trade and Agglomeration
615	Howard Gospel Paul Willman	Comparatively Open: Statutory Information Disclosure for Consultation and Bargaining in Germany, France and the UK
614	Andrew B. Bernard Stephen Redding Peter K. Schott Helen Simpson	Relative Wage Variation and Industry Location
613	David Marsden	Unions and Procedural Justice: An Alternative to the Common Rule
612	David G. Blanchflower Alex Bryson	The Union Wage Premium in the US and the UK
611	Stephen Gibbons Stephen Machin	Valuing Rail Access Using Transport Innovation
610	Johannes Hörner L.Rachel Ngai Claudia Olivetti	Public Enterprises and Labor Market Performance
609	Nikolaus Wolf	Endowments, Market Potential, and Industrial Location: Evidence from Interwar Poland (1918-1939)
608	Ellen E. Meade David Stasavage	Publicity of Debate and the Incentive to Dissent: Evidence from the US Federal Reserve
607	Ghazala Azmat Maia Güell Alan Manning	Gender Gaps in Unemployment Rates in OECD Countries

606	Henry G. Overman L. Alan Winters	The Geography of UK International Trade
605	Stephen Machin Stephen Wood	Looking for HRM/Union Substitution: Evidence from British Workplaces
604	Maarten Goos Alan Manning	Lousy and Lovely Jobs: the Rising Polarization of Work in Britain
603	Nan-Kuang Chen Hsiao-Lei Chu	Collateral Value and Forbearance Lending
602	Ricardo Peccei Helen Bewley Howard Gospel Paul Willman	Is it Good To Talk? Information Disclosure and Organisational Performance in the UK Incorporating evidence submitted on the DTI discussion paper 'High Performance Workplaces – Informing and Consulting Employees'
601	Andy Charlwood	The Anatomy of Union Decline in Britain 1990-1998
600	Christopher A. Pissarides	Unemployment in Britain: A European Success Story
599	Stephen Bond Dietmar Harhoff John Van Reenen	Corporate R&D and Productivity in Germany and the United Kingdom
598	Michael Storper Anthony J. Venables	Buzz: Face-to-Face Contact and the Urban Economy
597	Stephen Gibbons Alan Manning	The Incidence of UK Housing Benefit: Evidence from the 1990s Reforms
596	Paul Gregg Maria Gutiérrez-Domènech Jane Waldfogel	The Employment of Married Mothers in Great Britain: 1974-2000

The Centre for Economic Performance Publications Unit
Tel 020 7955 7673 Fax 020 7955 7595 Email info@cep.lse.ac.uk
Web site http://cep.lse.ac.uk