the independence between (x_{i1}, x_{i2}) and c_i cannot be exploited within the fixed-effects framework. Unfortunately, the additional information that (x_{i1}, x_{i2}) and c_i are independent of each other does not reduce difficulty in estimating the index θ: It is not yet clear whether θ can be $\sqrt{n}$-consistently estimated even with the random-effects assumption when the random effects are nonparametrically specified. On the other hand, estimating β becomes easier as a result of the constancy of the generalized propensity score Pr $[x_{i1}, x_{i2}|c_i]$: Presence of unobserved c_i in the model was rendered irrelevant due to the constancy of the generalized propensity score. [See Imbens (1999) for a discussion on the generalized propensity score.]

It is interesting to note that, in the panel probit model (1), estimation of β is not necessarily simple unless the index structure is discarded altogether. It is useful to note that the new target parameter β could be estimated consistently using index structure *if* consistent estimators of θ and $\mathcal{L}$ are given. Here, $\mathcal{L}$ denotes the distribution of c_i. We may alternatively write (2) as

$$\int (\Phi(c+\theta) - \Phi(c))\, d\mathcal{L}(c), \tag{3}$$

which can in principle be estimated by using consistent estimators of θ and $\mathcal{L}$. Estimation of β using the alternative characterization (3) requires consistent estimation of an additional parameter $\mathcal{L}$, a parameter that was not given too much attention in the past. The problem is that not many consistent estimators of $\mathcal{L}$ are available. It is not yet clear whether the model satisfies the primitive conditions for consistency of the nonparametric maximum likelihood estimator (NPMLE) as discussed by Heckman and Singer (1984). The difficulty in estimating the target parameter using the expression (3), which is based on the index structure, is in sharp contrast to the ease of the estimation strategy using the expression (2), for which the index structure is irrelevant.

The preceding discussion suggests that the success of Angrist's perspective critically hinges on the structure of treatment assignment *and* careful reexpression of the new target parameter. If the joint distribution of c_i and (x_{i1}, x_{i2}) is completely unknown, it is clear that changing the target parameter does not ease the difficulty of estimation. Angrist's perspective therefore requires substantial effort in modeling such joint distribution. Whether such a modeling effort will be successful in dealing with nonlinear panel problems remains to be seen.

ADDITIONAL REFERENCES

Chamberlain, G. (1984), "Panel Data," in *Handbook of Econometrics*, eds. Z. Griliches and M. D. Intriligator, Amsterdam: North-Holland, pp. 1247–1318.

——— (1992), "Binary Response Models for Panel Data: Identification and Information," unpublished manuscript.

Heckman, J., and Singer, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52, 271–320.

Imbens, G. (1999), "The Role of Propensity Score in Estimating Dose-Response Functions," Technical Working Paper 237, National Bureau of Economic Research, Cambridge, MA.

Manski, C. (1987), "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

Comment

Guido W. Imbens

Department of Economics, University of California at Los Angeles, Los Angeles, CA 90095
(*imbens@econ.ucla.edu*)

It is a pleasure to comment on this article by Joshua Angrist, whose applications of instrumental-variables methods (Angrist 1989; Angrist and Krueger 1991) have been a source of inspiration for my own work in this area. As with Angrist's previous work on instrumental variables, the current article raises some controversial issues and makes a number of important points. Here I offer some comments on three of them. First, I shall discuss the issues raised in Section 1, "Causal Effects and Structural Parameters," concerning the goals of statistical inference. Angrist argues that many questions of interest are most easily formulated in terms of comparisons between realized and potential outcomes, the latter defined as outcomes that would have been observed under alternative states of nature. I shall explore some of the implications of this view for empirical practice and econometric theory. Second, I shall offer some remarks on the role of economic theory in specification and identification of econometric models, again reinforcing Angrist's point regarding the importance of formulating the key assumptions in terms of potential outcomes. Third, I shall discuss some of the issues related to the limited dependent nature of outcome variables for empirical practice, in particular in the presence of covariates. Partly motivated by the widespread perception of fundamental difficulties in applying instrumental-variables methods to data with limited dependent outcome variables, Angrist argues that standard linear model techniques are generally applicable. I agree with Angrist's position that most of these perceived problems are exaggerated but suggest that principled inference should nevertheless take account of the limited dependent nature of the outcome variables and use nonlinear models.

1. CAUSAL ESTIMANDS

In his textbook discussion of the difference between structural and reduced-form estimates, Goldberger (1997) wrote, following Marshak (1953), that the ultimate goal of

Journal of Business & Economic Statistics
January 2001, Vol. 19, No. 1

econometrics is to provide predictions. More specifically, in my view, the goal is to provide predictions of policy interventions. Using both economic theory and data, economists wish to inform policy discussions by providing predictions of states of the world under different policy choices. Based on comparisons of such predictions, policy makers can then choose among the different policies using some social welfare measure as objective function (e.g., Heckman and Smith 1997). Angrist argues that such questions are most easily formulated in terms of potential outcomes. Here I want to elaborate on that view.

Consider, as an example the problem faced by a policy maker contemplating a new tax in a market. To evaluate this policy, the policy maker wishes to take into account the effect of the tax on the quantity traded. Economic theory suggests that this effect depends on the slope of the supply and demand functions. The first step is therefore the estimation of these slopes, and in the remainder of this discussion I shall focus on this component of the policy-evaluation problem. In principle the policy maker may be interested in the entire distribution of the quantity traded under various taxes. Let us assume, however, that for purposes of evaluation of the policies it is sufficient to know the average effect of the policy on the quantity traded. If there are only two values for the policy—for example, no tax or a tax—the difference between these two averages is the key quantity of interest. Following Rubin (1974) I will refer to this as the *estimand.*

Note that the choice of estimand is distinct from the statistical question of the specification of the model. Often the statistical model is specified in such a way that a single parameter corresponds to the estimand. For example, in a structural interpretation of the linear regression model, the coefficients correspond to the effect of changing the covariates by a single unit. Such one-to-one correspondence, however, is the exception rather than the rule. Wooldridge (1992) made this point in the context of Box–Cox regression models. Such models are often used when a linear representation for $E[Y|X]$ is inappropriate. The Box–Cox regression model generalizes this linear form to $E[Y(\lambda)|X] = X'\beta$, where

$$Y(\lambda) = \begin{cases} (Y^{\lambda}-1)/\lambda & \lambda \neq 0, \\ \ln Y & \lambda = 0. \end{cases}$$

Although consistent estimators for β exist under these assumptions, Wooldridge stressed that because (a) the interpretation of β changes with the value of λ and (b) knowledge of β and λ is not sufficient for recovering $E[Y|X]$, there is no reason for economists to be interested in estimates of β under these assumptions. In other words, β cannot be the sole focus of the researcher because the question it answers changes with the value of nuisance parameters. Wooldridge then suggested an alternative specification that always allows the researcher to recover the conditional expectation $E[Y|X]$.

In empirical work this distinction between the estimand and the parameters of the statistical model is consistent with the now common practice of reporting estimates of average derivatives in binary response models rather than reporting estimates of the logit or probit coefficients. Unlike a linear regression model, there is no direct link from one of the coefficients in the logit or probit model to average causal effects, and thus there is no intrinsic interest in such coefficients.

This view is at odds, however, with a large part of the semiparametric literature. An exception is the work by Stoker (e.g., Stoker 1986), who focused on estimation of index coefficients in settings where these are proportional to average derivatives and thus directly linked to changes in predictions. Consider, for example, the work on semiparametric estimation of binary response models. In this literature, such models are estimated without making logistic or probit assumptions, instead only making conditional mean or median assumptions in a latent index interpretation (e.g., Manski 1985). This literature, however, has begged the question of why economists should be interested in the coefficient estimates in these models in the absence of a direct link between these coefficients and the choice probabilities or their derivatives. Similarly, some of the models with fixed effects in panel data with limited dependent variables have focused on estimation of parameters that in themselves do not allow for estimation of conditional expectations or their derivatives and thus do not allow for estimation of causal effects. See Arellano and Honoré (in press) for a survey of many of these methods.

2. IDENTIFICATION

After deciding on the estimand, the next step is to make substantive assumptions on the process that generated the data. This is where economic, as opposed to statistical, theory plays a key role. Theoretical considerations may suggest that certain variables have no direct causal effect on others because they do not enter into agents' utility function, nor do they affect the constraints these agents face. For example, in some markets it may be reasonable to postulate the existence of demand and supply function and assume that their intersection determines observed prices and quantities. In that case it may be argued that certain variables—for example, weather conditions in agricultural markets—affect supply at fixed prices but not demand because weather conditions do not affect utility of the buyers nor do they constrain their choices given prices. Similarly, theoretical considerations may suggest which variables, determine agents' fertility choices and which variables, are excluded from such choices, as in the structural models described in Section 1.2 of Angrist.

For the purpose of considering such exclusion restrictions, as well as other assumptions, it is important to formulate them in a way that economic theory can be brought to bear on them. This makes the formulation in terms of counterfactuals or potential outcomes that Angrist advocates particularly appropriate. The potential outcomes describe outcomes in different environments, and as such are the primitives of economic analyses, as well as choices under different sets of constraints, which are the result of agents solving constrained optimization problems. Since economic theory studies such optimization problems, it is therefore well equipped to assess assumptions formulated directly in terms of these potential outcomes. An example of the formulation of the critical assumptions in terms of such potential outcomes is Angrist, Imbens, and Rubin (1996, AIR from here on). In contrast, latent index models, although under some conditions mathematically equivalent to the potential outcome framework (e.g., Vytlacil 1999), formulate the critical assumptions in terms of associations between observed variables and unobserved residuals, which appears more difficult to contemplate [see Imbens (1997) for a discussion of the confusion such formulations have caused in the statistics literature].

It is rare that economic theory is specific enough to determine the exact value of the estimand. More typical is that the theory is consistent with a range of values for the estimand. Observations on agents' choices and outcomes may be helpful in narrowing down this range. The econometrician's task is to link the data to the estimand. Typically a number of additional assumptions are made at this stage. Almost always it is assumed that there is only limited dependence, or no dependence at all, between choices made by different agents, and identification focuses on the link between the joint distribution of the observables, estimable in large samples, and the estimand. Two possibilities arise at this stage. Sometimes the estimand can be expressed as a functional of the joint distribution of the observables, in which case the estimand is identified. A leading example is where the estimand is the average treatment effect and theory suggests that assignment to treatment is random, or at least random conditional on a set of observed covariates (unconfounded assignment, selection on observables). Alternatively, the assumptions suggested by economic theory do not allow for the direct link between the distribution of observables and the estimand. In that case the researcher faces some choices. One option, advocated in a series of papers by Manski (see, for a general discussion, Manski 1995), is to estimate the range of values of the estimand consistent with the data given the substantive assumptions. Another option, followed in the current article by Angrist, is the local average-treatment-effect approach developed by Imbens and Angrist (1994) to consider what aspects of the estimand are identified given data and assumptions. In instrumental-variables settings, the population average treatment effect is often not identified, but the average effect for a specific subpopulation may be. In that case one may choose to estimate the average treatment effect for this subpopulation and leave the extrapolation to the principal estimand to the researcher, possibly aided by theoretical considerations. As Heckman wrote, "It is a great virtue of the LATE parameter that it makes the investigator stick to the data at hand, and separate out the aspects of an estimation that require out of sample extrapolation or theorizing from aspects of an estimation that are based on observable data" (Heckman 1999, p. 832).

Let us consider the case studied by Angrist, with its focus on the effect of having more than two children on labor supply. Angrist argues that the second birth being a multiple birth (e.g., twins) is a valid instrument for this effect. In terms of the AIR formulation, this requires a multiple birth to be as good as randomly assigned, and the absence of a systematic direct effect on labor supply other than through its effect on the number of children. Such assumptions may be controversial. For example, fertility treatments may lead to a systematic association between multiple births and choices made by couples, violating the first assumption. Even if we accept these assumptions, however, they only imply that the average causal effect of more kids on labor supply is identified for women who had a third child solely because their second birth was a multiple birth (compliers in the AIR terminology). In my view it is unlikely that this is the population of primary interest. Nevertheless, it is the only subpopulation the data are informative about in the sense of point identification under the substantive assumptions, and it would appear to offer some guidance regarding the population average causal effect to policy makers similar to the way in the medical world results from clinical trials in homogenous subpopulations are regarded as useful because they are viewed as indicative of population average causal effects.

3. LIMITED DEPENDENT VARIABLES

Typically economic theory offers some guidance concerning the determinants of certain outcomes without specifying the exact form or strength of their relationship. In that case statistical modeling is required to complete the specification. Consider the example Angrist studies with binary outcome, binary endogenous regressor, a binary instrument, and covariates. Angrist suggests as one possible approach estimating the average treatment effect through a linear probability model with instrumenting for an endogenous regressor. The benefits of the linear probability approach stemming from the linearity and robustness against misspecification of the first stage appear to me largely illusory. At this point the statistical modeling is only intended to provide flexible approximations to the underlying conditional distributions. This is a fundamentally different role from that played by the substantive assumptions that are essential for identification. Appeals to consistency under specific parameterizations therefore appear irrelevant—in a larger sample one may well wish to use a more flexible specification because less smoothing is required. In addition to finding the alleged benefits of the linear probability model unpersuasive, I find its disadvantages troubling. Within small subpopulations characterized by extreme values of the covariates, the smoothing implicit in linear probability models is likely to lead to unattractive predictions compared to predictions based on nonlinear models that respect the limited-dependent-variable nature of the outcomes.

An alternative approach is followed in the study of the effect of flu shots on hospitalization rates using randomized incentives for vaccination by Hirano, Imbens, Rubin, and Zhou (2000, HIRZ from here on). Given their assumptions, extensions of those made by AIR to the case with exogenous covariates, there are three subpopulations—compliers (units who change treatment status in response to a change in the value of the instrument), always-takers (who always take the treatment, irrespective of the value of the instrument), and never-takers (who never take the treatment, irrespective of the value of the instrument). HIRZ modeled the conditional distribution of these three "types" conditional on covariates as a trinomial distribution:

$$\Pr(\text{Type}_i = c|X_i = x) = \frac{\exp(x'\psi_c)}{1+\exp(x'\psi_c)+\exp(x'\psi_a)},$$

$$\Pr(\text{Type}_i = a|X_i = x) = \frac{\exp(x'\psi_a)}{1+\exp(x'\psi_c)+\exp(x'\psi_a)},$$

and

$$\Pr(\text{Type}_i = n|X_i = x)$$
$$= 1 - \Pr(\text{Type}_i = c|X_i = x) - \Pr(\text{Type}_i = a|X_i = x).$$

Now compare this setup to the selection models Angrist describes in Section 3. In the selection models, the equation describing the endogenous regressor is $D_i = 1\{\gamma_0 + \gamma_1 Z_i + \gamma_2' X_i > \eta_i\}$. Suppose that the instrument is binary and that γ_1 is positive. Then the two models are very similar, with units with $\gamma_0 + \gamma_1 + \gamma_2' X_i > \eta_i$ in the selection model classified as always-takers in the potential outcome framework (because, irrespective of the value of the instrument, $D_i = 1$ for such units), units with $\gamma_0 + \gamma_2' X_i < \eta_i$ classified as never-takers (because, irrespective of the value of the instrument, $D_i = 0$ for such units), and the units with $\gamma_0 + \gamma_2' X_i < \eta_i < \gamma_0 + \gamma_1 + \gamma_2' X_i$ classified as compliers.

One advantage of the trinomial model is that it easily generalizes to provide an arbitrarily good fit to any conditional trinomial distribution by including higher-order terms and interactions in the covariates. If there are no substantive reasons to impose additional restrictions one should not impose them implicitly in the specification of the statistical model. In particular, in the selection model it is not sufficient to add higher-order terms to the covariate vector to provide an arbitrarily good fit to the trinomial distribution. Such an approximation would have to involve heteroscedasticity and other distributional extensions that are not straightforward to implement in the selection model.

Conditional on the individual's type, HIRZ specified the outcome distributions given covariates as logistic regression models. Again the aim is to provide a flexible approximation to the conditional distribution in a manner that does not impose any implicit restrictions. Given that for a binomial distribution the logistic regression model can be thought of as providing a linear approximation to the log odds ratio, this choice is again an appealing one. An alternative is the probit model, which also provides a good approximation. Less attractive here is the linear probability model since it requires inequality restrictions on the parameters if the implicit estimates of the probabilities are to be bounded between 0 and 1.

In cases with other limited dependent variables, alternative nonlinear models may be appropriate. For example, if the outcomes are durations, subject to censoring, models specified in terms of hazard functions (e.g., Lancaster 1979) may be convenient for dealing with such data.

ADDITIONAL REFERENCES

Angrist, J. (1989), "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence From Social Security Administrative Records," *American Economic Review*, 80, 313–335.

Angrist, J., and Krueger, A. (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Ecomonics*, 106, 979–1014.

Arellano, M., and Honoré, B. (in press), "Panel Data," in *Handbook of Econometrics*, eds. J. Heckman and E. Leamer, Amsterdam: Elsevier, North-Holland.

Goldberger, A. (1991), *A Course in Econometrics*, Cambridge, MA: Cambridge University Press.

Heckman, J. (1999), "Instrumental Variables: Response to Angrist and Imbens." *Journal of Human Resources*, 34, 828–837.

Heckman, J., and Smith, J. (1997), "Evaluating the Welfare State," in *Econometrics and Economics in the 20th Century: The Ragnar Frisch Centenary*, ed. S. Strom, New York, Cambridge University Press, pp. 214–318.

Hirano, K., Imbens, G., Rubin, D., and Zhou, A. (2000), "Estimating the Effect of Flu Shots in a Randomized Encouragement Design," *Biostatistics*, 1, 69–88.

Imbens, G. (1997), Book Review of *The Foundations of Econometric Analysis*, by D. Hendry and M. Morgan, *Journal of Applied Econometrtics*, 12, 91–94.

Lancaster, T. (1979), "Econometric Methods for the Analysis of Duration Data," *Econometrica*,

Manski, C. (1985), "Semiparametric Analysis of Discrete Response," *Journal of Econometrics*, 27, 313–333.

——— (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.

Marshak, J. (1953), "Economic Measurements for Policy and Prediction," in *Studies in Econometric Method*, eds. W. Hood and T. Koopmans, New York, Wiley, pp. 1–26.

Stoker, T. (1986), "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461–1481.

Vytlacil, E. (1999), "Independence, Monotonicity, and Latent Index Models: An Equivalency Result," unpublished manuscript, University of Chicago, Dept. of Economics.

Wooldridge, J. (1992), "Some Alternatives to the Box–Cox Regression Model," *International Economic Review*, 33, 935–955.

Comment

Robert A. Moffitt

Department of Economics, Johns Hopkins University, Baltimore, MD 21218, and National Bureau of Economic Research, Cambridge, MA (*moffitt@jhu.edu*)

The Problem. Although the article by Angrist ranges across a number of issues, much of the discussion, and the article title, suggests that the problem of concern is that instrumental variables (IV) cannot be used in one of three common models. Let the first model be $y = \alpha + \beta d + x\delta + \epsilon$, where y is an absolutely continuous variable but d is binary, and where x is independent of ϵ but d is not. Then β can be consistently estimated with IV (Heckman and Robb 1985). Let the second model be $y^* = \alpha + \beta d^* + x\delta + \epsilon$, where y^* and d^* are continuous and where $y = 1(y^* > 0)$ and $d = d^*$ are the observed variables. The parameters of this model can likewise be estimated by IV with some auxiliary assumptions (Newey 1986; see Blundell and Smith 1993 for a review of alternative methods). But let the third model be $y^* = \alpha + \beta d + x\delta + \epsilon$, where

Journal of Business & Economic Statistics
January 2001, Vol. 19, No. 1