

Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments

Joshua D. Angrist and Alan B. Krueger

The method of instrumental variables is a signature technique in the econometrics toolkit. The canonical example, and earliest applications, of instrumental variables involved attempts to estimate demand and supply curves.¹ Economists such as P.G. Wright, Henry Schultz, Elmer Working and Ragnar Frisch were interested in estimating the elasticities of demand and supply for products ranging from herring to butter, usually with time series data. If the demand and supply curves shift over time, the observed data on quantities and prices reflect a set of equilibrium points on both curves. Consequently, an ordinary least squares regression of quantities on prices fails to identify—that is, trace out—either the supply or demand relationship.

P.G. Wright (1928) confronted this issue in the seminal application of instrumental variables: estimating the elasticities of supply and demand for flaxseed, the source of linseed oil.² Wright noted the difficulty of obtaining estimates of the elasticities of supply and demand from the relationship between price and quantity

¹ See Goldberger (1972) and Morgan (1990) for a discussion of the origins of instrumental variables and related methods. Bowden and Turkington (1984) provide a more technical discussion of instrumental variables. The term “instrumental variables” originated with Olav Reiersøl (1945); Morgan cites an interview in which Reiersøl attributed the term to his teacher, Ragnar Frisch.

² In the early 1920s, Wright’s son, Sewall Wright, developed “causal path analysis,” a method-of-moments-type technique for estimating recursive structural models and simultaneous equations. P.G. Wright showed that path analysis and instrumental variables were equivalent in his simultaneous equations application. It is quite likely that Sewall Wright deserves much of the credit for his father’s use of instrumental variables.

■ *Joshua D. Angrist is Professor of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Alan B. Krueger is Bendheim Professor of Economics and Public Policy, Princeton University, Princeton, New Jersey.*

alone. He suggested (p. 312), however, that certain “curve shifters”—what we would now call instrumental variables—can be used to address the problem: “Such additional factors may be factors which (A) affect demand conditions without affecting cost conditions or which (B) affect cost conditions without affecting demand conditions.” A variable he used for the demand curve shifter was the price of substitute goods, such as cottonseed, while a variable he used for the supply curve shifter was yield per acre, which can be thought of as primarily determined by the weather.

Specifically, an instrumental variables estimate of the demand elasticity can be constructed by dividing the sample covariance between the log quantity of flaxseed and the yield per acre by the sample covariance between the log price of flaxseed and the yield per acre. This estimate is consistent as long as yield per acre is uncorrelated with the error in the demand equation and correlated with price. Replacing yield per acre with the price of substitutes in this calculation generates an instrumental variables estimate of the supply elasticity. Intuitively, weather-related shifts in yield are used to trace out the demand curve, while changes in the price of substitutes are used to shift the demand curve so as to trace out the supply curve.

Wright (1928, p. 314) observed: “Success with this method depends on success in discovering factors of the type A and B.” He used six different supply shifters to estimate the demand curve and then averaged the six instrumental variables estimates. The resulting average elasticity of demand for flaxseed was -0.80 . His average instrumental variables estimate of the elasticity of supply was 2.4. Wright’s econometric advance went unnoticed by the subsequent literature. Not until the 1940s were instrumental variables and related methods rediscovered and extended.

Wright’s (1928) method of averaging the different instrumental variables estimates does not necessarily produce the most efficient estimate; other estimators may combine the information in different instruments to produce an estimate with less sampling variability. The most efficient way to combine multiple instruments is usually two-stage least squares, originally developed by Theil (1953).³ In the first stage, the “endogenous” right-hand side variable (price, in this application) is regressed on all the instruments. In the second stage, the predicted values of price, based on the data for the instruments and the coefficients estimated from the first-stage regression, are then either plugged directly into the equation of interest in place of the endogenous regressor or, equivalently, used as an instrument for the endogenous regressor. In this way, two-stage least squares takes the information in a set of instruments and neatly boils it down to a single instrument.⁴

³ The relative efficiency of two-stage least squares turns on a number of auxiliary assumptions, such as homoskedastic errors. See Wooldridge’s paper in this issue for a discussion of alternative generalized method of moments estimators.

⁴ Typically, a number of “exogenous” conditioning variables also appear in both the supply and demand equations. These exogenous covariates do not play the role of instruments, but nevertheless should be included in both the first- and second-stage regressions. Two-stage least squares can also be used if there

Instrumental Variables and Measurement Error

Instrumental variables methods were also pioneered to overcome measurement error problems in explanatory variables.⁵ Measurement error can arise for many reasons, including the limited ability of statistical agencies to collect accurate information and the deviation between the variables specified in economic theory and those collected in practice. If an explanatory variable is measured with additive random errors, then the coefficient on that variable in a bivariate ordinary least squares regression will be biased toward zero in a large sample. The higher the proportion of variability that is due to errors, the greater the bias. Given an instrument that is uncorrelated with the measurement error and the equation error (that is, the equation error from the model with the correctly measured data), but correlated with the correctly measured variable, instrumental variables provide a consistent estimate even in the presence of measurement error.

Friedman's (1957) celebrated analysis of the consumption function can be interpreted as an application of instrumental variables in this context. Annual income is a noisy measure of permanent income, so a regression of consumption on annual income yields too small an estimate of the marginal propensity to consume from permanent income. To overcome this measurement problem, Friedman grouped his data by city, which is equivalent to using a two-stage least squares procedure. The first stage is implicitly a regression of annual income on a set of dummies indicating each of the cities. The fitted values from this regression would be average income by city, so that regressing micro consumption data on fitted income values, as is done in two-stage least squares, is the same as a weighted regression using city average data, where the weights are the number of observations per city.

While two-stage least squares and other instrumental variables estimators are consistent, they are not unbiased. Instrumental variables estimates are not unbiased because they involve a ratio of random quantities, for which expectations need not exist nor have a simple form. In contrast, expectations of ordinary least squares estimates typically exist and are easily calculated. Textbooks sometimes gloss over the distinction between unbiasedness and consistency, but the difference can matter in practice. Unbiasedness means the estimator has a sampling distribution centered on the parameter of interest in a sample of any size, while consistency only means that the estimator converges to the population parameter as the sample size grows. Since instrumental variables estimates are consistent, but not unbiased, researchers using instrumental variables should aspire to work with large samples.

is more than one endogenous regressor in an equation, provided there are at least as many instruments as endogenous regressors (Bowden and Turkington, 1984).

⁵ Wald's (1940) method of fitting straight lines was specifically developed to overcome errors-in-variables problems. Durbin (1954) showed that Wald's method is a special case of instrumental variables. See also Geary (1949). In this symposium, Hausman provides an overview of measurement error problems.

The precise form of the asymptotic distribution of an instrumental variables estimator (that is, the sampling distribution in very large samples) depends on a number of technical conditions. Most modern software packages include options for “robust standard errors” that are asymptotically valid under reasonably general assumptions. It is important to remember, however, that in practice these standard errors are only approximate.

Instrumental Variables and Omitted Variables

Although instrumental variables methods are still widely used to estimate systems of simultaneous equations and to counteract bias from measurement error, a flowering of recent work uses instrumental variables to overcome omitted variables problems in estimates of causal relationships. Studies of this type are primarily concerned with estimating a narrowly defined causal relationship, such as the effect of schooling, training or military service on earnings; the impact of smoking or medical treatments on health; the effect of social insurance programs on labor supply; or the effect of policing on crime. The observed association between the outcome and explanatory variable of interest in these and many other examples is likely to be misleading in the sense that it partly reflects omitted factors that are related to both variables. If these factors could be measured and held constant in a regression, the omitted variables bias would be eliminated. In practice, however, economic theory typically does not specify all of the variables that should be held constant while estimating a relationship, and it is difficult to measure all of the relevant variables accurately even if they are specified.

One solution to the omitted variables problem is to assign the variable of interest randomly. For example, social experiments are sometimes used to assign people to a job training program or to a control group. Random assignment assures that participation in the program (among those in the assignment pool) is not correlated with omitted personal or social factors. Randomized experiments are not always possible, however. It would not be feasible to force a randomly chosen group of people to quit smoking or to attend school for an extra year, or to assign the value of the minimum wage randomly across states. On the other hand, it may be possible to find, or even to create, a degree of exogenous variation in variables like schooling, smoking and minimum wages. Instrumental variables offer a potential solution in these situations.

To see how instrumental variables can solve the omitted variables problem, suppose that we would like to use the following cross-sectional regression equation to measure the “rate of return to schooling,” denoted ρ :

$$Y_i = \alpha + \rho S_i + \beta A_i + \varepsilon_i.$$

In this equation, Y_i is person i 's log wage, S_i is the person's highest grade of schooling completed, and A_i is a measure of ability or motivation. (For simplicity,

we take A_i to be a single variable, although it could be a vector of variables.) Although the problem of estimating this equation is straightforward in principle, data on A_i are typically unavailable, and researchers are unsure what the right controls for ability or motivation would be in any case.⁶ Without additional information, the parameter of interest, ρ , is not identified; that is, we cannot deduce it from the joint distribution of earnings and schooling alone.

Suppose, however, we have a third variable, the instrument, denoted Z_i , which is correlated with schooling, but otherwise unrelated to earnings. That is, Z_i is uncorrelated with the omitted variables and the regression error, ε_i . Then an instrumental variables estimate of the payoff to schooling is the sample analog of $\text{Cov}(Y_i, Z_i) / \text{Cov}(S_i, Z_i)$. The instrumental variables methods allow us to estimate the coefficient of interest consistently and free from asymptotic bias from omitted variables, without actually having data on the omitted variables or even knowing what they are. (If there is more than one valid instrument, the coefficient of interest can be estimated by two-stage least squares.) Intuitively, instrumental variables solve the omitted variables problem by using only part of the variability in schooling—specifically, a part that is uncorrelated with the omitted variables—to estimate the relationship between schooling and earnings.

Instruments that are used to overcome omitted variables bias are sometimes said to derive from “natural experiments.”⁷ Recent years have seen a resurgence in the use of instrumental variables in this way—that is, to exploit situations where the forces of nature or government policy have conspired to produce an environment somewhat akin to a randomized experiment. This type of application has generated some of the most provocative empirical findings in economics, along with some controversy over substance and methods.

A good instrument is correlated with the endogenous regressor for reasons the researcher can verify and explain, but uncorrelated with the outcome variable for reasons beyond its effect on the endogenous regressor. Maddala (1977, p. 154) rightfully asks, “Where do you get such a variable?” Like most econometrics texts, he does not provide an answer. In our view, good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest.

In the case of schooling, human capital theory suggests that people make schooling choices by comparing the costs and benefits of alternatives. Thus, one possible source of instruments would be differences in costs due, say, to loan policies or other subsidies that vary independently of ability or earnings potential. A second source of variation in educational attainment is institutional constraints.

⁶ The expected coefficient on schooling from a regression that omits the A_i variable is $\rho + \beta\gamma$, where γ is the regression coefficient from a hypothetical regression of A_i on schooling. It should be apparent that if the omitted variable is uncorrelated with schooling, or uncorrelated with earnings conditional on schooling, the coefficient on schooling is an unbiased estimate of ρ even if A_i is omitted from the equation.

⁷ A natural experiment can be studied without application of instrumental variables methods; in this case, reduced form estimates would be presented.

Angrist and Krueger (1991) exploit this kind of variation in a paper that typifies the use of “natural experiments” to try to eliminate omitted variables bias.

The rationale for the Angrist and Krueger (1991) approach is that because most states required students to enter school in the calendar year in which they turned six, school start age is a function of date of birth. Those born late in the year are young for their grade. In states with a December 31st birthday cutoff, children born in the fourth quarter enter school at age $5\frac{3}{4}$, while those born in the first quarter enter school at age $6\frac{3}{4}$. Furthermore, because compulsory schooling laws typically require students to remain in school until their 16th birthdays, these groups of students will be in different grades when they reach the legal dropout age. In essence, the combination of school start age policies and compulsory schooling laws creates a natural experiment in which children are compelled to attend school for different lengths of time depending on their birthdays.

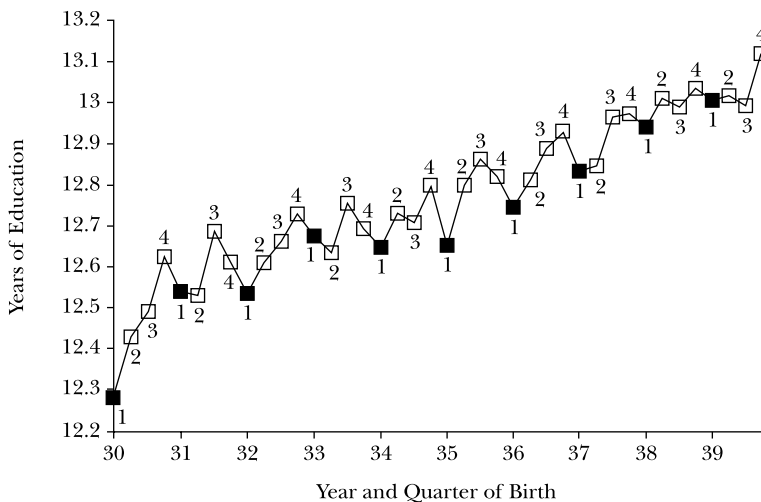
Using data from the 1980 census, we looked at the relationship between educational attainment and quarter of birth for men born from 1930 to 1959. The overall pattern is that younger birth cohorts finished more schooling. Figure 1 displays the relationship between education and quarter of birth for men born in the 1930s. The figure clearly shows that men born early in the calendar year tend to have lower average schooling levels. We selected this ten-year birth cohort because men this age tend to have a relatively flat age-earnings profile. But the pattern of less education for men born early in the year holds for men born in the 1940s and 1950s, as well. Because an individual’s date of birth is probably unrelated to the person’s innate ability, motivation or family connections (ruling out astrological effects), date of birth should provide a valid instrument for schooling.

Figure 2 displays average earnings by quarter of birth for the same sample. In essence, this figure shows the “reduced form” relationship between the instruments and the dependent variable. Older cohorts tend to have higher earnings, because earnings rise with work experience. But the figure also shows that, on average, men born in early quarters of the year almost always earn less than those born later in the year. Importantly, this reduced form relationship parallels the quarter-of-birth pattern in schooling. An examination of the reduced-form and first-stage estimates, either in graphical or tabular form, often provides insights concerning the causal story motivating a particular set of instrumental variable estimates. In this case, it is clear that the differences in education and earnings associated with quarter of birth are discrete blips, rather than smooth changes related to the gradual effects of aging.

The intuition behind instrumental variables in this case is that differences in earnings by quarter of birth are assumed to be accounted for solely by differences in schooling by quarter of birth, so that the estimated return to schooling is simply the appropriately rescaled difference in average earnings by quarter of birth. Only a small part of the variability in schooling—the part associated with quarter of birth—is used to identify the return to education. In our formal statistical estimates, we found that men born in the first quarter have about one-tenth of a year less schooling than men born in later quarters and that men born in the first quarter

Figure 1

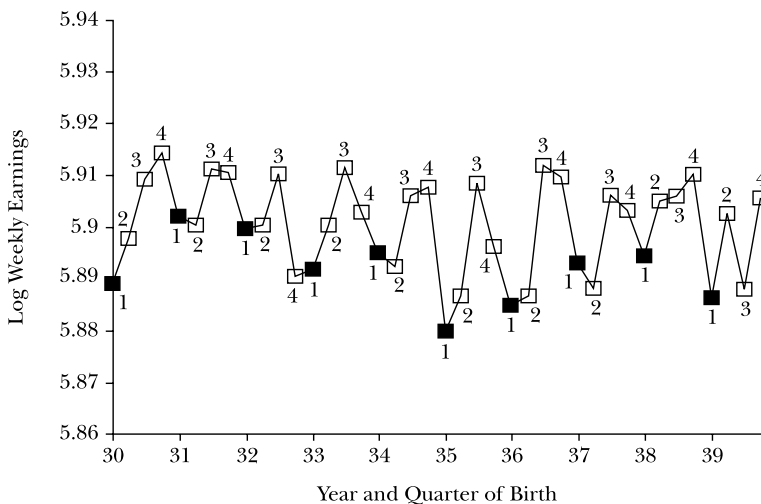
Mean Years of Completed Education, by Quarter of Birth



Source: Authors' calculations from the 1980 Census.

Figure 2

Mean Log Weekly Earnings, by Quarter of Birth



Source: Authors' calculations from the 1980 Census.

earn about 0.1 percent less than men born in later quarters. The ratio of the difference in earnings to the difference in schooling, about 0.10, is an instrumental variables estimate of the proportional earnings gain from an additional year of schooling.

As it turns out, this estimate of the change in earnings due to additional education differs little from a simple ordinary least squares regression of education on earnings in our data. This finding suggests that there is little bias from omitted ability variables in the ordinary least squares estimate of the effect of education on earnings, probably because omitted variables in the earnings equation are weakly correlated or uncorrelated with education. In other applications, the instrumental variables estimates and ordinary least squares estimates are quite different. For example, Angrist and Lavy (1999) estimate the effect of class size on student achievement using instruments constructed from Maimonides' rule, a bureaucratic ceiling on class size that induces sharp differences in average class size in Israeli schools. Ordinary least squares estimates in this case show either no effect or a beneficial effect of larger classes, while the corresponding instrumental variable estimates reveal a statistically significant and sizeable beneficial effect of smaller classes.

A common criticism of the natural experiments approach to instrumental variables is that it does not spell out fully the underlying theoretical relationships (for example, Rosenzweig and Wolpin, 2000). In the Angrist and Krueger (1991) application, for example, the theoretical relationship between education and earnings is not developed from an elaborate theoretical model; instead, it depends on the institutional details of the education system. Nevertheless, interest in the causal effect of education on earnings is easy to motivate. Moreover, the natural experiments approach to instrumental variables is fundamentally grounded in theory, in the sense that there is usually a well-developed story or model motivating the choice of instruments. These stories have implications that can be used to support or refute a behavioral interpretation of the resulting instrumental variables estimates.

For example, the interpretation of the patterns in Figures 1 and 2 as resulting from the interaction of school start-age policy and compulsory schooling is supported by our finding that quarter of birth is unrelated to earnings and educational attainment for those with a college degree or higher. This group is unconstrained by compulsory schooling laws, so if quarter of birth was related to education or earnings in this sample, the rationale motivating the use of quarter of birth as an instrument would have been refuted. The finding that the identification strategy was not refuted by this test suggests that factors other than compulsory schooling are not responsible for the correlation between education and the instrument in the full sample.

The rich implications and potential refutability of instrumental variables analyses based on natural experiments are an important part of what makes the approach attractive. We would argue that this approach contrasts favorably with studies that provide detailed but abstract theoretical models, followed by identification based on implausible and unexamined choices about which variables to exclude from the model and assumptions about what statistical distribution certain variables follow. Indeed, one of the most mechanical and naïve, yet common, approaches to the choice of instruments uses atheoretical and hard-to-assess assumptions about dynamic relationships to construct instruments from lagged vari-

ables in time series or panel data. The use of lagged endogenous variables as instruments is problematic if the equation error or omitted variables are serially correlated. In this regard, Wright's (1928) use of the more plausible exogenous instrument "yield per acre" seems well ahead of its time.

Interpreting Estimates with Heterogeneous Responses

One difficulty in interpreting instrumental variables estimates is that not every observation's behavior is affected by the instrument. As we have stressed, instrumental variables methods can be thought of as operating by using only part of the variation in an explanatory variable—that is, by changing the behavior of only some people. For example, in the Angrist and Krueger (1991) study just discussed, the quarter-of-birth instrument is most relevant for those who are at high probability of quitting school as soon as possible, with little or no effect on those who are likely to proceed on to college.

Another example that makes this point is Angrist's (1990) use of Vietnam-era draft lottery numbers as an instrument to estimate the effect of military service on earnings later in life. The draft lottery numbers randomly assigned to young men in the early 1970s were highly correlated with the probability of being drafted into the military, but not correlated with other factors that might change earnings later. The military draft presumably affected the behavior of those who would not have joined the military otherwise. But most of those who served in Vietnam were true volunteers who would have served regardless of their draft lottery number. An estimate using draft lottery numbers as instruments is therefore based on the experience of draftees only. This may not capture the effects of military service on the civilian earnings of volunteers.

In other words, instrumental variables provide an estimate for a specific group—namely, people whose behavior can be manipulated by the instrument. The quarter-of-birth instruments used by Angrist and Krueger (1991) generate an estimate for those whose level of schooling was changed by that instrument. Similarly, the draft lottery instrument provides estimates of a well-defined causal effect for a subset of the treated group: men whose behavior was changed by the draft lottery "experiment."

This issue arises in many studies using instrumental variables, and it is discussed formally in papers by Imbens and Angrist (1994) and related work. They show that with a dummy endogenous variable, instrumental variables methods estimate causal effects for those whose behavior would be changed by the instrument if it were assigned in a randomized trial. That is, the effect is estimated for subjects who will take the treatment if assigned to the treatment group, but otherwise not take the treatment. This parameter is known as the Local Average Treatment Effect, or LATE, for short. In some cases, the experiences of the group of "compliers" captured by LATE are representative of those of the entire treated group. If everyone in the population has the same response to a particular inter-

vention or treatment, as is commonly assumed, the distinction between LATE and other parameters does not matter. But with “heterogeneous treatment effects,” the parameter identified by instrumental variables may differ from the average effect of interest.⁸

It should be noted that this specificity of estimates is endemic to empirical research. All statistical methods, from the simplest regressions to the most complex structural models, have elements of this limitation when used to analyze phenomena with heterogeneous responses. Nevertheless, many interventions and relationships can be fruitfully studied using estimated effects for specific subsamples, provided the possible limitations to generalizing the results are understood and explored. Indeed, this lack of immediate generality is probably the norm in medical research based on clinical trials, yet much progress has been made in that field.

Our view is that instrumental variables methods often solve the first-order problem of eliminating omitted variables bias for a well-defined population. Since the sample size and range of variability in many empirical studies are quite limited, extrapolation to other populations is naturally somewhat speculative and often relies heavily on theory and common sense. (A fertilizer that helps corn to grow in Iowa will probably have a beneficial effect in California as well, though one can't be sure.) Moreover, the existence of heterogeneous treatment effects would be a reason for analyzing more natural experiments, not fewer, to understand the source and extent of heterogeneity in the effect of interest.

It is also worth emphasizing that the population one learns about in a natural experiment is often of intrinsic interest. For example, the Angrist and Krueger (1991) instrumental variables estimates, which are identified by differences in schooling for people affected by compulsory schooling laws, are relevant for assessing the economic rewards to increases in schooling induced by legal and institutional changes from policies designed to keep children from dropping out of high school. In the case of the draft lottery, even if the instrumental variables estimates do not necessarily tell us about the effect of military experience on earnings for both volunteers and draftees, knowing the effect of being drafted on later civilian earnings is important. Moreover, the story behind the instrumental variables analysis often opens up other avenues of inquiry. For example, Angrist (1990) interpreted the civilian earnings penalty associated with Vietnam-era service as due to a loss of labor market experience. If true, the resulting estimates have predictive validity for the consequences of compulsory military service in other times and places.

⁸ The theoretical result that instrumental variables methods identify LATE requires a technical assumption known as “monotonicity.” This assumption means that the instrument only moves the endogenous regressor in one direction. With draft lottery instruments, for example, monotonicity implies that being draft-eligible makes a person at least as likely to serve in the military as he would be if he were draft-exempt. This seems reasonable, and is automatically satisfied by traditional latent index models for endogenous treatments.

Potential Pitfalls

What can go wrong with instrumental variables? The most important potential problem is a bad instrument, that is, an instrument that is correlated with the omitted variables (or the error term in the structural equation of interest in the case of simultaneous equations). Especially worrisome is the possibility that an association between the instrumental variable and omitted variables can lead to a bias in the resulting estimates that is much greater than the bias in ordinary least squares estimates. Moreover, seemingly appropriate instruments can turn out to be correlated with omitted variables on closer examination. For example, the weather in Brazil probably shifts the supply curve for coffee, providing a plausible instrument to estimate the effect of price on quantity demanded. But weather in Brazil might also shift the demand for coffee if sophisticated commercial buyers at the New York Coffee, Sugar and Coca Exchange, where coffee futures are traded, use weather data to adjust holdings in anticipation of price increases that may not materialize in fact.

Another concern is the possibility of bias when instruments are only weakly correlated with the endogenous regressor(s). This possibility was first noted by Nagar (1959) and emphasized by Bound, Jaeger and Baker (1995). In fact, instrumental variables estimates with very weak instruments tend to be centered on the corresponding ordinary least squares estimate (Sawa, 1969).

Several solutions to the weak instruments problem have been proposed. First, the bias of two-stage least squares is proportional to the degree of overidentification. In other words, if K instruments are used to estimate the effect of G endogenous variables, the bias is proportional to $K - G$. Using fewer instruments therefore reduces bias. In fact, if the number of instruments is equal to the number of endogenous variables, the bias is approximately zero. A variety of technical fixes and diagnostic tests have also been proposed for the weak instrument problem.⁹

Concerns about weak instruments can be mitigated most simply by looking at the reduced form equation, that is, the ordinary least squares regression of the dependent variable of interest on the instruments and exogenous variables. These estimates are unbiased, even if the instruments are weak. Because the reduced form effects are proportional to the coefficient of interest, one can determine the sign of the coefficient of interest and guesstimate its magnitude by rescaling the reduced

⁹ One solution is to use the Limited Information Maximum Likelihood (LIML) estimator. Although LIML and two-stage least squares have the same asymptotic distribution and are algebraically equivalent in just-identified models, in overidentified models their finite-sample distributions can be very different. Most importantly, LIML is approximately unbiased in the sense that the *median* of its sampling distribution is generally close to the population parameter being estimated (Anderson et al., 1982). Alternatives to LIML include the approximately unbiased split-sample and jackknife instrumental variables estimators (Angrist and Krueger, 1995; Angrist, Imbens, and Krueger, 1999; Blomquist and Dahlberg, 1999), bias-corrected estimation (Sawa, 1973; Bekker, 1994), inference procedures discussed by Staiger and Stock (1997) and Hahn and Hausman (1999), and Bayesian smoothing of the first stage (Chamberlain and Imbens, 1996).

form using plausible assumptions about the size of the first-stage coefficient(s). Most importantly, if the reduced form estimates are not significantly different from zero, the presumption should be that the effect of interest is either absent or the instruments are too weak to detect it. The plausibility of the magnitude of the reduced form effects should also be considered.

We conclude our review of pitfalls with a discussion of functional form issues for both the first and second stages in two-stage least squares estimation. Researchers are sometimes tempted to use probit or logit to generate first-stage predicted values in applications with a dummy endogenous regressor. But this is not necessary and may even do some harm. In two-stage least squares, consistency of the second-stage estimates does not turn on getting the first-stage functional form right (Kelejian, 1971). So using a linear regression for the first-stage estimates generates consistent second-stage estimates even with a dummy endogenous variable. Moreover, using a nonlinear first stage to generate fitted values that are plugged directly into the second-stage equation does not generate consistent estimates unless the nonlinear model happens to be *exactly* right, a result which makes the dangers of misspecification high.¹⁰

Nonlinear second-stage estimates with continuous or multivalued regressors are similarly tricky, requiring a correctly specified functional form in order to interpret the estimates easily. Even if the underlying second-stage relationship is nonlinear, linear instrumental variables estimates such as two-stage least squares typically capture an average effect of economic interest analogous to the LATE parameter for dummy endogenous regressors (Angrist and Imbens, 1995; Card, 1995; Heckman and Vytlačil, 1998). Thus, two-stage least squares is a robust estimation method that provides a natural starting point for instrumental variables applications. The importance of functional form issues can be assessed in a more detailed secondary analysis by experimenting with alternative instruments and examining suitable graphs.

Nature's Stream of Experiments

Trygve Haavelmo (1944, p. 14) drew an analogy between two sorts of experiments: “those we should like to make” and “the stream of experiments that nature is steadily turning out from her own enormous laboratory, and which we merely watch as passive observers.” He also lamented, “[U]nfortunately—most economists do not describe their designs of experiments explicitly.” The defining characteristic

¹⁰ To be more specific, researchers risk specification error if they *plug in* fitted values from a logit, probit, or other nonlinear equations directly in the second step of a two-stage least squares procedure. Fitted values from a nonlinear model may still be used *as an instrument* for an endogenous dummy variable, provided a linear model is used to generate first-stage predictions of the endogenous dummy variable from these nonlinear fitted values and all other exogenous covariates in the second-stage equation.

of many recent applications of instrumental variables to the omitted variables problem is the attention devoted to describing and to assessing the underlying quasi-experimental design. This can be seen as an explicit attempt to use observational data to mimic randomized experiments as closely as possible.

Some economists are pessimistic about the prospects of finding a substantial number of useful natural experiments. Michael Hurd (1990), for example, called the search for natural experiments to test the effect of Social Security on labor supply “overly cautious” and warned, “if applied to other areas of empirical work [this method] would effectively stop estimation.” We make no claim that natural experiments are the only way to obtain useful results, only that they have the potential to increase our understanding of important economic relationships greatly. Table 1 provides a sampling of some recent studies that have used instrumental variables techniques to analyze a natural experiment or a researcher-generated randomized experiment. It is hard to conclude that empirical work has effectively stopped.

The first panel in Table 1 illustrates the breadth of application of the natural experiments idea in recent empirical work. Other examples can be found in the surveys by Meyer (1995) and Rosenzweig and Wolpin (2000). Some of the examples are more convincing than others. But all are distinguished by a serious attempt to substantiate the underlying assumptions used to infer causality. There is more “theory” behind these attempts than in many ostensibly structural models where the justification for including or excluding certain variables is neither explicitly described nor evaluated.

The second panel in Table 1 illustrates another important development: the use of instrumental variables in randomized experiments. Instrumental variables are useful in experiments when, either because of practical or ethical considerations, there is incomplete compliance in the treatment or control groups. In randomized evaluations of training programs, for example, some treatment group members may decline training while some control group members may avail themselves of training through channels outside the experiment. Similarly, in medical trials, doctors may be willing randomly to offer, but not to impose, incentives that change behaviors like smoking or taking a new medication.

Even in experiments with compliance problems, instrumental variables can be used to estimate the effect of interventions such as job training or medical treatments. The instrumental variable in such cases is a dummy variable indicating randomized assignment to the treatment or control group, and the endogenous right-hand-side variable is an indicator of actual treatment status. In a training evaluation, for example, the actual treatment status variable would be a dummy variable that equals one for each treatment and control group member who participated in training, and it would be zero for all those who did not participate in training. This approach yields a consistent estimate of the causal effect of the treatment for the population that complies with their random assignment—that is, the population of “compliers” (Imbens and Angrist, 1994). As in natural experiments, the instrument is used to exploit an exogenous source of variation—created

Table 1

Examples of Studies That Use Instrumental Variables to Analyze Data From Natural and Randomized Experiments

| <i>Outcome Variable</i> | <i>Endogenous Variable</i> | <i>Source of Instrumental Variable(s)</i> | <i>Reference</i> |
|----------------------------------|--|--|---------------------------------------|
| <i>1. Natural Experiments</i> | | | |
| Labor supply | Disability insurance replacement rates | Region and time variation in benefit rules | Gruber (2000) |
| Labor supply | Fertility | Sibling-Sex composition | Angrist and Evans (1998) |
| Education, Labor supply | Out-of-wedlock fertility | Occurrence of twin births | Bronars and Grogger (1994) |
| Wages | Unemployment insurance tax rate | State laws | Anderson and Meyer (2000) |
| Earnings | Years of schooling | Region and time variation in school construction | Duflo (2001) |
| Earnings | Years of schooling | Proximity to college | Card (1995) |
| Earnings | Years of schooling | Quarter of birth | Angrist and Krueger (1991) |
| Earnings | Veteran status | Cohort dummies | Imbens and van der Klaauw (1995) |
| Earnings | Veteran status | Draft lottery number | Angrist (1990) |
| Achievement test scores | Class size | Discontinuities in class size due to maximum class-size rule | Angrist and Lavy (1999) |
| College enrollment | Financial aid | Discontinuities in financial aid formula | van der Klaauw (1996) |
| Health | Heart attack surgery | Proximity to cardiac care centers | McClellan, McNeil and Newhouse (1994) |
| Crime | Police | Electoral cycles | Levitt (1997) |
| Employment and Earnings | Length of prison sentence | Randomly assigned federal judges | Kling (1999) |
| Birth weight | Maternal smoking | State cigarette taxes | Evans and Ringel (1999) |
| <i>2. Randomized Experiments</i> | | | |
| Earnings | Participation in job training program | Random assignment of admission to training program | Bloom et al. (1997) |
| Earnings | Participation in Job Corps program | Random assignment of admission to training program | Burghardt et al. (2001) |
| Achievement test scores | Enrollment in private school | Randomly selected offer of school voucher | Howell et al. (2000) |
| Achievement test scores | Class size | Random assignment to a small or normal-size class | Krueger (1999) |
| Achievement test scores | Hours of study | Random mailing of test preparation materials | Powers and Swinton (1984) |
| Birth weight | Maternal smoking | Random assignment of free smoker's counseling | Permutt and Hebel (1989) |

by explicit random assignment in these cases—to estimate the effect of interest. The use of such researcher-generated instruments is growing and reflects the accelerating convergence of classical experimentation and observational research methods.

Our view is that progress in the application of instrumental variables methods depends mostly on the gritty work of finding or creating plausible experiments that can be used to measure important economic relationships—what statistician David Freedman (1991) has called “shoe-leather” research. Here the challenges are not primarily technical in the sense of requiring new theorems or estimators. Rather, progress comes from detailed institutional knowledge and the careful investigation and quantification of the forces at work in a particular setting. Of course, such endeavors are not really new. They have always been at the heart of good empirical research.

■ *We are grateful to Anders Björklund, Brad De Long, David Freedman, Kevin Hallock, Jerry Hausman, Timothy Taylor and Michael Waldman for helpful comments.*

References

- Anderson, Patricia and Bruce Meyer.** 2000. “The Effects of the Unemployment Insurance Payroll Tax on Wages, Employment, Claims, and Denials.” *Journal of Public Economics*. October, 78: 1-2, pp. 81–106.
- Anderson, T. W., Naoto Kunitomo and Takamitsu Sawa.** 1982. “Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator.” *Econometrica*. July, 50:4, pp. 1009–027.
- Angrist, Joshua D.** 1990. “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records.” *American Economic Review*. June, 80:3, pp. 313–36.
- Angrist, Joshua D. and William N. Evans.** 1998. “Children and their Parents’ Labor Supply: Evidence from Exogenous Variation in Family Size.” *American Economic Review*. June, 88:3, pp. 450–77.
- Angrist, Joshua D. and Guido W. Imbens.** 1995. “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity.” *Journal of the American Statistical Association*. 90:430, pp. 431–42.
- Angrist, Joshua D. and Alan B. Krueger.** 1991. “Does Compulsory School Attendance Affect Schooling and Earnings?” *Quarterly Journal of Economics*. November, 106:4, pp. 979–1014.
- Angrist, Joshua D. and Alan B. Krueger.** 1995. “Split-Sample Instrumental Variables Estimates of the Returns to Schooling.” *Journal of Business and Economic Statistics*. 13:2, pp. 225–35.
- Angrist, Joshua D. and Alan B. Krueger.** 1999. “Empirical Strategies in Labor Economics,” in *Handbook of Labor Economics, Volume IIIA*. Orley Ashenfelter and David Card, eds. Amsterdam: North-Holland, pp. 1277–1366.
- Angrist, Joshua D. and Victor Lavy.** 1999. “Using Maimonides’ Rule to Estimate the Effects of Class Size on Scholastic Achievement.” *Quarterly Journal of Economics*. May, 114:2, pp. 533–76.
- Angrist, Joshua D., Guido W. Imbens and Alan B. Krueger.** 1999. “Jackknife Instrumental Variables Estimation.” *Journal of Applied Econometrics*. 14:1, pp. 57–67.
- Bekker, Paul A.** 1994. “Alternative Approximations to the Distributions of Instrumental Variables Estimators.” *Econometrica*. 62:3, pp. 657–81.
- Blank, Rebecca.** 1991. “The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Re-

- view." *American Economic Review*. 81:5, pp. 1041–067.
- Blomquist, Sören and Matz Dahlberg.** 1999. "Small Sample Properties of LIML and Jackknife IV Estimators: Experiments with Weak Instruments." *Journal of Applied Econometrics*. January/February, 14:1, pp. 69–88.
- Bloom, Howard S. et al.** 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources*. Summer, 32:3, pp. 549–76.
- Bound, John, David Jaeger and Regina Baker.** 1995. "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association*. 90:430, pp. 443–50.
- Bowden, Roger J. and Darrell A. Turkington.** 1984. *Instrumental Variables*. Cambridge: Cambridge University Press.
- Bronars, Stephen G. and Jeff Grogger.** 1994. "The Economic Consequences of Unwed Motherhood: Using Twins as a Natural Experiment." *American Economic Review*. 84:5, pp. 1141–156.
- Burghardt, John et al.** 2001. *Does Job Corps Work? Summary of the National Job Corps Study*. Princeton, N.J.: Mathematica Policy Research, Inc.
- Card, David E.** 1995. "Earnings, Schooling and Ability Revisited," in *Research in Labor Economics, Volume 14*. Solomon W. Polachek, ed. Greenwich, Conn. and London: JAI Press, pp. 23–48.
- Chamberlain, Gary and Guido W. Imbens.** 1996. "Hierarchical Bayes Models with Many Instrumental Variables." Harvard University Department of Economics, Discussion Paper No. 1781, December.
- Duflo, Esther.** 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." *American Economic Review*. 91:4, pp. 795–813.
- Durbin, J.** 1954. "Errors in Variables." *Review of the International Statistical Institute*. 22, pp. 23–32.
- Evans, William N. and Jeanne Ringel.** 1999. "Can Higher Cigarette Taxes Improve Birth Outcomes?" *Journal of Public Economics*. April, 72:1, pp. 135–54.
- Freedman, David.** 1991. "Statistical Models and Shoe Leather," in *Sociological Methodology 1991*. Peter Marsden, ed. Washington, D.C.: American Sociological Association, chapter 10.
- Friedman, Milton.** 1957. *A Theory of the Consumption Function*. Princeton: Princeton University Press.
- Geary, Robert C.** 1949. "Determination of Linear Relations Between Systematic Parts of Variables with Errors of Observations, the Variances of Which are Unknown." *Econometrica*. January, 17, pp. 30–58.
- Goldberger, Arthur S.** 1972. "Structural Equation Methods in the Social Sciences." *Econometrica*. November, 40:6, pp. 979–1001.
- Gruber, Jonathan.** 2000. "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy*. December, 108:6, pp. 1162–183.
- Haavelmo, Trygve.** 1944. "The Probability Approach in Econometrics." *Econometrica*. July, 12: Supplement, pp. 1–115.
- Hahn, Jinyong and Jerry Hausman.** 1999. "A New Specification Test for the Validity of Instrumental Variables." MIT Working Paper No. 99-10, May.
- Heckman, James and Edward Vytlacil.** 1998. "Instrumental Variables Methods for the Correlated Random Coefficient Model: Estimating the Average Rate of Return to Schooling when the Return is Correlated with Schooling." *Journal of Human Resources*. 33:4, pp. 974–87.
- Howell, W. G. et al.** 2000. "Test-Score Effects of School Vouchers in Dayton, New York, and Washington: Evidence from Randomized Field Trials." Paper presented at the annual meeting of the American Political Science Association, Washington, D.C. September.
- Hurd, Michael D.** 1990. "Research on the Elderly: Economic Status, Retirement, and Consumption and Savings." *Journal of Economic Literature*. 28:2, pp. 565–637.
- Imbens, Guido W. and Joshua D. Angrist.** 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica*. 62:2, pp. 467–75.
- Imbens, Guido W. and Wilbert van der Klaauw.** 1995. "Evaluating the Cost of Conscription in the Netherlands." *Journal of Business and Economic Statistics*. April, 13:2, pp. 207–15.
- Kelejian, H. H.** 1971. "Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables." *Journal of the American Statistical Association*. 66, pp. 373–74.
- Kling, Jeffrey R.** 1999. "The Effect of Prison Sentence Length on the Subsequent Employment and Earnings of Criminal Defendants." Princeton University, Woodrow Wilson School, Discussion Paper 208.
- Krueger, Alan B.** 1999. "Experimental Estimates of Education Production Functions."

- Quarterly Journal of Economics*. May, 114:2, pp. 497–532.
- Levitt, Steven D.** 1997. “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime.” *American Economic Review*. June, 87:3, pp. 270–90.
- Maddala, G. S.** 1977. *Econometrics*. New York: McGraw-Hill.
- McClellan, Mark, Barbara J. McNeil and Joseph P. Newhouse.** 1994. “Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?” *Journal of the American Medical Association*. September, 272:11, pp. 859–66.
- Meyer, Bruce D.** 1995. “Natural and Quasi-Experiments in Economics.” *Journal of Business and Economic Statistics*. 13:2, pp. 151–61.
- Meyer, Bruce D., W. Kip Viscusi and David L. Durbin.** 1995. “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment.” *American Economic Review*. June, 85:3, pp. 322–40.
- Morgan, Mary S.** 1990. *The History of Econometric Ideas*. Cambridge: Cambridge University Press.
- Nagar, A. L.** 1959. “The Bias and Moment Matrix of the General k -Class Estimators of the Parameters in Simultaneous Equations.” *Econometrica*. October, 27, pp. 575–95.
- Permutt, Thomas and J. Richard Hebel.** 1989. “Simultaneous Equation Estimation in a Clinical Trial of the Effect of Smoking on Birth Weight.” *Biometrics*. June, 45, pp. 619–22.
- Powers, Donald E. and Spencer S. Swinton.** 1984. “Effects of Self-Study for Coachable Test Item Types.” *Journal of Educational Psychology*. April, 76:2, pp. 266–78.
- Reiersøl, Olav.** 1941. “Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis.” *Econometrica*. January, 9, pp. 1–24.
- Reiersøl, Olav.** 1945. “Confluence Analysis by Means of Instrumental Sets of Variables.” *Arkiv för Matematik, Astronomi och Fysik*. 32a:4, pp. 1–119.
- Rosenzweig, Mark R. and Kenneth I. Wolpin.** 2000. “Natural ‘Natural Experiments’ in Economics.” *Journal of Economic Literature*. December, 38:4, pp. 827–74.
- Sawa, Takamitsu.** 1969. “The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators.” *Journal of the American Statistical Association*. 64:327, pp. 923–37.
- Sawa, Takamitsu.** 1973. “An Almost Unbiased Estimator in Simultaneous Equations Systems.” *International Economic Review*. 14:1, pp. 97–106.
- Staiger, Douglas and James H. Stock.** 1997. “Instrumental Variables Regression with Weak Instruments.” *Econometrica*. 65:3, pp. 557–86.
- Theil, H.** 1953. “Repeated Least Squares Applied to Complete Equation Systems.” The Hague: Central Planning Bureau.
- van der Klaauw, Wilbert.** 1996. “A Regression-Discontinuity Evaluation of the Effect of Financial Aid Offers on College Enrollment.” Mimeo, Department of Economics, New York University.
- Wald, Abraham.** 1940. “The Fitting of Straight Lines if Both Variables are Subject to Error.” *Annals of Mathematical Statistics*. 11, pp. 284–300.
- Wright, Phillip G.** 1928. *The Tariff on Animal and Vegetable Oils*. New York: MacMillan.