13

Instrumental variables regression

Christopher Muller, Christopher Winship and Stephen L. Morgan

INTRODUCTION

Because of the presence of unobserved variables, social scientists are often unable to use standard methods such as stratification, matching, or regression to estimate the causal effect of a treatment D on an outcome Y. Instrumental variables (IV) regression provides one potential solution to this problem. IV regression involves finding an exogenous variable, the instrument Z, that affects Y only through the causal variable D. The researcher identifies the causal effect of D on Y by estimating the portion of variation in the outcome Y associated with the treatment D that is attributable to the exogenous variable Z.

In this chapter we introduce IV regression. Our discussion is based in part on the presentation in chapters 3 and 7 of Morgan and Winship (2007).[1] Our primary goal is to provide the reader with a good conceptual understanding of IV regression.[2] In order to provide conceptual clarity, we rely in part on directed acyclic graphs (DAGs) to represent the causal relationships among variables (see Pearl, 2000, 2009). Our secondary goal is to demonstrate the broad applicability of IV regression. To this end, we present a number of examples from economic history, several of which use natural experiments (Diamond and Robinson, 2010; Rosenzweig and Wolpin, 2000), rather than the standard examples from labor economics research.

We begin with a brief introduction to DAGs. We then discuss binary and continuous instruments. Throughout, we provide both algebraic and graphical introductions to IV regression. We then discuss whether IV assumptions can be tested with data, ways to adjudicate between ordinary least squares (OLS) and IV estimation, how IV assumptions can be relaxed, and two-stage least squares estimation. We close with a didactic example using European Social Survey (ESS) data on Italy and a brief consideration of some important limitations of IV regression.

INTRODUCTION TO THE METHOD

Introduction to directed acyclic graphs

In his 2000 book, *Causality: Models, Reasoning, and Inference*, Judea Pearl lays out a powerful and extensive graphical theory of causality. Here, we present only the elements that are relevant to understanding IV estimators. Pearl has shown that graphs provide a powerful way of thinking about causal systems of variables and the identification strategies that can be used to estimate the effects within them.

As with standard path models, the basic goal of drawing a causal system as a DAG is to represent explicitly all causes of the outcome of interest. Each symbol in a causal graph represents a random variable. To the reader acquainted with traditional linear path models, much of this material will be familiar. There are, however, important and subtle differences between traditional path models and DAGs.

In this chapter, symbols that are placed in brackets are unobserved random variables. All other symbols represent observed random variables. Causes are represented by directed edges $\rightarrow$ (i.e. single-headed arrows), such that an edge from one symbol to another signifies that the variable at the origin of the directed edge causes the variable at the terminus.[3] Missing arrows encode the identifying assumptions of an analysis. The absence of an arrow between two variables represents the strong assumption of the absence of a causal relation.

DAGs are more general than standard paths models in that the functional relationship between variables need not be linear.[4] However, unlike standard path models, a DAG does not permit the representation of simultaneous causation. Only directed edges are permissible: direct causation can run in only one direction, as in $X \rightarrow Y$. Furthermore, a DAG is defined to be an acyclic graph. Accordingly, no directed paths emanating from a causal variable can return to that variable.

In some circumstances, it is useful to use a curved and dashed bidirected edge as a shorthand device to indicate that two variables are mutually dependent on one or more (typically unobserved) common causes. When this representation is used, the resulting graph is no longer formally a DAG. Because the bidirected edge is simply a semantic substitution, however, the graph can usually be treated as if it were a DAG. Such shorthand can be helpful in suppressing a complex set of background causal relationships that are irrelevant to the empirical analysis at hand. Nonetheless, these bidirected edges should not be interpreted in any other way than as we have just stated. They are not indicators of mere correlations between the variables that they connect, nor do they signify the possibility that the two variables have a direct effect on each other.

Figure 13.1 uses DAGs to present the three basic patterns of causal relationships that can occur for any three variables that are related to each other: a chain of mediation, a fork of mutual dependence, and an inverted fork of mutual causation. The first two types of relationship are conventional. For the graph in Figure 13.1a, A affects B through A's causal effect on C and C's causal effect on B. This type of a causal chain renders the variables A and B unconditionally associated. For the graph in Figure 13.1b, A and B are both caused by C. Here, A and B are also unconditionally associated because they mutually depend on C.

For the third graph in Figure 13.1c, A and B are again connected by a pathway through C. But now A and B are both causes of C. Pearl labels C a 'collider' variable. Formally, a variable is a collider along a particular path if it has two arrows running into it. Figuratively, in the third graph the causal effects of A and B 'collide' with each other at C. Collider variables are common in social science applications: any endogenous variable that has two or more causes is a collider along some path.

A path containing a collider variable does not generate an unconditional association between the variables that cause the collider variable. For the mutual causation graph in Figure 13.1c, the pathway between A and B through C does not generate an unconditional association between A and B. Pearl's language is quite helpful here. The path $A \rightarrow C \leftarrow B$ does not generate an association between A and B because the collider variable C 'blocks' the possible causal effects of A and B on each other.

$$\begin{tikzcd} A \arrow[r] & C \arrow[r] & B \end{tikzcd}$$

(a) Mediation

$$\begin{tikzcd} & C \arrow[dl] \arrow[dr] & \\ A & & B \end{tikzcd}$$

(b) Mutual dependence

$$\begin{tikzcd} A \arrow[dr] & & B \arrow[dl] \\ & C & \end{tikzcd}$$

(c) Mutual causation

Figure 13.1 Basic patterns of causal relationships for three variables

The importance of considering collider variables is a key insight of Pearl's framework, and it is closely related to the familiar concerns of selecting on the dependent variable and conditioning on an endogenous variable.[5] Even though collider variables do not generate unconditional associations between the variables that determine them, the incautious handling of colliders can create conditional dependence that can sabotage a causal analysis. Understanding collider variables also helps clarify when conditioning strategies can and cannot be used to identify causal effects (Elwert and Winship, 2014).

One of the most common modeling strategies for estimating causal effects in quantitative research is to analyze a putative causal relationship within groups defined by one or more variables. Whether referred to as subgroup analysis, subclassification, stratification, or tabular decomposition, the usual motivation is to analyze the data after conditioning on membership in groups identified by values of a variable that is thought to be related to both the causal variable and the outcome variable.

From a graphical perspective, the result of such a modeling strategy is to generate simplified subgraphs for each subgroup or stratum of the data that correspond to each value of the conditioning variable. This procedure is analogous to disconnecting the conditioning variable from all other variables it points to in the original graph and rewriting the graph as many times as there are values for the conditioning variable. In the mutual dependence graph in Figure 13.1b, conditioning on C results in separate graphs for each value of C; in each of these subgraphs, A and B are disconnected. The reasoning here should be obvious: if analysis is carried out for a group in which all individuals have a particular value for the variable C, then the variable C is constant within the group and cannot be associated with A or B. Thus, from a graphical perspective, conditioning means transforming one graph into a simpler set of component graphs where fewer causes are represented.

As a technique for estimating a causal effect, conditioning is a very powerful and very general strategy. But one very important qualification must be noted: conditioning on a collider variable induces an association between its causes. Rather than eliminating it, conditioning on a collider creates association in the new subgraph where it previously did not exist.

The intuition behind this consequence of conditioning on a collider can be demonstrated with a simple example of selecting on the dependent variable. Consider a hypothetical university where in order to get tenure one must either work hard or do innovative work, but not necessarily both. Hard work and innovative work are separately sufficient to get tenure, but neither is necessary. Assume that among faculty at this hypothetical university there is no relationship between working hard and doing innovative work. Now consider only the faculty who are tenured. If a faculty member does not work hard, we know that he must do innovative work; similarly, if he does not do innovative work, we know that he must work hard. In short, among tenured faculty, working hard and doing innovative work are negatively associated.

Conditioning on a variable therefore has very different effects depending on whether the variable is a collider or a non-collider variable. In the first case, conditioning on a variable potentially eliminates association. In the second case, association is induced. A simple way of thinking about this is that if one conditions on a non-collider variable, this 'blocks' any association that is transmitted across the paths to which it is connected. Paths containing collider variables do not transmit association because a collider 'blocks' the path. However, if one conditions on a collider, this unblocks the paths of which they are a part, enabling these paths to transmit association.

Causal effect estimation with a binary instrumental variable

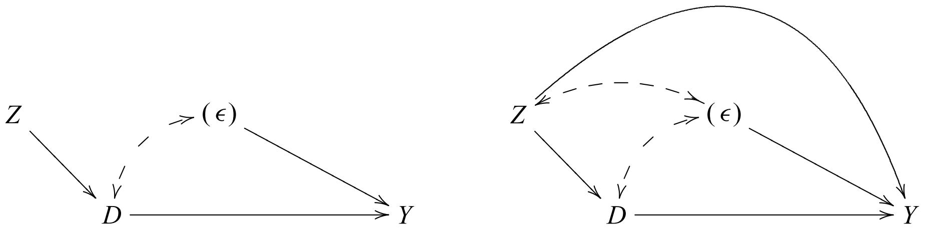
Consider the causal regression setup

$$Y = \alpha + \delta D + \epsilon, \tag{13.1}$$

where Y is the outcome variable, D is a binary treatment variable, α is an intercept, and ϵ is a summary random variable that stands for all other causes of Y. When equation (13.1) is considered to be a structural or causal model, the parameter δ represents the causal effect of D on Y. If δ does not have a subscript (i) either explicitly or implicitly, then the model indicates that the causal effect of D on Y is constant across individuals. As we discuss later in this chapter, this is a strong assumption that, if false, has important implications for how an IV estimate should be interpreted.

Now consider a binary variable, Z, such that the probability that D is equal to 1 rather than 0 is a function of Z. Figure 13.2 depicts two possible ways the variable Z could be related to both D and Y. Note first that, for both causal diagrams, the presence of the set of paths denoted by $D \leftarrow\text{----}\rightarrow \epsilon \rightarrow Y$ prevents a least squares regression of Y on D (or any other conditioning strategy) from generating a consistent estimate of the effect of D on Y. For the graph in Figure 13.2a, Z has an association with Y only through D. For the graph in Figure 13.2b, however, Z has an association with Y through D and also through common causes that determine both Z and ϵ.[6]

Consider how Z can be used to estimate the causal effect of D on Y in the first but not the second causal diagram. We know that for the set of causal relationships in both parts of Figure 13.2, the probability that D is equal to 1 rather than 0 is a function of the value of Z. But D still varies when conditioning on the value of Z because there are common causes that determine both D and ϵ. In addition, in Figure 13.2b, Z directly affects Y. In this case, Z should be included in equation (13.1) as a cause of Y. If we continue to maintain the assumption that the effect of D on Y is a constant structural effect δ, then it is not necessary to relate all of the variation in D to all of the variation in Y in order to obtain a consistent estimate of the causal effect. The



(a) Z is a valid instrumental variable for D

(b) Z is not a valid instrumental variable for D

Figure 13.2 Causal diagrams for Z as a potential IV

covariation in D and Y that is generated by the common causes of D and ϵ can be ignored if a way of isolating the variation in D and Y that is causal can be found.

Can the variation in D and Y that is generated by variation in Z be used to consistently estimate the causal effect of D on Y in this way? The answer to this question depends crucially on whether or not Z has an association with Y independent of its association with Y through D. For the graph in Figure 13.2a, this strategy will succeed because Z is associated with Y only through D. In the language of econometrics, the instrument satisfies the 'exclusion restriction' with respect to equation (13.1). For the graph in Figure 13.2b, this strategy will fail because Z and Y share common causes, as represented by the bidirected edge $Z \leftarrow\!\!-\!-\!-\!\!\rightarrow \epsilon$ and because Z has its own direct causal effect on Y. As a result, the variation that Z appears to generate in both D and Y is also generated by unobserved common causes of Z and by Z's direct effect on Y.

To see this result a bit more formally, take the population-level expectation of equation (13.1), $E[Y] = E[\alpha + \delta D + \epsilon] = \alpha + \delta E[D] + E[\epsilon]$, and rewrite it as a difference equation in Z:

$$\begin{aligned} &E[Y|Z=1] - E[Y|Z=0] \\ &\qquad = \delta(E[D|Z=1] - E[D|Z=0]) + (E[\epsilon|Z=1] - E[\epsilon|Z=0]). \end{aligned} \tag{13.2}$$

Equation (13.2) is now focused narrowly on the variation in Y, D, and ϵ that exists across levels of Z.[7] Now take equation (13.2) and divide both sides by $E[D|Z=1] - E[D|Z=0]$. This yields

$$\begin{aligned} &\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} \\ &\qquad = \frac{\delta(E[D|Z=1] - E[D|Z=0]) + (E[\epsilon|Z=1] - E[\epsilon|Z=0])}{E[D|Z=1] - E[D|Z=0]}. \end{aligned} \tag{13.3}$$

If the data are generated by the set of causal relationships depicted in Figure 13.2a, then Z has no association with ϵ, and $E[\epsilon|Z=1] - E[\epsilon|Z=0]$ in equation (13.3) is equal to 0. Consequently, the right-hand side of equation (13.3) simplifies to δ:

$$\frac{E[Y|Z=1] - E[Y|Z=0]}{E[D|Z=1] - E[D|Z=0]} = \delta. \tag{13.4}$$

Under these conditions, the ratio of the population-level association between Y and Z and between D and Z is equal to the causal effect of D on Y. This result suggests that, if Z is in fact associated with D but not associated with ϵ (or with Y, except through D), then the following sample-based estimator will equal δ in an infinite sample:

$$\hat{\delta}_{\text{IV, Wald}} = \frac{E_N[y_i|z_i = 1] - E_N[y_i|z_i = 0]}{E_N[d_i|z_i = 1] - E_N[d_i|z_i = 0]} \,. \tag{13.5}$$

As suggested by its subscript, this is the IV estimator, which is known as the Wald estimator when the instrument is binary. Although the Wald estimator is consistent for δ in this scenario, the assumption that δ is an invariant structural effect is crucial for this result.[8] As already noted, this is a strong assumption. We explain when and how this assumption can be relaxed later in this chapter.[9]

More insight can be gained about the Wald estimator by separately considering the definitions of the numerator and denominator in equation (13.4). The denominator is equal to what is known as the standard or naïve estimator of the causal effect of Z on D (Morgan and Winship, 2007, p. 44). That is, it is the mean outcome for D when $Z = 1$ minus the mean outcome for D when $Z = 0$. The numerator is equal to the standard or naïve estimator for the causal effect of Z on Y. That is, it is the mean outcome for Y when $Z = 1$ minus the mean outcome for Y when $Z = 0$. Equation (13.4) shows that the ratio of these two estimates provides an estimate of δ, the causal effect of D on Y.

For completeness, return to consideration of Figure 13.2b, in which Z has a non-zero association with ϵ. In order to simplify the algebra, assume that Z does not directly affect Y, but that it has a non-zero association with ϵ. In this situation, $E[\epsilon|Z = 1] - E[\epsilon|Z = 0]$ in equations (13.2) and (13.3) cannot equal 0, and thus (13.3) does not reduce further to (13.4). Instead, it reduces only to

$$\frac{E[Y|Z = 1] - E[Y|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} = \delta + \frac{E[\epsilon|Z = 1] - E[\epsilon|Z = 0]}{E[D|Z = 1] - E[D|Z = 0]} \,. \tag{13.6}$$

Here, the ratio of the population-level association between Y and Z and between D and Z does not equal the causal effect of D on Y but rather the causal effect of D on Y plus the last term on the right-hand side of equation (13.6). The Wald estimator in equation (13.5) is not consistent or asymptotically unbiased for δ in this situation. Instead, it converges to the right-hand side of equation (13.6), which is equal to δ plus a bias term that is a function of the net association between Z and ϵ.

A binary instrumental variable example from economic history

Banerjee and Iyer (2005) study the long-term effect of historical institutions on economic development. In particular, they are interested in the legacy of colonial land tenure systems established in British India. The authors propose that historically landlord-based districts will economically underperform historically non-landlord-based districts today because landlords' historical control over revenue collection generated a durable class-based antagonism that frustrates collective action.[10] They estimate the effect of historical non-landlord control over revenue collection on development today using the timing of British control as an instrumental variable.

Using historical evidence, Banerjee and Iyer (2005) make a strong case that there are no unobserved common causes of districts' historical use of landlord-based tenure systems and low levels of economic development today. In fact, they show that the districts with the lowest levels of historical development were those least likely to adopt a landlord system of revenue collection. Thus, if anything, OLS would be expected to underestimate the historical effect of a district's colonial land tenure system. The authors therefore estimate the effect of land tenure using both OLS and IV.

The authors argue that because of the influence of individual colonial administrators, districts where the British took over land revenue collection between 1820 and 1856 were much more likely to have non-landlord systems. Administrator Sir Thomas Munro, for instance, believed that the non-landlord *raiyatwari* system prevented landlords from absorbing surplus agricultural yields. Imposing the *raiyatwari* system, he claimed, would increase agricultural productivity by improving cultivators' incentives. After the Madras Board of Revenue, which favored landlords, prevailed over Munro, he 'traveled to London … and managed to convince the Court of Directors of the East India Company of the merits of the individual-based *raiyatwari* system; they then ordered the Madras Board of Revenue to implement this policy all over the province after 1820' (Banerjee and Iyer, 2005, p. 1195). Since the assignment of non-landlord land-tenure systems to Madras districts was based largely on the ideologies of individual administrators such as Munro until there was a policy reversal in 1857, Banerjee and Iyer (2005, p. 1191) conclude that 'areas where the land revenue collection was taken over by the British between 1820 and 1856 (but not before or after) are much more likely to have a non-landlord system, for reasons that have nothing to do with factors that directly influence agricultural investment and yields' (see also Wilson, 2011). This statement entails removing the bidirected edge $Z \leftarrow\!\!-\,-\,-\!\!\rightarrow \epsilon$ and the directed edge $Z \rightarrow Y$ from Figure 13.2b.

Banerjee and Iyer (2005) construct an instrument, Z, scoring one if a district was colonized between 1820 and 1856 and zero otherwise, and use this instrument to predict their treatment, D, the proportion of a district under non-landlord tenure.[11] Key to their IV identification strategy is the assumption, argued using historical evidence, that there is no bidirected edge connecting Z and their outcome Y, which indexes agricultural production, nor any direct effect of Z on Y. Munro and his allies, in other words, did not select districts for the *raiyatwari* system based on their inherent capacity for high agricultural yields.[12]

Table 13.1 reproduces the OLS and IV estimates for eight different dependent variables measuring agricultural investments and productivity. It shows that, consistent with the authors' hypothesis, the more districts historically relied on non-landlord systems of colonial land tenure, the more developed they are today. The coefficients of the IV estimates are larger than those of the OLS estimates, although not statistically significantly so. The authors argue that the OLS estimates are biased downward because the districts most likely to adopt non-landlord systems were initially the least agriculturally productive.

As we discuss later in this chapter, when district-level effect heterogeneity is present, the particular causal effect the IV estimates reported in Table 13.1 identify is not the average causal effect for all districts. Instead, these estimates identify the average causal effect for districts whose system of land tenure was dictated by the date of British control. Before discussing this interpretation of IV estimates in more detail, we turn to a more complete accounting of the traditional IV literature.

Traditional instrumental variable estimators

As detailed by Goldberger (1972), Bowden and Turkington (1984), and Heckman (2000), IV estimators were developed first in the 1920s by biologists and economists analyzing equilibrium price determination in market exchange (see Working, 1925, 1927; Wright, 1921, 1925). After subsequent development in the 1930s and 1940s (e.g. Haavelmo, 1942; Reiersol, 1941; Schultz, 1938; Wright, 1934), IV estimators were brought into widespread use in economics by researchers associated with the Cowles commission (see Hood and Koopmans, 1953; Koopmans and Reiersol, 1950). The structural equation tradition in sociology shares similar origins to that of the IV literature (see Duncan, 1975). The most familiar deployment of IV estimation in the

Table 13.1 Regression of agricultural investments and productivity on non-landlord proportion

Dependent variable$^{\alpha}$	OLS	IV
Agricultural investments		
Proportion of gross cropped area irrigated	0.065*	0.216
	(0.034)	(0.137)
Fertilizer use (kg/ha)	10.708***	26.198**
	(3.345)	(13.244)
Proportion of rice area under HYV	0.079*	0.411**
	(0.044)	(0.163)
Proportion of wheat area under HYV	0.092**	0.584***
	(0.046)	(0.163)
Proportion of other cereals area under HYV	0.057*	0.526***
	(0.031)	(0.129)
Agricultural productivity		
log (yield of 15 major crops)	0.157**	0.409
	(0.071)	(0.261)
log (rice yield)	0.171**	0.554*
	(0.081)	(0.285)
log (wheat yield)	0.229***	0.706***
	(0.067)	(0.214)
Number of districts	166	166

$^{\alpha}$Each cell represents the coefficient from a regression of the dependent variable on the non-landlord proportion. Standard errors, corrected for district-level clustering, are in parentheses. HYV = high-yielding varieties. All models include year fixed effects, geographic controls, and controls for the date of British land revenue control. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
Source: Banerjee and Iyer (2005).

extant sociological research is as the order condition for identification of a system of structural equations (see Bollen, 1989, 1995, 1996a,b, 2001; Fox, 1984).

Consider the same basic ideas presented earlier for the Wald estimator in equation (13.1):

$$Y = \alpha + \delta D + \epsilon. \tag{13.7}$$

The OLS estimator of the regression coefficient on D is

$$\hat{\delta}_{\text{OLS, bivariate}} \equiv \frac{\text{Cov}_N(y_i, d_i)}{\text{Var}_N(d_i)}, \tag{13.8}$$

where $\text{Cov}_N(\cdot)$ and $\text{Var}_N(\cdot)$ denote unbiased, sample-based estimates from a sample of size N of the population-level covariance and variance.

Again suppose that a correlation between D and ϵ renders the least squares estimator biased and inconsistent for δ in equation (13.7). If least squares cannot be used to effectively estimate δ, an alternative IV estimator can be attempted, with an instrument Z, as in

$$\hat{\delta}_{\text{IV}} \equiv \frac{\text{Cov}_N(y_i, z_i)}{\text{Cov}_N(d_i, z_i)} \tag{13.9}$$

where Z can now take on more than two values. If the instrument Z is correlated with D but uncorrelated with ϵ, then the IV estimator in equation (13.9) is consistent for δ in equation (13.7).[13]

One way to see why IV estimators yield consistent estimates is to again consider the population-level relationships between Y, D, and Z, as in equations (13.1)–(13.4). Manipulating equation (13.1) as before, one can write the covariance between the outcome Y and the instrument Z as

$$\text{Cov}(Y, Z) = \delta\text{Cov}(D, Z) + \text{Cov}(\epsilon, Z), \tag{13.10}$$

again assuming that δ is a constant structural effect. Dividing by $\text{Cov}(D, Z)$ then yields

$$\frac{\text{Cov}(Y, Z)}{\text{Cov}(D, Z)} = \frac{\delta\text{Cov}(D, Z) + \text{Cov}(\epsilon, Z)}{\text{Cov}(D, Z)}, \tag{13.11}$$

which is directly analogous to equation (13.3). When $\text{Cov}(\epsilon, Z)$ is equal to 0 in the population, then the right-hand side of equation (13.11) simplifies to δ. This suggests that

$$\frac{\text{Cov}(y_i, z_i)}{\text{Cov}(d_i, z_i)} \xrightarrow{p} \delta \tag{13.12}$$

if $\text{Cov}(\epsilon, Z) = 0$ in the population and if $\text{Cov}(D, Z) \neq 0$. This would be the case for the causal diagram in Figure 13.2a. But here the claim is more general and holds for cases in which Z is many-valued (and, in fact, for cases in which D is many-valued as well, assuming that the linear specification in equation (13.7) is appropriate).

In economic history, IVs are considered all-purpose remedies for least squares estimators that yield biased and inconsistent estimates, whether because of obvious omitted variables or subtle patterns of self-selection. In the following section we describe three examples of traditional IV estimation from economic history.

Traditional instrumental variables examples from economic history

Nunn (2008) studies the long-term effect of the slave trades on economic development in Africa, using African countries' overland and sailing distance from slaves' ultimate destination as instrumental variables. Drawing on historical documents and shipping records, he constructs estimates of the number of slaves exported from the territories of current African countries. Normalizing these estimates using countries' current land area, he finds a robust negative relationship between historical slave exports and per capita GDP in 2000.

One potential concern with using an OLS estimate to summarize the relationship between historical slave exports and current economic performance is that there might be an unobserved common cause of both the treatment and the outcome. Perhaps Europeans, for instance, sought out regions with historically low levels of development for involvement in the slave trades. Like Banerjee and Iyer (2005), Nunn (2008) uses historical evidence to show that selection, if anything, ran in the opposite direction of the treatment effect.[14]

As a second strategy for addressing his concern about omitted variables, Nunn (2008) uses instrumental variables. He measures the overland and sailing distance of each country from slaves' ultimate destination. The validity of this instrument rests on the assumption that slaves were taken, for example, from the west coast of Africa because this region was close to the West Indies. If instead the West Indies were selected as a slave destination because of their proximity to the west coast of Africa, the instrument, Z, would be an outcome rather than a cause of the causal variable D, rendering it invalid. This would be akin to removing the edge $Z \rightarrow D$ from Figure 13.2 and replacing it with the edge $Z \leftarrow$ D. Nunn (2008) provides considerable historical evidence that regions of slave demand were selected because their climates were suitable for sugar and tobacco cultivation or because they contained gold and silver mines,

Table 13.2 Regression of log real per capita GDP in 2000 on slave exports

Causal variable[α]	OLS	IV	OLS	IV
ln(exports/area)	–0.076***	–0.286*	–0.108***	–0.248***
	(0.029)	(0.153)	(0.037)	(0.071)
First-stage F-statistic		1.73		4.01
Number of observations	52	52	42	42

[α] Standard errors are in parentheses. All models include colonizer fixed effects and geographic controls. Columns three and four report estimates from the same model using a smaller sample excluding island and North African countries. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Source: Nunn (2008).

Table 13.3 Regression of log real per capita GDP in 1995 on current institutions

Causal variable[α]	OLS	IV
Average protection against expropriation risk	0.40	1.10
	(0.06)	(0.46)
Number of observations	64	64

[α] Standard errors are in parentheses. 'Average protection against expropriation risk 1985–1995' is the measure of current institutions. All models include indicator variables for Asia, Africa, and other continents as well as controls for latitude.

Source: Acemoglu et al. (2001).

not because they were geographically proximate to the regions that supplied slaves. Since an African country's distance from the historical location of slave demand should have no effect on its current development other than through its effect on the country's historical involvement in the slave trades, Nunn (2008) argues that the exclusion restriction is satisfied.

Table 13.2 reproduces the paper's main OLS and IV estimates. As in Banerjee and Iyer (2005), both the IV coefficients and standard errors are larger than the corresponding OLS coefficients and standard errors, but here the difference in the estimates is statistically significant.[15] The downward bias in the OLS estimates in Nunn (2008) should be expected given historical evidence that the most prosperous societies selected into the slave trades.

Papers estimating the effect of historical conditions such as slavery or colonial institutions on development often take their cue from an influential paper by Acemoglu et al. (2001). The authors estimate the effect of colonial institutions on the economic development of countries colonized by Europeans, using an index of settler mortality as an IV. The problems with using OLS estimation in this analysis should be familiar by now: 'It is quite likely that rich countries can choose or afford better institutions', or that 'economies that are different for a variety of reasons will differ both in their institutions and their income per capita' (Acemoglu et al., 2001, pp. 1369ff.). For Acemoglu et al. (2001), achieving causal identification therefore necessitates circumventing the bidirected edge, $D \dashleftarrow\dashrightarrow \epsilon \rightarrow Y$, connecting institutions, D, to development, Y, in Figure 13.2.

The authors propose that countries colonized by Europeans received one of roughly two types of institutions. The first type, *extractive* states, provided little support for property rights and few checks and balances against expropriation by elites or the government. The second type, *settler* states, following the European example, championed property rights and placed substantial constraints on executive power. Settler states promoted stronger institutions that, according to the argument, generate durable economic prosperity.

Whether a country received an extractive or a settler state, the authors argue, depended in part upon whether Europeans ultimately could settle there. Some countries' disease environment prohibited extensive European settlement. Assuming that there is a strong relationship between historical and current institutions, an index of European settler mortality can therefore be used as an instrument for current institutions. The edge $Z \to Y$ can be removed because the indigenous population had developed immunity to the diseases, such as malaria and yellow fever, that most commonly took the lives of European settlers. Table 13.3 reproduces the results of the most parameterized models reported in Acemoglu et al. (2001). As in the previous two examples, the IV estimates are larger and less precisely estimated than the OLS estimates, but they retain statistical significance.[16]

A final example of instrumental variables regression in economic history is provided by Tabellini (2010). Rather than investigate the effects of historical institutions, Tabellini estimates the effect of culture on economic performance, using two measures of historical institutional strength as IVs. Whereas the identification strategies of Banerjee and Iyer (2005), Nunn (2008), and Acemoglu et al. (2001) could be straightforwardly summarized using Figure 13.2, Tabellini's analysis is slightly more complex. We therefore depict the author's assumptions in the DAG drawn in Figure 13.3.

Like Banerjee and Iyer (2005), Tabellini (2010) studies regions rather than countries. The author constructs a regional measure of culture by extracting the principal component from individual respondents' answers to survey questions about their levels of trust, regard for values such as obedience and respect, and belief that they have control over their lives. Like the effect of institutions on development, the effect of culture on development might be confounded by an unobserved common cause of both variables. One possibility is that historical economic prosperity generated both more trust – as well as other relevant cultural attributes – *and* a higher level of current development. To block the possible backdoor path running from culture to output, Tabellini (2010) conditions on a measure of urbanization in 1850.[17] Figure 13.3 shows that conditioning on urbanization blocks the path Culture $\leftarrow$ Urbanization $\rightarrow$ Output.

The concern that a separate unobserved common cause of culture and output might confound the analysis motivates Tabellini's (2010) second identification strategy. Since the author is concerned that there are unobserved common causes of culture and current development other than historical development, we might also be concerned that there are unobserved common causes, U, of culture and historical development. Common causes, W, of historical and current development, moreover, might confound the relationship between urbanization and output. Figure 13.4 adds these unobserved causes to the DAG drawn in Figure 13.3. This DAG shows that in the presence of these unobserved causes, conditioning on Urbanization in 1850 will unblock the path Culture $\dashleftarrow U \dashrightarrow$ Urbanization $\dashleftarrow W \dashrightarrow$ Output on which it is a collider.

Aware of this limitation, Tabellini (2010) adopts an instrumental variables strategy. He constructs two instruments for culture: one measuring a region's historical literacy rate and the other measuring the quality of its historical institutions. The author acknowledges that historical literacy rates, in violation of the exclusion restriction, might have a direct effect on output. Reasoning that a region's historical literacy rate will only affect output through its effect on culture or current education, he conditions on current school enrollment to block the path Culture $\leftarrow$ Literacy $\rightarrow$ Enrollment $\rightarrow$ Output. In addition, all of the paper's regressions include country fixed effects. This rules out the possibility that differences in national historical institutions drive the results. Banerjee and Iyer (2005), however, show that historical institutions at the regional level can have an effect on output independent of what is captured by the measures of culture

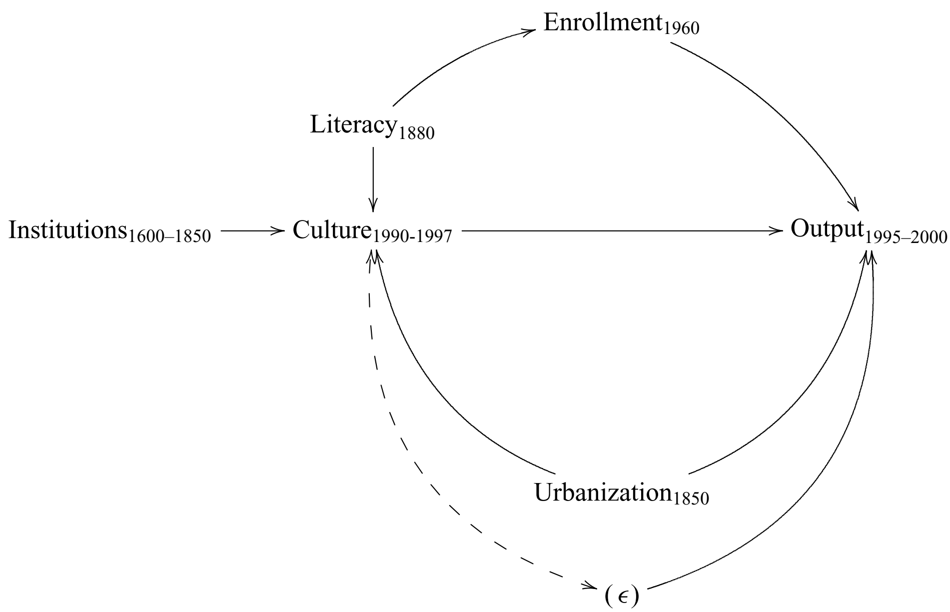


Figure 13.3 The assumptions of Tabellini (2010) in graphical form

used in Tabellini (2010). Were there a direct effect of historical regional institutions on output, the backdoor path Institutions $\dashrightarrow$ Output, as depicted in Figure 13.4, would remain unblocked, in violation of the exclusion restriction.

Our objective in this discussion is not to question the results reported in Tabellini (2010).[18] The plausibility of the identifying assumptions depends on the reader's substantive knowledge of the case. Rather, we aim to show how the simple graphical language of DAGs can render these assumptions transparent, better enabling readers to assess their validity. This is especially important for analyses using instrumental variables, since a commonly used test of the IV assumption yields misleading answers.[19] We turn to this issue in the next section.

ADVANCED ISSUES

Can data be used to test the IV identifying assumption?

The basic assumption underlying the three studies just summarized is that Z has no causal effect on Y except indirectly through D. It may seem intuitive that this assumption can be tested in the following way. If Z has no direct effect on Y, we may suppose that there will be no association between Z and Y conditional on D. Thus, if we regress Y on Z and D, we should find no association between Z and Y if the IV assumption is correct.

In Figure 13.2a, an unblocked path connects D and Y, but Z is a valid IV that identifies the causal effect of D on Y. The instrument is valid because it causes D and because it is unconditionally unassociated with the common cause of D and Y. To test this assumption, some authors estimate the relationship between Z and Y, using D as a control variable. The rationale for this test is that conditioning on D will block the indirect relationship between Z and Y through D. Accordingly, if the only association between Z and Y is indirect through D, there

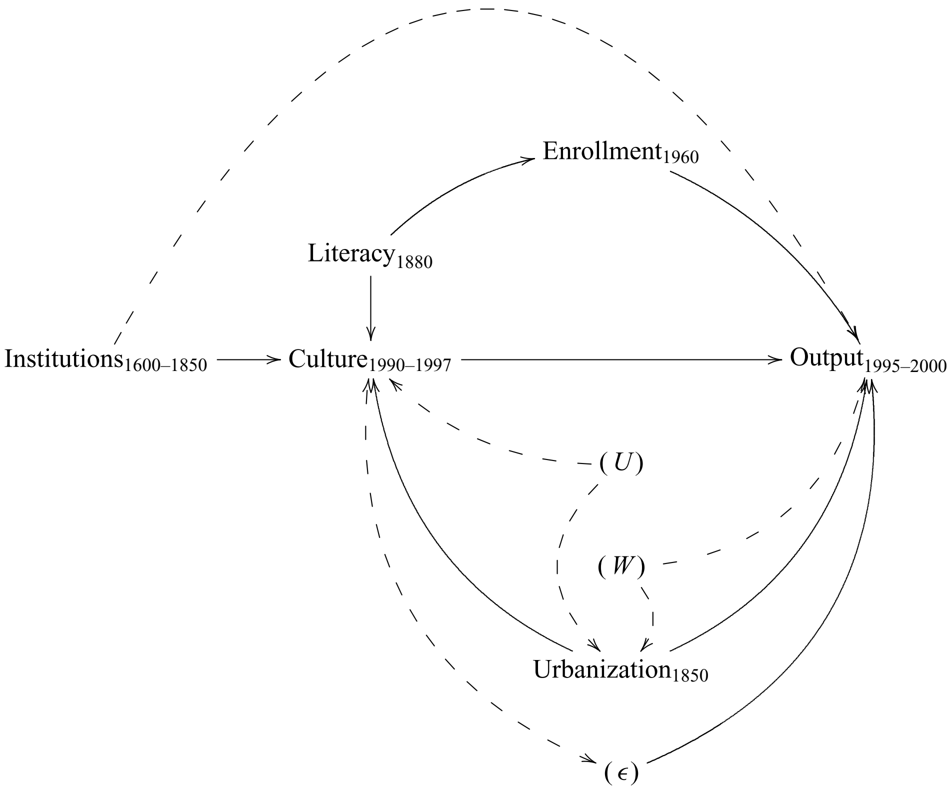


Figure 13.4 Relaxing the assumptions of Tabellini (2010)

should be no association between Z and Y after conditioning on D. If such a net association is detected, it would seem reasonable to conclude that the IV identifying assumption must be false.

It is certainly true that if the IV assumption is invalid, then Z and Y will be associated after conditioning on D. But the converse does not hold. Z and Y will always be associated after conditioning on D when the IV assumption is valid. This follows from the fact that D in Figure 13.2a is a collider that is mutually caused by Z and the unobserved common cause connecting D to Y. As we note in the previous section, conditioning on a collider induces an association between its causes. Thus conditioning on D in this DAG creates an association between Z and the unobserved common cause of D and Y, even though the IV identifying assumption is valid. Not finding a relationship between Z and Y after conditioning on D does not undermine the results of an IV analysis, since the only way not to observe an association in this regression is if the exclusion restriction is satisfied and there is no unobserved common cause confounding the relationship between D and Y (Elwert and Winship, 2014). But in this situation there is no need to use an IV to estimate the effect of D on Y. OLS will produce an unbiased and consistent estimate.[20]

OLS versus IV

The results reproduced in Tables 13.1–13.3 have an important commonality: since the IV estimator uses only a portion of the covariation between the treatment, D, and the outcome, Y, the IV estimates have far larger standard errors than the OLS estimates. If D is uncorrelated with ϵ, then the OLS estimates will be consistent and, according to the Gauss–Markov theorem, efficient, yielding smaller standard errors than the IV estimates. How, then, should we decide whether to use OLS or IV? Recall once more the causal regression setup introduced in equation (13.1),

$$Y = \alpha + \delta D + \epsilon, \tag{13.13}$$

as well as what is known as the first-stage equation predicting the treatment, D, using the instrumental variable Z,

$$D = \gamma + \beta Z + v. \tag{13.14}$$

If D is uncorrelated with ϵ, both the OLS and IV estimators will consistently estimate δ in equation (13.13). Our first intuition thus might be simply to compare the OLS and IV estimates. Different results would suggest that D is endogenous, leading us to prefer the IV estimator. If the estimates were similar, we would prefer OLS because it has smaller standard errors. The problem is that due to sampling error our estimates will never be exactly the same. Thus, we need a more formal test. This test is known as a Hausman test (Hausman, 1978).

Equation (13.14) allows us to segment D into two components: one, $\hat{D}$, that is a function of Z and therefore uncorrelated with ϵ; and another, $\hat{v}$, that may be associated with ϵ. If $\hat{v}$ is associated with Y, we have evidence that D is endogenous. We can test this formally by simply regressing Y on D and $\hat{v}$ and calculating a standard t test of the significance of the coefficient, λ, on $\hat{v}$:

$$Y = \alpha + \delta D + \lambda \hat{v} + w. \tag{13.15}$$

The estimate of δ in this equation will be equivalent to the IV estimate. In estimating the effect of D in equation (13.15), we are controlling for the component of D that may be associated with the error term, $\hat{v}$. This solves the endogeneity problem. If the effect of $\hat{v}$ is zero, however, then there is no need to control for it. In this case, OLS estimation will be both consistent and typically much more efficient.

IV estimation as LATE estimation

Following the adoption of a counterfactual perspective, a group of econometricians and statisticians clarified what IVs identify when unit-level causal effects are heterogeneous. In this section, we discuss the connections between traditional IV estimators and potential-outcome-defined treatment effects (Imbens and Angrist, 1994; Angrist and Imbens, 1995; Angrist et al., 1996; Imbens and Rubin, 1997). The key innovation here is the definition of a new treatment effect parameter: the local average treatment effect (LATE).

Until this point, we have assumed that the effect of D on Y is a constant structural effect. In this section, we relax this assumption and consider a different interpretation of the Wald estimator.[21] Recall the definition of Y in the fundamental problem of causal inference:

$$\begin{aligned} Y &= Y^0 + (Y^1 - Y^0)D \\ &= Y^0 + \delta D \\ &= \mu^0 + \delta D + v^0, \end{aligned} \tag{13.16}$$

where $\mu^0 \equiv E[Y^0]$ and $v^0 \equiv Y^0 - E[Y^0]$. Note that δ is now defined as $Y^1 - Y^0$, unlike its structural representation in equations (13.1) and (13.7) where δ was implicitly assumed to be constant across all units.

To understand when an IV estimator can be interpreted as an average causal effect estimator, Imbens and Angrist (1994) developed a counterfactual framework to classify units into those that respond positively to an instrument, those that remain unaffected by an instrument, and those that 'rebel' against an instrument. Their innovation was to define potential treatment assignment variables, $D^{Z=z}$, for each state z of the instrument Z. When D and Z are binary variables, there are four possible groups of units in the population.[22] These can be summarized by a four-category latent variable $\tilde{C}$ for compliance status:

$$\begin{aligned}
\text{compliers } (\tilde{C} = c): &\quad D^{Z=0} = 0 \text{ and } D^{Z=1} = 1,\\
\text{defiers } (\tilde{C} = d): &\quad D^{Z=0} = 1 \text{ and } D^{Z=1} = 0,\\
\text{always takers } (\tilde{C} = a): &\quad D^{Z=0} = 1 \text{ and } D^{Z=1} = 1,\\
\text{never takers } (\tilde{C} = n): &\quad D^{Z=0} = 0 \text{ and } D^{Z=1} = 0.
\end{aligned}$$

Although this terminology is best suited for describing experiments on individuals, it can accommodate natural experiments affecting aggregate units. Take Banerjee and Iyer (2005), for instance. Districts that adopted non-landlord systems only if their revenue control was taken over by the British between 1820 and 1856 can be defined as compliers ($\tilde{C} = c$). Districts that adopted non-landlord systems only if their revenue control was *not* taken over between 1820 and 1856 are defiers ($\tilde{C} = d$). Districts that would have adopted a non-landlord system irrespective of their date of revenue control are always takers ($\tilde{C} = a$). And districts that would never have adopted non-landlord systems are never takers ($\tilde{C} = n$).[23]

Analogous to the definition of the observed outcome, Y, the observed treatment indicator variable D can then be defined as

$$\begin{aligned}
D &= D^{Z=0} + (D^{Z=1} - D^{Z=0})Z \\
&= D^{Z=0} + \kappa Z,
\end{aligned} \tag{13.17}$$

where $\kappa \equiv D^{Z=1} - D^{Z=0}$. The parameter κ in equation (13.17) is the unit-level causal effect of the instrument on D. If the instrument represents encouragement to take the treatment, such as the adoption of non-landlord tenure systems in colonial India, then κ can be interpreted as the unit-level compliance inducement effect of the instrument. Accordingly, $\kappa = 1$ for compliers and $\kappa = -1$ for defiers. For always takers and never takers, $\kappa = 0$ because neither group responds to the instrument.

Given these definitions of potential outcome variables and potential treatment variables, a valid instrument Z for the causal effect of D on Y must satisfy three assumptions in order to identify a LATE:

$$\text{independence assumption: } (Y^1, Y^0, D^{Z=1}, D^{Z=0}) \perp\!\!\!\perp Z; \tag{13.18}$$

$$\text{nonzero effect of instrument assumption: } \kappa \neq 0 \text{ for all } i; \tag{13.19}$$

$$\text{monotonicity assumption: either } \kappa \geq 0 \text{ for all } i \text{ or } \kappa \leq 0 \text{ for all } i. \tag{13.20}$$

The independence assumption (13.18) is analogous to the assumption that $\text{Cov}(Z, \epsilon) = 0$ in the traditional IV literature. It stipulates that the instrument must be independent of the potential outcomes and potential treatments. Knowing the value of the instrument for unit i must not yield any information about the potential outcome of unit i under either treatment state. Moreover, knowing the realized value of the instrument for unit i must not yield any information about the probability of being in the treatment under the alternative hypothetical values of the instrument. A valid instrument predicts observed treatment status (D), but it does not predict potential treatment status ($D^{Z=z}$).

Assumptions (13.19) and (13.20) are about unit responses to shifts in the instrument. The assumption of a non-zero effect of Z on D is a stipulation that the instrument must predict treatment assignment for at least some units. There must be at least some compliers or some defiers in the population of interest. The monotonicity assumption further specifies that the effect of Z on D must be either weakly positive or weakly negative for all individuals i. Thus, there may be either defiers or compliers in the population, but not both.

If these three assumptions obtain, then an instrument Z identifies the LATE: the average causal effect of the treatment for the subset of the population whose treatment selection is induced by the instrument. If $\kappa \geq 0$ for all i, then the Wald estimator from equation (13.5) converges to a particular LATE:

$$\hat{\delta}_{\text{IV, Wald}} \xrightarrow{p} E[\delta|\tilde{C} = c], \tag{13.21}$$

which is equal to $E[Y^1 - Y^0|D^{Z=1} = 1, D^{Z=0} = 0]$ and is therefore the average causal effect among compliers. In contrast, if $\kappa \leq 0$ for all i, then the Wald estimator from equation (13.5) converges to the opposite of LATE:

$$\hat{\delta}_{\text{IV, Wald}} \xrightarrow{p} E[\delta|\tilde{C} = d], \tag{13.22}$$

which is equal to $E[Y^1 - Y^0|D^{Z=1} = 0, D^{Z=0} = 1]$ and is therefore the average causal effect among defiers. In either case, the treatment effects of always takers and never takers are not informed in any way by the IV estimate. Returning to Banerjee and Iyer (2005), we can thus interpret the IV estimates reproduced in Table 13.1 as the average causal effect for districts whose system of land tenure was set by the date when the British took control of revenue collection.

Two-stage least squares estimation

Most statistical software features packages that will automatically generate IV estimates. Identical point estimates can also be generated using what is known as two-stage least squares estimation. In the first stage, the analyst regresses the treatment, D, on the instrument, Z, all exogenous covariates, X, and a summary random variable, v, representing all other causes of D:

$$D = \gamma + \beta_1 Z + \beta_2 X + v. \tag{13.23}$$

In the second stage, the analyst uses the predicted values from the first-stage equation, $\hat{D}$, to predict the outcome conditional on the same exogenous covariates:

$$Y = \alpha + \delta_1 \hat{D} + \delta_2 X + \epsilon. \tag{13.24}$$

Often two-stage least squares is described as purging D of the component that is associated with v and then estimating the effect of D on Y using the purged D, or $\hat{D}$.

Because of its intuitive interpretation, we use a two-stage least squares framework to introduce our example below. However, in general it is not advisable to perform each stage manually, as the resulting standard errors and test statistics will be incorrect.[24]

EXAMPLE ANALYSIS

One of the primary theoretical motivations for the analysis presented in Tabellini (2010) is the claim, first advanced by Banfield (1958) and later studied by Putnam (1993), that differences

Table 13.4 Italian regions

Tabellini (2010)	ESS
Abruzzo – Molise – Basilicata	Abruzzo
	Molise
	Basilicata
Calabria	Calabria
Campania	Campania
Emilia-Romagna	Emilia-Romagna
Lazio	Lazio
Liguria	Liguria
Lombardia	Lombardia
Piemonte – Valle d'Aosta	Piemonte
	Valle d'Aosta
Puglia	Puglia
Sicilia – Sardegna	Sicilia
	Sardegna
Toscana	Toscana
Trentino-Alto Adige – Veneto – Fruili-Venezia Giulia	Trentino-Alto Adige
	Veneto
	Friuli-Venezia Giulia
Umbria – Marche	Umbria
	Marche

in economic development across the regions of Italy can be partially attributed to differences in citizens' generalized trust. In this didactic example, we use data from the European Social Survey (ESS) to test this hypothesis at the individual level. Our analysis thus provides a loose theoretical replication of the results reported in Tabellini (2010).

Despite a well-developed theoretical literature, the exact mechanisms by which Italian citizens' generalized trust might affect regional economic development remain incompletely understood. One possible way a lack of trust in others might hamper economic progress is by impeding borrowing and lending practices. Fortunately, the ESS enables us to test this hypothesis. All ESS survey respondents were asked to answer the following question: 'If for some reason you were in serious financial difficulties and had to borrow money to make ends meet, how difficult or easy would that be?' Respondents were given five choices, coded as follows: 'Very difficult' (1); 'Quite difficult' (2); 'Neither easy nor difficult' (3); 'Quite easy' (4); 'Very easy' (5). We treat these responses as continuous and run linear models throughout our analysis.

To predict citizens' beliefs about the ease or difficulty of borrowing money in times of need, we use the same trust measures used by Tabellini (2010). Both the World Values Survey and the ESS ask respondents: 'Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted.' We treat this, our causal variable, as continuous as well.

Because the relationship between generalized trust and beliefs about borrowing money might be confounded by an unobserved common cause, we condition our estimates on several additional variables measured at the individual level. Specifically, we control for respondents' age in years, gender (coded as 1 for men and 0 for women), nativity (1 if native; 0 otherwise), and education (divided into five ascending categories).

Next we merge our data set with that of Tabellini (2010), linking individuals to information about the historical region in which they reside. There are 20 Italian regions in the ESS data set, each of which falls into one of the 13 historical regions identified by Tabellini (2010). Table 13.4 shows how Italy's current regions were organized historically.

We follow Tabellini (2010) and use a measure of historical institutions at the regional level as an instrument for individual respondents' levels of trust.[25] Tabellini (2010) argues that because these historical institutions have no modern equivalent, they should not have a direct effect on development. The author uses country fixed effects to block the effect of national historical institutions; likewise we restrict our analysis to a single country. We emphasize that the identifying assumption here is a strong one: as shown by Banerjee and Iyer (2005), it is possible for historical institutions at the regional level to have a direct effect on development. It might be that in our data historical regional institutions affected both individuals' level of generalized trust and their expectations about how easy or difficult it would be to borrow money in times of need. If so, the exclusion restriction would be violated. It is also important to bear in mind that despite our large sample of individuals, our design effectively leaves us only 13 observations at the regional level.[26]

We carry out our IV analysis using a traditional two-stage least squares approach. In the first stage, we write respondents' (i) degree of generalized trust as a function of the quality of the institutions governing the historical region (r) in which they reside and a vector of individual-level covariates, X:

$$\text{trust}_{i,r} = \alpha_1 + \beta \text{institutions}_r + \gamma_1 X_{i,r} + v_{i,r}. \tag{13.25}$$

We then substitute the predicted values ($\widehat{\text{trust}}$) from equation (13.25) for individuals' actual survey responses:

$$y_{i,r} = \alpha_2 + \delta \widehat{\text{trust}_{i,r}} + \gamma_2 X_{i,r} + \epsilon_{i,r}, \tag{13.26}$$

where y is an individual's belief about the ease or difficulty of borrowing money in times of need. We adjust the standard errors in all models to account for the clustering of individuals in regions.

Table 13.5 reports parameter estimates from equations (13.25) and (13.26), as well the OLS equations that correspond to them. Column 1 reports the bivariate OLS relationship between an individual's belief that others can generally be trusted and their belief that borrowing money in times of need is easy. As expected, we observe a positive relationship between these beliefs. The more people subscribe to the belief that 'most people can be trusted', the easier they believe it is to borrow money in times of need. Column 2 reports the bivariate IV relationship. Both the coefficient and standard error increase by more than 10 times and the relationship remains statistically significant. We can interpret this coefficient as the local effect of the portion of the total variation in trust that is attributable to historical institutions. Columns 3 and 4 show that both the OLS and the IV estimates are robust to the inclusion of individual-level controls. The first-stage results indicate that historical institutions are a strong predictor of individual respondents' levels of generalized trust, as measured both by the significance of the coefficient on institutions and by the F-statistic testing the strength of the instrument.[27] Provided the exclusion restriction is satisfied, these results provide additional support for Tabellini's (2010) thesis that trust affects economic development. People in regions with stronger historical institutions have more faith that people in general can be trusted, and this leads them to have greater confidence that they will be able to borrow money in times of need.

PITFALLS OF TRADITIONAL IV ESTIMATION: WEAK INSTRUMENTS

By using only a portion of the covariation in the causal variable and the outcome variable, IV estimators use only a portion of the information in the data. This represents a direct loss in statistical power, and as a result IV estimates typically have substantially more expected

Table 13.5 Regression of borrowing beliefs on generalized trust

Independent variables$^{\alpha}$	OLS	IV	OLS	IV
Trust	0.063***	0.682**	0.040**	0.682**
	(0.011)	(0.165)	(0.010)	(0.169)
Age			0.005***	0.005
			(0.001)	(0.002)
Male			0.198***	0.211*
			(0.025)	(0.079)
Native			–0.030	0.079
			(0.159)	(0.144)
Education			0.234***	–0.042
			(0.032)	(0.072)
Intercept	2.670***	–0.061	1.869***	–0.365
	(0.100)	(0.761)	(0.160)	(0.616)
First stage				
Institutions		0.243***		0.225***
		(0.048)		(0.040)
Age				0.000
				(0.003)
Male				–0.018
				(0.103)
Native				–0.110
				(0.265)
Education				0.415***
				(0.038)
Intercept		4.715***		3.775***
		(0.065)		(0.311)
F - Statistic		25.76		31.76
Number of observations	2522	2522	2516	2516

$^{\alpha}$Standard errors, clustered by region, in parentheses. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

sampling variance than other estimators. The problem can be especially acute in some cases. It has been shown that instruments that only weakly predict the causal variable of interest should be avoided entirely, even if they generate point estimates with acceptably small estimated standard errors (Bound et al., 1995). In brief, the argument here is fourfold. (A) In finite samples, IV point estimates can always be computed because sample covariances are never exactly equal to zero. (B) As a result, an IV point estimate can be computed even for an instrument that is invalid because it does not predict the endogenous variable in the population (i.e. even if $\mathrm{Cov}(D, Z) = 0$ in the population, rendering equation (13.11) underdefined because its denominator is equal to 0). (C) At the same time, the formulas for calculating the standard errors of IV estimates fail in such situations, giving artificially small standard errors (when in fact the true standard error for the undefined parameter is infinity); and (D) the bias imparted by a small violation of the assumption that the IV affects the outcome variable only by way of the causal variable can explode if the instrument is weak.[28] To see this last result, consider equation (13.6), which depicts the expected bias in the Wald estimator for a binary IV as the term

$$\frac{E[\epsilon|Z=1] - E[\epsilon|Z=0]}{E[D|Z=1] - E[D|Z=0]}. \tag{13.27}$$

When the identifying assumption is violated, the numerator of equation (13.27) is non-zero because Z is associated with Y through ϵ. The bias is then an inverse function of the strength of the instrument; the weaker the instrument, the smaller the denominator and the larger the bias. If the denominator is close to zero, even a tiny violation of the identifying assumption can generate a large amount of bias. And, unfortunately, this relationship is independent of the sample size. Thus, even though a weak instrument may suggest a reasonable point estimate, and one with an acceptably small estimated standard error, the IV estimate may contain no genuine information whatsoever about the true causal effect of interest (see Hahn and Hausman, 2003; Staiger and Stock, 1997; Wooldridge, 2002).[29]

CONCLUSION

In this chapter we have provided an introduction to instrumental variables regression. We began by using directed graphs to demonstrate the logic underlying IVs. In order to show the broad applicability of IVs for applied researchers, we examined a number of examples of IV analysis in economic history. We noted the strong assumptions required for IVs to give consistent estimates of a causal effect and showed why a common approach to testing these assumptions is flawed. We introduced readers to a simple test that can be used to determine whether to prefer OLS or IV estimates and demonstrated that when the treatment effect is heterogeneous across units, IV analysis only estimates the treatment effect for compliers – those units that are affected by the instrument. We then introduced two-stage least squares estimation and presented a simple didactic example using the European Social Survey. Code to replicate this example in Stata and R is available in this chapter's online appendix. We closed with a brief discussion of some important limitations of IV regression.

FURTHER READING

The interested reader should consult more thorough and technical discussions in Angrist and Pischke (2009), Wooldridge (2002), Morgan and Winship (2007), Bollen (2012), and Murray (2006).

NOTES

1 Specifically, the chapter includes lightly edited material from pages 61–67, 187–190, 193–194, and 196–203 of Morgan and Winship (2007), but adds new discussions of the use of IVs in economic history.
2 Bollen (2012) provides an excellent discussion of many of the technical aspects of IV regression, including a broader treatment of the utility of instrumental variables in structural equation modeling.
3 In Pearl's framework, each variable is assumed to have an implicit probability distribution net of the causal effects reprented by the directed edges. This position is equivalent to assuming that background causes of each variable that exist are independent of the causes explicitly represented in the graph by directed edges.
4 For IV estimation, however, one requires what Pearl labels a linearity assumption. What this assumption means depends on the assumed distribution of the variables. For example, it would be satisfied in Figure 13.2 if the causal effect of Z on D is linear and the causal effect of D on Y is linear.
5 Selecting on the dependent variable and conditioning on an endogenous variable are two typical cases of the same fundamental problem: conditioning on the common outcome of two causes.
6 We have drawn the two bidirected edges separately for Figure 13.2b for simplicity. The same reasoning holds in more complex situations in which some of the common causes of Z and ϵ are also common causes of D and/or Y (and so on).
7 Equation (13.2) is generated in the following way. First, write $E[Y] = \alpha + \delta\, E[D] + E[\epsilon]$ conditional on the two values of Z, yielding $E[Y|Z = 1] = \alpha + \delta\, E[D|Z = 1] + E[\epsilon \mid Z = 1]$ and $E[Y\,|Z = 0] = \alpha + \delta\, E[D\,|Z = 0] +$

$E[\epsilon \mid Z = 0]$. Note that α and δ are constant structural effects and thus do not vary across individuals. Now, from $E[Y|Z = 1] = \alpha + \delta\, E[D|Z = 1] + E[\epsilon \mid Z = 1]$ subtract $E[Y|Z = 0] = \alpha + \delta\, E[D|Z = 0] + E[\epsilon \mid Z = 0]$. The parameter α is eliminated by the subtraction, and δ can be factored out of its two terms, resulting in equation (13.2).

8 As for the origin of the Wald estimator, it is customarily traced to Wald (1940) by authors such as Angrist and Krueger (1999). As we discuss later, the Wald estimator is only consistent, not generally unbiased, in a finite sample.

9 The reasons may well be clear to the reader already. In moving from equation (13.1) to equation (13.2), covariation in Y and D within levels of Z was purged from the equation. But if D has a causal effect on Y that varies across all individuals, it may be that whatever contrast is identified through Z may not be relevant to all members of the population.

10 All cultivable land in British India was governed by one of three systems: the *zamindari* or *malguzari* system, in which a class of landlords collected and set the terms of revenue for peasants under their jurisdiction; the *raiyatwari* system, under which revenue was collected directly from cultivators and usually varied according to annual outputs; and the *mahalwari* system, in which revenue was collected from collective village bodies.

11 In other specifications, they construct an indicator scoring one for districts mostly under non-landlord systems.

12 The authors include controls for the length of time under British rule to block a possible causal path from Z to Y.

13 Notice that substituting d_i for z_i in equation (13.9) results in the least squares regression estimator in equation (13.8). Thus, the OLS estimator implicitly treats D as an instrument for itself.

14 In order to be considered for trade with Europeans, societies needed to have passed a minimal threshold of institutional development. Societies with highly developed institutions tended to be densely populated and economically prosperous (Acemoglu et al., 2002). Thus, the societies most likely to enter the slave trade with Europeans were least likely to suffer from historical economic underdevelopment.

15 We describe a simple test of the difference between OLS and IV coefficients later in this chapter.

16 The authors attribute the larger IV estimates to measurement error in the institutions variables. See Card (2001) for a discussion of why IV estimates are consistently larger than OLS estimates in the literature on returns to schooling.

17 Urbanization is a standard proxy for historical economic development in economic history (see Acemoglu et al., 2005).

18 Indeed, we provide a loose theoretical replication of these results later in the chapter.

19 See Pearl (2009, pp. 274ff.) for an alternative test of the IV identifying assumption.

20 For completeness, consider what the test reveals when the IV assumption is invalid. Suppose that Figure 13.2a is augmented by an unobserved cause E and then two edges $E \to Z$ and $E \to \epsilon$. In this case Z and Y would be associated within levels of D for two reasons: (a) conditioning on the collider D generates a net association between Z and ϵ (and hence Z and Y); and (b) the common cause of E of Z and ϵ generates an unconditional association between Z and Y.

21 Here we denote potential outcome random variables for a binary cause D as Y^1 and Y^0. Accordingly, $Y = Y^1$ if $D= 1$ and $Y = Y^0$ if $D = 0$.

22 These four groups are considered principal strata in the framework of Frangakis and Rubin (2002).

23 In the IV regressions reported in Banerjee and Iyer (2005), the date of revenue control is actually used to predict the non-landlord *proportion* – not the dichotomous non-landlord indicator used elsewhere in the analysis.

24 See Gelman and Hill (2007, p. 223) for R code to correct two-stage least squares standard errors generated manually.

25 Tabellini (2010, pP. 697) uses a measure of ‘constraints on the executive’ to measure historical political institutions. This variable was constructed using information from historical sources.

26 This situation is akin to having only 13 group-level observations in a multilevel framework (Gelman and Hill, 2007) or an independent variable with low variance.

27 We discuss problems with weak instruments in the following section.

28 Complications (A)–(D) are all closely related. Situation (D) can be considered a less extreme version of the three-part predicament depicted in (A)–(C).

29 There is no consensus about how large an association between an IV and a treatment variable must be before an analysis can proceed safely. Many researchers rely on the rule of thumb that first-stage F-statistics testing the significance of the instrument should be around 10 or higher (Staiger and Stock, 1997; Stock et al., 2002).

REFERENCES

Acemoglu, D., Johnson, S. and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91, 1369–1401.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *Quarterly Journal of Economics*, 117, 1231–1294.

Acemoglu, D., Johnson, S. and Robinson, J. A. (2005). The rise of Europe: Atlantic trade, institutional change, and economic growth. *American Economic Review*, 95, 546–579.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association*, 90, 431–442.

Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91, 444–455.

Angrist, J. D. and Krueger, A. B. (1999). Empirical strategies in labor economics. In O. C. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, volume 3 (pp. 1277–1366). Amsterdam: Elsevier.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press.

Banerjee, A. and Iyer, L. (2005). History, institutions, and economic performance: the legacy of colonial land tenure systems in India. *American Economic Review*, 95, 1190–1213.

Banfield, E. C. (1958). *The Moral Basis of a Backward Society*. New York: Free Press.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: a least-squares estimator. *Sociological Methodology*, 25, 223–251.

Bollen, K. A. (1996a). An alternative two-stage least squares (2sls) estimator for latent variable equations. *Psychometrika*, 61, 109–121.

Bollen, K. A. (1996b). A limited-information estimator for lisrel models with or without heteroscedastic errors. In G. A. Marcoulides and R. E. Schumacker (eds), *Advanced Structural Equation Modeling: Issues and Techniques* (pp. 227–241). Mahwah, NJ: Lawrence Erlbaum.

Bollen, K. A. (2001). Two-stage least squares and latent variable models: simultaneous estimation and robustness to misspecifications. In R. Cudeck, S. du Toit and D. Sörbom (eds), *Structural Equation Modeling: Present and Future – a Festschrift in Honor of Karl Jöreskog* (pp. 119–138). Lincolnwood, IL: Scientific Software International.

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37–72.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous variable is weak. *Journal of the American Statistical Association*, 90, 443–450.

Bowden, R. J. and Turkington, D. A. (1984). *Instrumental Variables*. Cambridge: Cambridge University Press.

Card, D. (2001). Estimating the return to schooling: progress on some persistant econometric problems. *Econometrica*, 69, 1127–1160.

Diamond, J. and Robinson, J. A. (eds) (2010). *Natural Experiments of History*. Cambridge, MA: Harvard University Press.

Duncan, O. D. (1975). *Introduction to Structural Equation Models*. New York: Academic Press.

Elwert, F. and Winship, C. (2014). Endogenous selection bias. *Annual Review of Sociology*, 40, 31–53.

Fox, J. (1984). *Linear Statistical Models and Related Methods: With Applications to Social Research*. New York: Wiley.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58, 21–29.

Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica*, 40, 979–1001.

Haavelmo, T. (1942). The statistical implications of a system of simultaneous equations. *Econometrica*, 11, 1–12.

Hahn, J. and Hausman, J. (2003). Weak instruments: diagnosis and cures in empirical economics. *American Economic Review*, 93, 118–125.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46, 1251–1271.

Heckman, J. J. (2000). Causal parameters and policy analysis in economics: a twentieth century retrospective. *Quarterly Journal of Economics*, 115, 45–97.

Hood, W. C. and Koopmans, T. (eds) (1953). *Studies in Econometric Method*. New York: Wiley.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62, 467–475.

Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*, 64, 555–574.
Koopmans, T. C. and Reiersol, O. (1950). The identification of structural characteristics. *Annals of Mathematical Statistics*, 21, 165–181.
Morgan, S. L. and Winship, C. (2007). *Counterfactuals and Causal Inference*. New York: Cambridge University Press.
Murray, M. (2006). Avoiding invalid instruments and coping with weak instruments. *Journal of Economic Perspectives*, 20, 111–132.
Nunn, N. (2008). The long-term effects of Africa's slave trades. *Quarterly Journal of Economics*, 123, 139–176.
Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Putnam, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton, NJ: Princeton University Press.
Reiersol, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9, 1–24.
Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural 'natural experiments' in economics. *Journal of Economic Literature*, 38, 827–874.
Schultz, H. (1938). *The Theory and Measurement of Demand*. Chicago: University of Chicago Press.
Staiger, D. and Stock, J. H. (1997). IV regression with weak instruments. *Econometrica*, 65, 557–586.
Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, 20, 518–529.
Tabellini, G. (2010). Culture and institutions: economic development in the regions of Europe. *Journal of European Economic Association*, 8, 677–716.
Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, 11, 284–300.
Wilson, N. H. (2011). From reflection to refraction: State administration in British India, circa 1770-1855. *American Journal of Sociology*, 116, 1437–1477.
Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.
Working, E. J. (1927). What do statistical 'demand curves' show? *Quarterly Journal of Economics*, 41, 212–235.
Working, H. (1925). The statistical determination of demand curves. *Quarterly Journal of Economics*, 39, 503–545.
Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.
Wright, S. (1925). *Corn and Hog Correlations*. Washington, DC: U.S. Department of Agriculture.
Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161–215.