treatment effects

The term 'treatment effect' refers to the causal effect of a binary (0–1) variable on an outcome variable of scientific or policy interest. Economics examples include the effects of government programmes and policies, such as those that subsidize training for disadvantaged workers, and the effects of individual choices like college attendance. The principal econometric problem in the estimation of treatment effects is selection bias, which arises from the fact that treated individuals differ from the non-treated for reasons other than treatment status per se. Treatment effects can be estimated using social experiments, regression models, matching estimators, and instrumental variables.

A 'treatment effect' is the average causal effect of a binary (0–1) variable on an outcome variable of scientific or policy interest. The term 'treatment effect' originates in a medical literature concerned with the causal effects of binary, yes-or-no 'treatments', such as an experimental drug or a new surgical procedure. But the term is now used much more generally. The causal effect of a subsidized training programme is probably the mostly widely analysed treatment effect in economics (see, for example, Ashenfelter, 1978, for one of the first examples, or Heckman and Robb, 1985 for an early survey). Given a data-set describing the labour market circumstances of trainees and a non-trainee comparison group, we can compare the earnings of those who did participate in the programme and those who did not. Any empirical study of treatment effects would typically start with such simple comparisons. We might also use regression methods or matching to control for demographic or background characteristics.

In practice, simple comparisons or even regression-adjusted comparisons may provide misleading estimates of causal effects. For example, participants in subsidized training programmes are often observed to earn less than ostensibly comparable controls, even after adjusting for observed differences (see, for example, Ashenfelter and Card, 1985). This may reflect some sort of omitted variables bias, that is, a bias arising from unobserved and uncontrolled differences in earnings potential between the two groups being compared. In

general, omitted variables bias (also known as selection bias) is the most serious econometric concern that arises in the estimation of treatment effects. The link between omitted variables bias, causality, and treatment effects can be seen most clearly using the potential-outcomes framework.

Causality and potential outcomes

The notion of a causal effect can be made more precise using a conceptual framework that postulates a set of potential outcomes that could be observed in alternative states of the world. Originally introduced by statisticians in the 1920s as a way to discuss treatment effects in randomized experiments, the potential outcomes framework has become the conceptual workhouse for non-experimental as well as experimental studies in many fields (see Holland, 1986, for a survey and Rubin, 1974; 1977, for influential early contributions). Potential outcomes models are essentially the same as the econometric *switching regressions* model (Quandt, 1958), though the latter is usually tied to a linear regression framework. Heckman (1976; 1979) developed simple two-step estimators for this model.

Average causal effects

Except in the realm of science fiction, where parallel universes are sometimes imagined to be observable, it is impossible to measure causal effects at the individual level. Researchers therefore focus on average causal effects. To make the idea of an average causal effect concrete, suppose again that we are interested in the effects of a training programme on the post-training earnings of trainees. Let Y_{1i} denote the potential earnings of individual i if he were to receive training and let Y_{0i} denote the potential earnings of individual i if not. Denote training status by a dummy variable, D_i. For each individual, we observe $Y_i = Y_{0i} + D_i(Y_{1i} - Y_{0i})$, that is, we observe Y_{1i} for trainees and Y_{0i} for everyone else.

Let $E[\cdot]$ denote the mathematical expectation operator, i.e., the population average of a random variable. For continuous random variables, $E[Y_i] = \int yf(y)dy$, where $f(y)$ is the density of Y_i. By the law of large numbers, sample averages converge to population averages so we can think of $E[\cdot]$ as giving the sample average in very large samples. The two most widely studied

average causal effects in the treatment effects context are the average treatment effect (ATE), $E[Y_{1i} - Y_{0i}]$, and the average treatment effect on the treated (ATET), $E[Y_{1i} - Y_{0i}/D_i = 1]$. Note that the ATET can be rewritten

$$E[Y_{1i} - Y_{0i} \mid D_i = 1] = E[Y_{1i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 1].$$

This expression highlights the counter-factual nature of a causal effect. The first term is the average earnings in the population of trainees, a potentially observable quantity. The second term is the average earnings of trainees had they not been trained. This cannot be observed, though we may have a control group or econometric modelling strategy that provides a consistent estimate.

Selection bias and social experiments

As noted above, simply comparing those who are and are not treated may provide a misleading estimate of a treatment effect. Since the omitted variables problem is unrelated to sampling variance or statistical inference, but rather concerned with population quantities, it too can be efficiently described by using mathematical expectation notation to denote population averages. The contrast in average outcomes by observed treatment status is

$$\begin{aligned} E[Y_i \mid D_i = 1] \ - E[Y_i \mid D_i = 0] &= E[Y_{1i} \mid D_i = 1] \ - E[Y_{0i} \mid D_i = 0] \\ &= E[Y_{1i} \mid Y_{0i} \mid D_i = 1] + \{E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]\} \end{aligned}$$

Thus, the naive contrast can be written as the sum of two components, ATET, plus selection bias due to the fact that the average earnings of non-trainees, $E[Y_{0i}|D_i = 0]$, need not be a good stand-in for the earnings of trainees had they not been trained, $E[Y_{0i}|D_i = 1]$.

The problem of selection bias motivates the use of random assignment to estimate treatment effects in social experiments. Random assignment ensures that the potential earnings of trainees had they not been trained – an unobservable quantity – are well-represented by the randomly selected control group. Formally, when D_i is randomly assigned, $E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i} - Y_{0i}]$. Replacing $E[Y_i|D_i = 1]$ and $E[Y_i|D_i = 0]$ with the corresponding sample analog provides a consistent estimate of ATE.

Regression and matching

Although it is increasingly common for randomized trials to be used to estimate treatment effects, most economic research still uses observational data. In the absence of an experiment, researchers rely on a variety of statistical control strategies and/or natural experiments to reduce omitted variables bias. The most commonly used statistical techniques in this context are regression, matching, and instrumental variables.

Regression estimates of causal effects can be motivated most easily by postulating a constant-effects model, where $Y_{1i} - Y_{0i} = \alpha$ (a constant). The constant-effects assumption is not strictly necessary for regression to estimate an average causal effect, but it simplifies things to postpone a discussion of this point. More importantly, the only source of omitted-variables bias is assumed to come from a vector of observed covariates, X_i, that may be correlated with D_i. The key assumption that facilitates causal inference (sometimes called an identifying assumption), is that

$$E[Y_{0i} \mid X_i, D_i] = X'_i \beta, \tag{1}$$

where β is a vector of regression coefficients. This assumption has two parts. First, Y_{0i} (and hence Y_{1i}, given the constant-effects assumption) is mean-independent of D_i conditional on X_i. Second, the conditional mean function for Y_{0i} given X_i is linear. Given eq. (1), it is straightforward to show that

$$E\{Y_i(D_i - E[D_i \mid X_i])\} / E\{D_i(D_i - E[D_i \mid X_i])\} = \alpha. \tag{2}$$

This is the coefficient on D_i from the population regression of Y_i on D_i and X_i (that is, the regression coefficient in an infinite sample). Again, the law of large numbers ensures that sample regression coefficients estimate this population regression coefficient consistently.

Matching is similar to regression in that it is motivated by the assumption that the only source of omitted variables or selection bias is the set of observed covariates, X_i. Unlike regression, however, treatment effects are constructed by matching individuals with the same covariates instead of through a linear model for the effect of covariates. The key identifying assumption is also weaker, in that the effect of covariates on Y_{0i} need not be linear. Instead of (1), the conditional independence assumption becomes

4

$$E[Y_{ji} \mid X_i, D_i] = E[Y_{ji} \mid X_i], \text{ for } j = 0, 1. \tag{3}$$

This implies

$$\begin{aligned} E[Y_{1i} - Y_{0i} \mid D_i = 1] &= E\{E[Y_{1i} \mid X_i, D_i = 1] - [Y_{0i} \mid X_i, D_i = 1] \mid D_i = 1\} \\ &= E\{E[Y_{1i} \mid X_i, D_i = 1] - [Y_{0i} \mid X_i, D_i = 0] \mid D_i = 1\} \end{aligned} \tag{4a}$$

and, likewise,

$$E[Y_{1i} - Y_{0i}] = E\{E[Y_{1i} \mid X_i, D_i = 1] - [Y_{0i} \mid X_i, D_i = 0]\} \tag{4b}$$

In other words, we can construct ATET or ATE by averaging X-specific treatment-control contrasts, and then reweighting these X-specific contrasts using the distribution of X_i for the treated (for ATET) or using the marginal distribution of X_i (for ATE). Since these expressions involve observable quantities, it is straightforward to construct consistent estimators from their sample analogs.

The conditional independence assumption that motivates the use of regression and matching is most plausible when researchers have extensive knowledge of the process determining treatment status. An example in this spirit is the Angrist (1998) study of the effect of voluntary military service on the civilian earnings of soldiers after discharge, discussed further below.

Regression and matching details

In practice, regression estimates can be understood as a type of weighted matching estimator. If, for example, $E[D_i \mid X_i]$ is a linear function of X_i (as it might be if the covariates are all discrete), then it is possible to show that eq. (2) is equivalent to a matching estimator that weights cell-by-cell treatment-control contrasts by the conditional variance of treatment in each cell (Angrist, 1998). This equivalence highlights the fact that the most important econometric issue in a study that relies on conditional independence assumptions to identify causal effects is the validity of these conditional independence assumptions, not whether regression or matching is used to implement them.

A computational difficulty that sometimes arises in matching models is how to find good matches for each possible value of the covariates when the covariates take on many values. For

example, beginning with Ashenfelter (1978), many studies of the effect of training programmes have shown that trainees typically experience a period of declining earnings before they go into training. Because lagged earnings is both continuous and multidimensional (since more than one period's earnings seem to matter), it may be hard to match trainees and controls with exactly the same pattern of lagged earnings. A possible solution in this case is to match trainees and controls on the *propensity score*, the conditional probability of treatment given covariates. Propensity-score matching relies on the fact that, if conditioning on X_i eliminates selection bias, then so does conditioning on $P[D_i = 1|X_i]$, as first noted by Rosenbaum and Rubin (1983). Use of the propensity score reduces the dimensionality of the matching problem since the propensity score is a scalar, though in practice it must still be estimated. See Dehejia and Wahba (1999) for an illustration.

Regression and matching example

Between 1989 and 1992, the size of the military declined sharply because of increasing enlistment standards. Policymakers would like to know whether the people – many of them black men – who would have served under the old rules but were unable to enlist under the new rules were hurt by the lost opportunity for service. The Angrist (1998) study was meant to answer this question. The conditional independence assumptions seems plausible in this context because soldiers are selected on the basis of a few well-documented criteria related to age, schooling, and test scores and because the control group also applied to enter the military.

Naive comparisons clearly overestimate the benefit of military service. This can be seen in Table 1, which reports differences-in-means, matching, and regression estimates of the effect voluntary military service on the 1988–91 Social Security-taxable earnings of men who applied to join the military between 1979 and 1982. The matching estimates were constructed from the sample analog of (4a), that is, from covariate-value-specific differences in earnings, weighted to form a single estimate using the distribution of covariates among veterans. The covariates in this case were the age, schooling, and test-score variables used to select soldiers from the pool of applicants. Although white veterans earn $1,233 more than non-veterans, this difference becomes negative once the adjustment for differences in covariates is made. Similarly, while

non-white veterans earn $2,449 more than non-veterans, controlling for covariates reduces this to $840.

Table 1   **Matching and regression estimates of the effects of voluntary military service in the United States**

Race	Average earnings in 1988–91	Differences in means	Matching estimates	Regression estimates	Regression minus matching
	(1)	(2)	(3)	(4)	(5)
Whites	14,537	1,233.4	– 197.2	– 88.8	108.4
		(60.3)	(70.5)	(62.5)	(28.5)
Non-whites	11,664	2,449.1	839.7	1,074.4	234.7
		(47.4)	(62.7)	(50.7)	(32.5)

Notes: Figures are in nominal US dollars. The table shows estimates of the effect of voluntary military service on the 1988–91 Social Security-taxable earnings of men who applied to enter the armed forces during 1979–82. The matching and regression estimates control for applicants' year of birth, education at the time of application, and Armed Forces Qualification Test (AFQT) score. There are 128,968 whites and 175,262 non-whites in the sample. Standard errors are reported in parentheses.

Source: Adapted from Angrist (1998, Tables II and V).

Table 1 also shows regression estimates of the effect of voluntary service, with the same covariates used in the matching estimates controlled for. These are estimates of α_r in the equation

$$Y_i = \sum_X d_{iX}\beta_X + a_r D_i + e_i,$$

where β_X is a regression-effect for $X_i = X$ and α_r is the regression parameter. This corresponds to a saturated model for discrete X_i. The regression estimates are larger than (and significantly different from) the matching estimates. But the regression and matching estimates are not very different economically, both pointing to a small earnings loss for White veterans and a modest gain for non-whites.

Instrumental variables estimates of treatment effects

The conditional independence assumption required for regression or matching to identify a treatment effect is often implausible. Many of the necessary control variables are typically unmeasured or simply unknown. Instrumental variables (IV) methods solve the problem of

missing or unknown controls, much as a randomized trial also obviates the need for regression or matching. To see how this is possible, begin again with a constant effects model without covariates, so $Y_{1i} - Y_{0i} = \alpha$. Also, let $Y_{0i} = \beta + \varepsilon_i$, where $\beta \equiv E[Y_{0i}]$. The potential outcomes model can now be written

$$Y_i = \beta + aD_i + \varepsilon_i, \tag{5}$$

where α is the treatment effect of interest. Because D_i is likely to be correlated with ε_i, regression estimates of eq. (5) do not estimate α consistently.

Now suppose that in addition to Y_i and D_i there is a third variable, Z_i, that is correlated with D_i, but unrelated to Y_i for any other reason. In a constant-effects world, this is equivalent to saying Y_{0i} and Z_i are independent. It therefore follows that

$$E[\varepsilon_i \mid Z_i], \tag{6}$$

a conditional independence restriction on the relation between Z_i and Y_{0i}, instead of between D_i and Y_{0i} as required for regression or matching strategies. The variable Z_i is said to be an IV or just 'an instrument' for the causal effect of D_i on Y_i.

Suppose that Z_i is also a 0–1 variable. Taking expectations of (5) with Z_i switched off and on, we immediately obtain a simple formula for the treatment effect of interest:

$$\{E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0]\} / \{E[D_i \mid Z_i = 1] - E[D_i \mid Z_i = 0]\} = \alpha. \tag{7}$$

The sample analog of this equation is sometimes called the Wald estimator, since it first appear in a paper by Wald (1940) on errors-in-variables problems. There are other more complicated IV estimators involving continuous, multi-valued, or multiple instruments. For example, with a multi-valued instrument, we might use the sample analog of $\mathrm{Cov}(Z_i, Y_i) / \mathrm{Cov}(D_i, Y_i)$. This simplifies to the Wald estimator when Z_i is 0–1. The Wald estimator captures the main idea behind most IV estimation strategies since more complicated estimators can usually be written as a linear combination of Wald estimators (Angrist, 1991).

IV example

To see how IV works in practice, it helps to use an example, in this case the effect of Vietnam-era military service on the earnings of veterans later in life (Angrist, 1990). In the 1960s and

early 1970s, young men were at risk of being drafted for military service. Concerns about fairness also led to the institution of a draft lottery in 1970 that was used to determine priority for conscription in cohorts of 19-year-olds. A natural instrumental variable for the Vietnam veteran treatment effect is draft-eligibility status, since this was determined by a lottery over birthdays. In particular, in each year from 1970 to 1972, random sequence numbers (RSNs) were randomly assigned to each birth date in cohorts of 19-year-olds. Men with lottery numbers below an eligibility ceiling were eligible for the draft, while men with numbers above the ceiling could not be drafted. In practice, many draft-eligible men were still exempted from service for health or other reasons, while many men who were draft-exempt nevertheless volunteered for service. So veteran status was not completely determined by randomized draft eligibility; eligibility and veteran status are merely correlated.

Table 2   IV estimates of the effects of military service on US white men born 1950

Earnings year	Earnings		Veteran status		Wald estimate of veteran effect
	Mean	Eligibility effect	Mean	Eligibility effect	
	(1)	(2)	(3)	(4)	(5)
1981	16,461	− 435.8	0.267	0.159	− 2,741
		(210.5)		(.040)	(1,324)
1970	2,758	− 233.8			− 1,470
		(39.7)			(250)
1969	2,299	− 2.0			
		(34.5)			

Notes: Figures are in nominal US dollars. There are about 13,500 observations with earnings in each cohort. Standard errors are shown in parentheses.

Sources: Adapted from Angrist (1990, Tables 2 and 3), and unpublished author tabulations. Earnings data are from Social Security administrative records. Veteran status data are from the Survey of Program Participation.

For white men who were at risk of being drafted in the 1970–71 draft lotteries, draft-eligibility is clearly associated with lower earnings in years after the lottery. This can be seen in Table 2, which reports the effect of randomized draft-eligibility status on average Social Security-taxable earnings in column (3). Column (1) shows average annual earnings for purposes of comparison. For men born in 1950, there are significant negative effects of eligibility status on earnings in 1970, when these men were being drafted, and in 1981, ten years

later. In contrast, there is no evidence of an association between eligibility status and earnings in 1969, the year the lottery drawing for men born in 1950 was held but before anyone born in 1950 was actually drafted.

Because eligibility status was randomly assigned, the claim that the estimates in column (3) represent the effect of *draft eligibility* on earnings seems uncontroversial. The only information required to go from draft-eligibility effects to veteran-status effects is the denominator of the Wald estimator, which is the effect of draft-eligibility on the probability of serving in the military. This information is reported in column (4) of Table 2, which shows that draft-eligible men were 0.16 more likely to have served in the Vietnam era. For earnings in 1981, long after most Vietnam-era servicemen were discharged from the military, the Wald estimates of the effect of military service amount to about 15 percent of earnings. Effects were even larger in 1970, when affected soldiers were still in the army.

IV with heterogeneous treatment effects

The constant-effects assumption is clearly unrealistic. We'd like to allow for the fact that some men may have benefited from military service while others were undoubtedly hurt by it. In general, however, IV methods fail to capture either ATE or ATET in a model with heterogeneous treatment effects. Intuitively, this is because only a subset of the population is affected by any particular instrumental variable. In the draft lottery example, many men with high lottery numbers volunteered for service anyway (indeed, most Vietnam veterans were volunteers), while many draft-eligible men nevertheless avoided service. The draft lottery instrument is not informative about the effects of military service on men who were unaffected by their draft-eligibility status. On the other hand, there is a sub-population who served solely because they were draft-eligible, but would not have served otherwise. Angrist, Imbens and Rubin (1996) call the population of men whose treatment status can be manipulated by an instrumental variable the set of *compliers*. This term comes from an analogy to a medical trial with imperfect compliance. The set of compliers are those who 'take their medicine', that is, they serve in the military when draft-eligible but they do not serve otherwise.

Under reasonably general assumptions, IV methods can be relied on to capture the effect of

treatment on compliers. The average effect for this group is called a local average treatment effect (LATE), and was first discussed by Imbens and Angrist (1994). A formal description of LATE requires one more bit of notation. Define potential treatment assignments D_{0i} and D_{1i} to be individual i's treatment status when Z_i equals 0 or 1. One of D_{0i} or D_{1i} is counterfactual since observed treatment status is

$$D_i = D_{0i} + Z_i(D_{1i} - D_{0i}).$$

The key identifying assumptions in this setup are (*a*) conditional independence, that is, that the joint distribution of $\{Y_{1i}, Y_{0i}, D_{1i}, D_{0i}\}$ is independent of Z_i; and (*b*) monotonicity, which requires that either $D_{1i} \geq D_{0i}$ for all i or vice versa. Monotonicity requires that, while the instrument might have no effect on some individuals, all of those who are affected should be affected in the same way (for example, draft eligibility can only make military service more likely, not less). Assume without loss of generality that monotonicity holds with $D_{1i} \geq D_{0i}$. Given these two assumptions, the Wald estimator consistently estimates LATE, written formally as $E[Y_{1i} - Y_{0i}|D_{1i} > D_{0i}]$. In the draft lottery example, this is the effect of military service on those veterans who served because they were draft eligible but would not have served otherwise. In general, LATE compliers are a subset of the treated. An important special case where LATE = ATET is when D_{0i} equals zero for everyone. This happens in a social experiment with imperfect compliance in the treated group and no one treated in the control group.

IV Details

Typically, covariates play a role in IV models, either because the IV identification assumptions are more plausible conditional on covariates or because of statistical efficiency gains. Linear IV models with covariates can be estimated most easily by two-stage least squares (2SLS), which can also be used to estimate models with multi-valued, continuous, or multiple instruments. See Angrist and Imbens (1995) or Angrist and Krueger (2001) for details and additional references.

<div style="text-align: right">Joshua D. Angrist</div>

See also instrumental variables with weak instruments; matching estimators; regression-

discontinuity analysis; Rubin causal model; selection bias and self-selection; two-stage least squares and the k-class estimator

Bibliography

Angrist, J. 1990. Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. *American Economic Review* 80, 313–35.

Angrist, J. 1991. Grouped-data estimation and testing in simple labor-supply models. *Journal of Econometrics* 47, 243-266.

Angrist, J. 1998. Estimating the labor market impact of voluntary military service using Social Security data on military applicants. *Econometrica* 66, 249–88.

Angrist, J. and Imbens, G. 1995. Two-stage least squares estimates of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90, 431–42.

Angrist, J., Imbens, G. and Rubin, D. 1996. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–55.

Angrist, J. and Krueger, A. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives* 15(4), 69–85.

Ashenfelter, O. 1978. Estimating the effect of training programs on earnings. *Review of Economics and Statistics* 6, 47–57.

Ashenfelter, O. and Card, D. 1985. Using the longitudinal structure of earnings to estimate the effect of training programs. *Review of Economics and Statistics* 67, 648–60.

Dehejia, R. and Wahba, S. 1999. Causal effects in nonexperimental studies: reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94, 1053–62.

Heckman, J. 1976. The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* 5, 475–92.

Heckman, J. 1979. Sample selection bias as a specification error. *Econometrica* 47, 153–61.

Heckman, James J. and Robb, R. 1985. Alternative methods for evaluating the impact of interventions. In J. Heckman & B. Singer (Eds.), *Longitudinal Analysis of Labor Market Data* (pp. 156-245). New York: Cambridge University Press.

Holland, P. 1986. Statistics and causal inference. *Journal of the American Statistical Association* 81, 945–70.

Imbens, G. and Angrist, J. 1994. Identification and estimation of local average treatment effects. *Econometrica* 62, 467–75.

Quandt, R. 1958. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53, 873–80.

Rosenbaum, P. and Rubin, D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* 70, 41–55.

Rubin, D. 1974. Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology* 66, 688–701.

Rubin, D. 1977. Assignment to a treatment group on the basis of a covariate. *Journal of Educational Statistics* 2, 1–26.

Wald, A. 1940. The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11, 284–300.