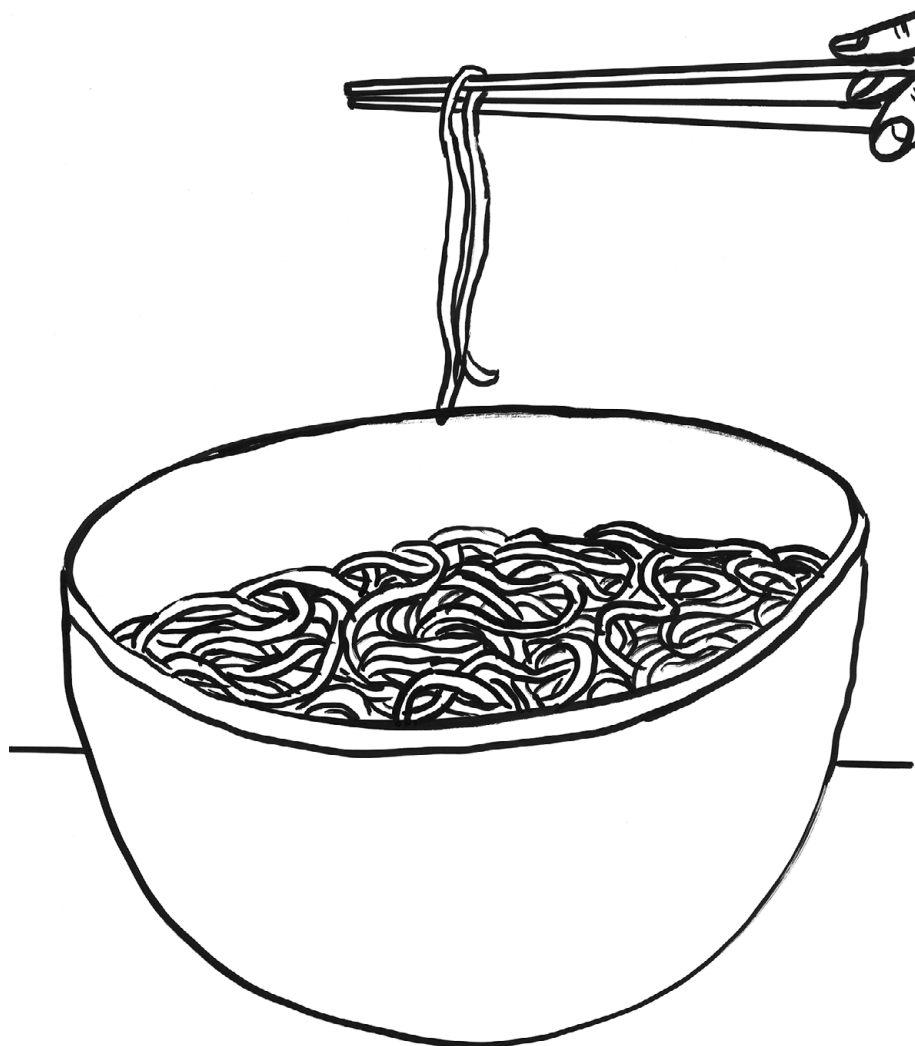LIFE IS LIKE A BOWL OF NOODLES

I WILL GIVE NO FURTHER EXPLANATION

7   Tools to Explain Wellbeing

There are three kinds of lies: lies, damned lies, and statistics.

Mark Twain

To explain the level and inequality of wellbeing, we use the standard tools of quantitative social science. These are mainly the techniques of multiple regression. In this chapter, we shall show how multiple regression can address the following issues.[1]

(1) What is the effect of different factors on the level of wellbeing (using survey data)?
(2) What problems arise in estimating this and how can they be handled?
(3) How far do different factors contribute to the observed inequality of wellbeing?
(4) How can experiments and quasi-experiments show us the effect of interventions to improve wellbeing?

So suppose that a person's wellbeing (W) is determined by a range of explanatory variables ($X_1, \ldots, X_N$) in an additive fashion. But in addition there is an unexplained residual (e), which is randomly distributed around an average value of zero. Then the wellbeing of the ith individual (W_i) is given by

$$W_i = a_0 + a_1 X_{i1} + \ldots + a_N X_{iN} + e_i,$$

which we can also write as

$$W_i = a_0 + \sum_{j=1}^{N} a_j X_{ij} + e_i. \tag{1}$$

In this equation, wellbeing is being explained by the X_js. So wellbeing is the 'dependent' variable (or left-hand variable) and the X_js are the 'independent' or (right-hand) variables. These right-hand variables can be of many forms. They can be continuous like income or the logarithm of income or like age or age squared. Or they can be binary variables like unemployment: you are either unemployed or not unemployed. These binary variables are often called dummy variables and they take

---

[1] The treatment is introductory and some readers will already know it all. If not, it will help you understand what you are doing when you use statistical software. For fuller expositions, see one of the excellent textbooks such as Angrist and Pischke (2008).

the value of 1 when you are in that state (e.g., unemployed) and the value of 0 when you are not in that state (e.g., not unemployed).

If we want to explain wellbeing, we have to discover the size of the effect of each thing that affects wellbeing. In other words, we have to discover the size of the a_js. For example, suppose

$$W_i = a_0 + a_1 \ \log \ \text{Income}_i + a_2 \ \text{Unemployed}_i + e_i. \tag{2}$$

From Chapter 8, you will find as benchmark numbers that $a_1 = 0.3$ and $a_2 = -0.7$. This means that when a person's log Income increases by one point, her wellbeing increases by 0.3 points (out of 10). Similarly, when a person ceases to be unemployed, her wellbeing increases by 0.7 points (ignoring any effect of a simultaneous change in income). And, if both things happen together, wellbeing increases by a whole point (0.3 + 0.7).

Estimating the Effect of a Variable

But how are we to estimate, as best we can, the true values of these a_j coefficients? The best unbiased way of doing this is to find the set of a_js that leaves the smallest sum of squared residuals e_i^2, across the whole sample of people being studied.[2] This is known as the method of Ordinary Least Squares (OLS). Standard programmes like STATA will do it for you automatically. However, there are 4 possible problems with such estimates when obtained from a cross-section of the population.

Omitted variables

Suppose that equation (2) is not the correct model but that another X variable should also be in the equation. Suppose, for example, that the right model is

$$W_i = a_0 + a_1 \ log \ \text{Income}_i + a_2 \ \text{Unemployed}_i + a_3 \ \text{Education}_i + e_i \tag{3}$$

where Education means years of education. Clearly education and income are positively correlated. So if a_1 and a_3 are positive, people with higher income will be getting higher wellbeing for 2 reasons:

the direct effect of income (a_1) and
the effect of education in so far as it is correlated with income.

Thus, equation (2) will give an exaggerated estimate of the direct effect (a_1) of income on wellbeing.[3] To leave out education is to leave out a confounding variable. And any such **confounding variable** must have two properties:

it is causally related to the dependent (LHS) variable and
it is correlated with an independent (RHS) variable.

---

[2] This is the best unbiased estimation system (with least standard errors on the estimated a_js), provided the errors are homoscedastic.
[3] The sign of the bias in a_1 equals the sign of a_3 times the sign of the correlation of X_1 and X_3.

If we lack data on the confounding variable, the classic way to overcome this problem is to use time-series panel data on the same people. Provided the omitted variable is constant over time, it can cause no problem, since we can now estimate how changes in income within the same person affect changes in her wellbeing. Thus, if we use time-series data, we cease to compare different individuals at the same point of time and we compare the same individual at different periods of time. Algebraically, we do this by expanding equation (2) to include multiple time periods (t) and adding a fixed effect dummy variable (f_i) for each individual. This picks up the effect of all the fixed characteristic of the individual (which for most adults will include education). Thus, we now explain the wellbeing of the ith person in the tth time period by

$$W_{it} = a_0 + a_1 \, log \, \text{Income}_{it} + a_2 \, log \, \text{Unemployed}_{it} + f_i + e_{it}. \tag{4}$$

There are standard programmes for including fixed-effects. A similar method to this is used for analysing the effect of experiments, but we shall come to this later.

Reverse causality

However, there is another problem. Suppose we are interested in the effect of income on wellbeing. But suppose that there is also the reverse effect – of wellbeing on income.[4] How can we be sure that, when we estimate equation (2), we are really estimating the effect of income on wellbeing rather than the reverse relationship or a mixture of the two? In other words is equation (2) in principle '**identifiable**'?

For an equation to be identifiable, it must **exclude** at least one of the variables that appears in the second relationship (the one that determines income).[5] But, even if it is identifiable, there is still the problem of getting a causal estimate of the effects of the endogenous variable.

The aim has to be to isolate that part of the endogenous variable that is due to something exogenous to the system. A variable that can isolate that part of the endogenous variable is called an **instrumental variable.** For example, if tax rates or minimum wages changed over time, these would be good instruments. Instrumental variables can also be used to handle the problem of omitted variables. In every case a good instrument

(i) is well related in a causal way to the variable it instruments and
(ii) should not itself appear in the equation, (i.e., it is not correlated with the error term in the equation).

There are programmes for the use of instrumental variables (IVs).

---

[4] For evidence on the reverse relationship (the effect of adolescent wellbeing on later earnings), see De Neve and Oswald (2012).

[5] In a three-equation system, it would need to exclude two variables from the rest of the system and so on.

Another way to isolate causal relationships is through the **timing** of effects. For example, income affects wellbeing in the next period rather than the current period. We can then identify its effect by regressing current wellbeing on income in the previous period. Similarly with unemployment. This gives us

$$W_{it} = a_0 + a_1 \, log \, \text{Income}_{i,t-1} + a_2 \text{Unemployed}_{i,t-1} + f_i + e_{it}. \qquad (5)$$

Measurement error

Another source of biased estimates is measurement error. If the **left-hand** variable has high measurement error, this will **not** bias the estimated coefficients a_j. But, if an **explanatory** variable X_j is measured with error, this will bias a_j towards zero. If the measurement error is known, this can be used to correct for the bias. But, if not, an instrumental variable can again come to the rescue, provided it is uncorrelated with the measurement error in the variable it is instrumenting.

Mediating variables

A final issue is this. A multiple regression equation such as (3) shows us the effect of each variable upon wellbeing **holding other things constant**. But suppose we are interested in the **total** effect of changing one variable upon wellbeing. For example, we might ask What is the total effect of unemployment upon wellbeing?

The total effect is clearly

- a_2, plus
- a_1 times the effect of unemployment upon log income.

That is one way you could estimate it. An alternative way is to take equation (2) and leave income out of the equation, so that the estimated coefficient on unemployment includes any effect that unemployment has on wellbeing via its effect on income.

In a case like this, income is a **mediating** variable. If we are only interested in the total effect of unemployment, we can simply leave the mediating variable out of the equation. Or we can estimate a system of structural equations consisting of (2) and the equation that determines income. This discussion brings out one crucial point in wellbeing research. We should always be very clear what question we are trying to answer. We should choose our equation or equations accordingly.

Standard errors and significance

All coefficients are estimated with a margin of uncertainty. Each estimated coefficient has a 'standard error' (se) around the estimated value. The true value will lie within 2 'standard errors' on either side of the estimated coefficients in 95% of samples. Thus the '95% confidence interval' for the a_j coefficient runs from $\widehat{a}_j - 2se_j$ to $\widehat{a}_j + 2se_j$, where $\widehat{a}_j$ means the estimated value of a_j. If this confidence interval does not include the value zero, the estimated coefficient is said to be 'significantly different from zero at the 95% level'.

For many psychologists, this issue of **significance** is considered crucial. It answers the question 'Does X affect W at all?' But for policy purposes the more important question is 'How much does X affect W?' So the coefficient itself is more interesting than its significance level. For any sample size, the estimated coefficient is the best available answer to the question of how much X changes W. And, if you increase the size of the sample, the expected value of the estimated a_j does not change but its standard error automatically falls (it is inversely proportional to the square root of the sample size). So in this book we focus more heavily on the size of coefficients than on their significance (though we sometimes show standard errors in brackets in the tables).

The question we have been asking thus far in this chapter is How does wellbeing change when an independent variable changes? In algebraic terms, we have been studying dW/dX_j? This is the type of number we need in order to evaluate a policy change. For example, suppose we increased the income of poor people by 20%, how much would their wellbeing change (on a scale of 0–10)? If $a_j = 0.3$, it would increase by 0.06 points (0.3×0.2). A quite different question is In which areas of life should we look hardest in the search for better policies?

The Explanatory Power of a Variable

If our main aim is to help the people with the lowest wellbeing (as we discussed in Chapter 2), then our focus should be on what explains the inequality of wellbeing. To see why, suppose first that wellbeing depends only on one variable X_1, with $W = \alpha_0 + \alpha_1 X_1$. Then the distribution of W depends only on the distribution of X_1. If W is unequal, it is because X_1 is unequal and α_1 is high. The higher the standard deviation (σ_1) of X_1 and the higher α_1, the greater the inequality of W. This is illustrated in Figure 7.1. For high variance of W, the numbers in misery correspond to the areas A and B. But for the low variance of W the numbers in misery correspond only to the area B.

A next natural step is to compare the standard deviation of $a_1\sigma_1$ with the standard deviation of wellbeing itself. Obviously, if they were equal in size, the spread of X_1 would be 'explaining' the whole spread of wellbeing σ_w – in other words, the two variables would be perfectly correlated. The correlation coefficient (r) between W and X_1 is therefore $a_1\sigma_1/\sigma_w$:

$$r = \frac{a_1\sigma_1}{\sigma_w} = \text{Correlation coefficient}$$

However, this can be either positive or negative depending on the sign of a_1. So a natural measure of the **explanatory power** of a right-hand variable is *the* squared value of r (which is also often written as R^2):

$$r^2 = \frac{a_1^2\sigma_1^2}{\sigma_w^2} = \text{Share of variance explained}$$

Since the denominator is the variance of wellbeing, this shows what proportion of the variance in wellbeing is explained by the variance of X_1.
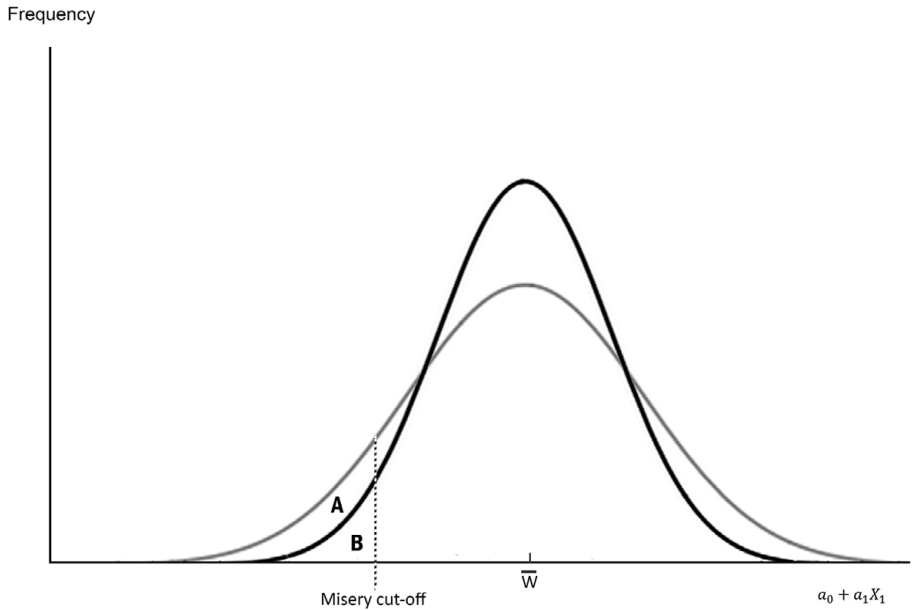
Frequency

Figure 7.1 How the numbers in misery are affected by $a_1\sigma_1$

In the real world, wellbeing depends on more than one variable (see equation [1]). The policy-maker may then ask Which of these variables is producing the largest amount of misery?[6] For this purpose, we need to compare the explanatory power of the different variable. This is done by computing for each variable its partial correlation coefficient with wellbeing. This partial correlation coefficient is normally described as β_j where

$$\beta_j = \frac{a_j\sigma_j}{\sigma_w} = \text{Partial correlation coefficient.}$$

This β-coefficient will appear frequently throughout this book.[7]

These β-coefficients are hugely interesting, as we shall see via two steps. First, starting from equation (1) we can readily derive the following equation.[8]

$$\frac{W_i - \bar{W}}{\sigma_w} = \sum \beta_j \frac{\left(X_{ij} - \bar{X}_j\right)}{\sigma_j} + \frac{e_i}{\sigma_w} \tag{6}$$

---

[6] As discussed in Chapter 18, the policy-maker can then develop policies in these areas and target them at those individuals least favoured in these variables. This is reasonably practical, while it is not really practical to identify the individuals with the lowest wellbeing and then target them with the best policies (though that would be the most logical approach).

[7] Sociologists often call it the 'path coefficient, p'.

[8] (i) Divide both sides by σ_w. (ii) Multiply and divide X_j by σ_j. (iii) Derive the average equation for the whole population and subtract it from the original equation (to eliminate a_0/σ_w).

Here we have **standardised** each variable by measuring it from its mean and dividing it by its standard deviation. These standardised equations appear many times in this book.[9]

But, to see the importance of these βs, we move on to a second equation, which is derived from (6).[10] This says

$$r^2 = \sum \beta_j^2 + \sum \sum \beta_g \beta_k r_{gk} \ (g \neq k); \tag{7}$$

r^2 is the proportion of the variance of W that is explained by the right-hand variables. And r_{gk} is the correlation coefficient between X_g and X_k.

Thus, the left-hand side is the share of the variance of wellbeing that is explained. The right-hand side consists of $\Sigma\beta_j^2$, which includes all the effects of the independent variation of the X_js, plus the effects of all their covariances. Thus β_j (or the partial correlation coefficient) measures the explanatory power of a variable (just as the correlation coefficient does in a simple bivariate relationship).

But some readers may wonder if this approach can handle independent variables that are binary. It can, because the standard deviation of a binary variable is simply $\sqrt{p\,(1-p)}$, where p is the proportion of people answering Yes to the binary question. For example, the standard deviation of Unemployed is $\sqrt{u(1-u)}$ where u is the unemployed rate. Thus, if X_j is Unemployed, its β coefficient is $a_j\sqrt{u(1-u)}/\sigma_w$.

Binary dependent variables

The matter is more complicated when it is the dependent variable that is binary. For example, suppose we divide the population into those who are in misery (with wellbeing below say 6) and the rest. How can we handle this? The most natural approach is, as normal, to regress the binary variable on all the other variables. This is what we often do in this book and, since it provides statistics of the standard kind, it is easy to understand.[11]

Box 7.1 Odds ratios

In analysing the effect of one binary variable on another binary variable, psychologists and sociologists often use the concept of an '**odds ratio**' rather than the values of a_j and β_j we have been discussing. Suppose, for example, we ask: How much more likely are unemployed people to be in misery, compared with people who are not unemployed? Imagine 100 people were distributed as follows (Table 7.1):

---

[9] The standardised value of a variable is often called its z-score. In other words, $\frac{X_{ij}-\bar{X}_j}{\sigma_j}$ is individual i's z-score for the variable X_j.

[10] To derive this, (i) square both sides of equation (6) and (ii) add up the equations, for all the individuals. (iii) Note that $r^2 = 1-\Sigma e_i^2/\sigma_w^2$.

[11] This linear probability model (LPM) has the problem that whereas the left-hand variable is either 1 or 0, the regression equation predicts all kinds of values for different individuals including some which are greater than 1 or less than 0. Thus, there is an alternative approach to binary dependent variables which assumes that a person has a given probability of being 1 or 0 as a function of the X_js and then chooses that function which makes what actually happened to appear as likely as possible. Depending on the functional form, this type of analysis is called either **Logit or Probit** analysis which is again available in STATA.

Box 7.1 (*cont.*)

Table 7.1 Distribution of 100 people by unemployment status and misery status

	In misery	**Not in misery**	**Total**
Unemployed	2	8	10
Not unemployed	9	81	90
Total	11	89	100

In this situation, the chance of an unemployed person being in misery is much higher than the chance of a non-unemployed person being in misery. The odds-ratio is

$$\frac{2}{8} / \frac{9}{81} \simeq 2.25.$$

But odds ratios do not answer either of the main questions we are addressing in this chapter. First, if we are interested in the effect on wellbeing of reducing unemployment, the proper measure of this effect is not the odds ratio but the **absolute difference** in the probabilities of misery between unemployed and non-unemployed people, that is, $0.2 - 0.1 = 0.1$. Second, if we are interested in the power of unemployment to **explain the prevalence** of misery, the correct statistic is the correlation coefficient between the two. So we shall not be showing odds ratios in this book, though the reader is able to compute them, given the necessary information.

Effect size of a binary independent variable

We have so far considered two ways in which to report regression results. One is to report the absolute effect of say unemployment on wellbeing in units of wellbeing. The other is to look at the relationship when both variables are standardised. However, there is the third approach that is often useful. This is to measure only the dependent variable in a standardised fashion. For example, we might ask 'When a person becomes unemployed, by how many standard deviations does his wellbeing go down?' This is a measure known as the **effect size** of the independent variable (sometimes knows as Cohen's *d*):

$$\text{Effect size} = \frac{\text{Absolute effect}}{\text{SD of depdendent variable}} = \text{Cohen's } d.$$

This is particularly useful when reporting the effect of an experiment.[12]

---

[12] Much less useful is the odds ratio (see Box 11.1). It is also useful to know that, if someone started at the median of a normal distribution and experienced a treatment with a given effect size, the resulting rise in the person's position in the distribution would be given approximately by

    Change in percentile points $= 40 \times$ Effect size

unless the effect size is very large.

Experiments

So far, we have been discussing the use of **naturalistic** data – mainly obtained by surveys of the population. As we have mentioned, it is often difficult to establish the causal effect of one variable on another from this type of data. The simplest way to establish a causal relationship is through a properly controlled experiment. Moreover, if you want to examine the effect of a policy that has never been tried before, it is the best way to get convincing evidence of its effects.

So how do we estimate the effects of being 'treated' in an experiment? Let's begin with a simple example. Suppose we want to try introducing a wellbeing curriculum into a school. Our aim is to see whether it makes any difference to those who receive it. So we would select two groups of pupils who were as similar to each other as possible. Then we would give the wellbeing curriculum to the treatment group (T) but not the control group (c). We would also measure the wellbeing of both groups before and after the treatment. So we would have the following values of wellbeing for each of four situations (Table 7.2).

To find the average effect of the treatment, we would compare the change in wellbeing experienced by the treatment group (T) with that experienced by the control group (C). Thus, the 'average treatment effect on the treated' (ATT) would be estimated as

$$ATT = (W_{T1} - W_{T0}) - (W_{C1} - W_{C0}). \tag{8}$$

In other words, the ATT is the '**difference in differences**', or for short the 'diff in diff'.

There may of course be many ways in which both groups changed between periods 0 and 1 – they will become older, they may experience a flu epidemic or whatever. But those changes should be similar for both groups. Thus the only observable thing that can produce a different change in wellbeing is the fact that Group T took the course and Group C did not.

Of course, there may also be some unobservable difference in experience, which means that the ATT is always estimated with a standard error. So, to put things into a more general form, let's imagine we have observations over a number of years. We then estimate

$$W_{it} = a_0 + a_1 T_{it} + v_t + f_i + e_{it}. \tag{9}$$

Here T_{it} is a variable which takes the value 1 in all periods after someone has taken the course, v_t is a year dummy, f_i is a person fixed effect and e_{it} is random noise.

Table 7.2 Average wellbeing of each group before and after the experiment

	Before	**After**
Treatment group (T)	W_{T0}	W_{T1}
Control group (C)	W_{C0}	W_{C1}

So far, we have assumed that in our experiment we can easily arrange for the treatment group and the control group to be reasonably similar. This is never in fact completely possible. But the method that gets us closest to it is '**random assignment**'.[13] In this case, we select an overall group for the experiment and then randomly assign people to either Group T or Group C (e.g., by tossing a coin for each individual). In this way, the groups are more likely to be similar than in any other way. Of course we can then check whether they differ in observable characteristics (X) and we can then allow in our equation for the possibility that these variables affect the measured *ATT*. Our equation then becomes

$$W_{it} = a_0 + a_1 T_{it} + a_2 T_{it} X_{it} + a_3 X_{it} + v_t + f_i + e_{it}. \tag{10}$$

Estimating equations like this are quite common.

However, randomisation between individuals is often not practicable. For example, suppose you wanted to test whether higher income transfers raised wellbeing enough to justify the cost. You could not randomly allocate money within a given population – it would be considered unfair since the transfer clearly benefits the recipient. You might, however, choose to transfer money to all eligible people in some **areas** and not in others, with the allocation between areas being random. This might not be considered unfair. Similarly, suppose you wanted to test the effects of improved teaching of life-skills in schools. Within a school it might be organisationally impossible to give improved teaching to some children and not others – or even to some classes. But you could use random assignment across schools. Or you could even argue that it is 'quasi-random' whether a child is born in Year t or Year $t − 1$; in this case, you could use children born in year t as a control group in the trial of a treatment applied to those born in year $t + 1$ (see Chapter 9). So all experiments should, if at all possible, use randomisation to reduce the unobservable differences between the treatment and control groups.

Selection bias

But suppose an innovation is made without an experiment and we then want to know its effects. For example, an exercise programme has been established, which some people have decided to adopt. Has it done them any good?

The only information that we have is for the period after the innovation. But we do also have information on people who did not opt in to the programme. So, can we answer our question by comparing the wellbeing of those who took the programme with those who didn't? Probably not, because the people who opted into the programme may have differed from those who didn't: they may well have started with higher wellbeing in the first place. So, if we just compared their final wellbeing with those of non-participants, the difference could be largely due to 'selection bias'.

[13] For the limits of this method, see Deaton and Cartwright (2018).

One method to deal with this is called Propensity Score Matching. In it we first take the whole sample of participants and non-participants and do a logit (or probit) analysis to identify that equation that best predicts whether they participate or not. From this analysis, we can say for every participant what was the probability they participated. We then find, for each participant, a non-participant with the same (or nearly the same) probability of participating. It is those non-participants who become the control group and we now compare their wellbeing with that of the treatment group. This gives us our estimate of the average treatment effect on the treated:

$$ATT = \tilde{W} \text{ of treated} - \bar{W} \text{ of matched sample.} \qquad (11)$$

Summary

(1) If $W = a_0 + \sum a_j X_j + e$, then the best unbiased way to estimate the values of the a_js is by Ordinary Last Squares (choosing the a_js to minimise the sum of squared residuals e^2).

(2) Omitted variables are confounders that can lead to biased estimates of the effect of the variables which are included.

(3) Time series estimation can eliminate any problem caused by omitted variables which are constant over time. Time series can also help to identify a causal effect if this takes place with a lag, so that for example X_{t-1} is affecting W_t.

(4) If a right-hand variable is endogenous, it should if possible be instrumented by an instrumental variable that is independent of the error in the equation. Instrumental variables can also help with omitted variables and measurement error.

(5) If an explanatory variable is measured with error, its estimated coefficient will be biased towards zero. This problem can again be solved by using an instrumental variable uncorrelated with the original measurement error.

(6) All regression estimates are estimated with 'standard errors' (se). The 95% confidence interval is the coefficient $\pm$ 2 se. Provided this interval does not include zero, the coefficient is 'significantly different from zero at the 95% level'. But the coefficient estimate is more interesting that its significance.

(7) To find the explanatory power of the different variables, we run the equation using standardised variables, that is, the original variables minus their mean and divided by their standard deviation. The resulting coefficients (β_j) – or partial correlation coefficients – reflect the explanatory power of the independent variation of each variable X_j. They are equal to $a_j \sigma_j / \sigma_w$ where w is the dependent variable.

(8) The surest way to determine a causal effect is by experiment. The best form of experiment is by random assignment. We then measure the wellbeing of the treatment and the control group before and after the experiment. This difference-in-difference measures the average treatment effect on the treated.

(9) Where random assignment is impossible, naturalistic data can be used and the outcome for the treatment group compared with a similar untreated group chosen by Propensity Score Matching.

(10) If the measured effect of a treatment is a (in units of the outcome variable W), the 'effect size' is a/σ_w.

We can now put these tools to work.

Further Reading

Angrist, J. D., and Pischke, J. S. (2008). *Mostly Harmless Econometrics*. Princeton University Press.