Public Finance
Public Finance Quarterly

In addition, all the major general-interest economics journals frequently publish articles that deal with public finance issues. These include, but are not limited to:

American Economic Review
Journal of Economic Perspectives
Journal of Political Economy
Quarterly Journal of Economics
Review of Economics and Statistics

Articles on public finance in these and many other journals are indexed in the *Journal of Economic Literature* and can be searched on the Internet.

In addition, students should consult the volumes included in the Brookings Institution's series *Studies of Government Finance*. These books include careful and up-to-date discussions of important public finance issues. The Congressional Budget Office also provides useful reports on current policy controversies. A list of documents is provided at its Web site, http://www.cbo.gov.

The working paper series of the National Bureau of Economic Research, available through university libraries, is another good source of recent research on public finance. The technical difficulty of these papers is sometimes considerable, however. Papers can be downloaded at its Web site, http://www.nber.org.

Vast amounts of data are available on government spending and taxing activities. The following useful sources of information are published by the US Government Printing Office and are available online as indicated:

Statistical Abstract of the United States (http://www.census.gov/compendia/statab/)
Economic Report of the President (http://www.gpoaccess.gov/eop/index.html/)
Budget of the United States (http://www.gpoaccess.gov/usbudget)
US Census of Governments (http://www.census.gov/govs/www/)

All the preceding are published annually, except for the *US Census of Governments*, which appears every five years. *Facts and Figures on Government Finance*, published annually by the Tax Foundation, is another compendium of data on government taxing and spending activities. For those who desire a long-run perspective, data going back to the 18th century are available in *Historical Statistics of the United States from Colonial Times to 1970* [US Government Printing Office]. Readers with a special interest in state and local public finance will want to read the reports issued by the US Advisory Commission on Intergovernmental Relations.

A great deal of public finance data is available on the Internet. A particularly useful site is *Resources for Economists on the Internet* (http://www.rfe.org). It lists and describes more than 900 Internet resources. The home page of the US Census Bureau (http://www.census.gov) is also very useful. Finally, for up-to-date information on tax policy issues, consult the Web site of the University of Michigan's Office of Tax Policy Research (http://www.otpr.org) and the Urban-Brookings Tax Policy Center (http://www.taxpolicycenter.org/).

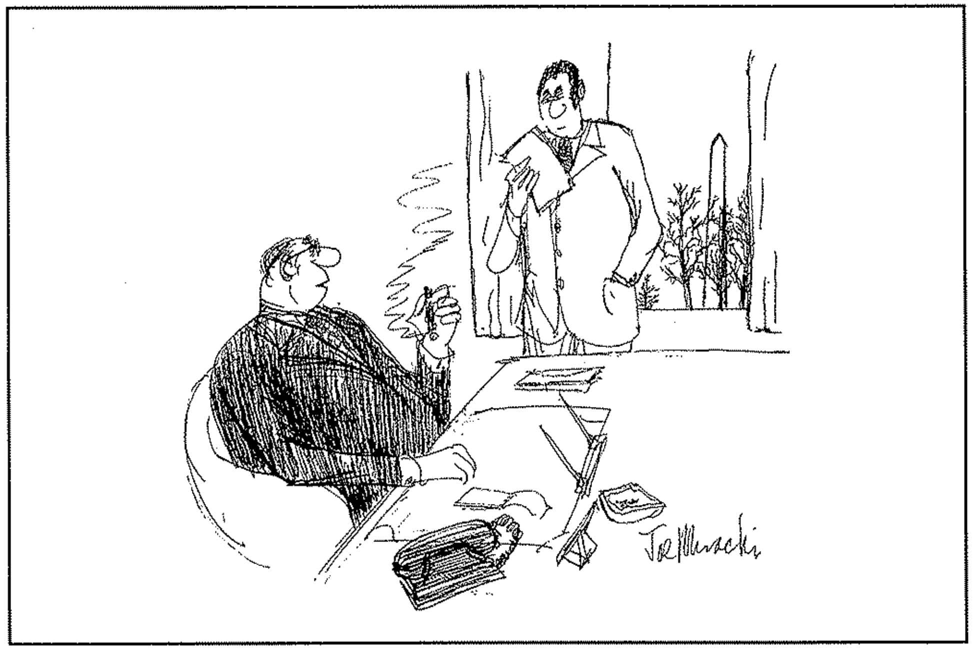
Chapter Two

Tools of Positive Analysis

Numbers live. Numbers take on vitality.

—JESSE JACKSON

A good subtitle for this chapter is "Why Is It So Hard to Tell What's Going On?" We constantly hear economists—and politicians—disagree vehemently about the likely consequences of various government actions. For example, when George W. Bush proposed reducing tax rates in 2006, many conservatives argued that lower tax rates create incentives for people to work harder. Many liberals were skeptical, arguing that taxes have little effect on work effort. Each side had economists testifying that their opinion was correct. Is the cynicism expressed in the cartoon below really surprising?



"That's the gist of what I want to say. Now get me some statistics to base it on."

This kind of discussion occurs virtually whenever economists and policymakers consider the impact of a government program. Economists debate whether environmental regulations improve health outcomes, whether government-provided health insurance decreases mortality, whether school vouchers improve test scores, whether tax reductions for corporations generate more investment, whether unemployment insurance leads to longer unemployment spells, and a slew of other important issues. This chapter discusses the tools that economists use to estimate the impact of government programs on individuals' behavior.

▸ THE ROLE OF THEORY

Economic theory is a useful starting point for analyzing the impact of government policy because it provides a framework for thinking about the factors that might influence the behavior of interest. Consider again the tax cuts proposed by President Bush, and suppose we are interested in their effect on annual hours of work. The theory of labor supply posits that the work decision is based on the rational allocation of time.[1] Suppose Mr. Rogers has only a certain number of hours in the day: How many hours should he devote to work in the market, and how many hours to leisure? Rogers derives satisfaction ("utility") from leisure, but to earn income he must work and thereby surrender leisure time. Rogers's problem is to find the combination of income and leisure that maximizes his utility.

Suppose Rogers's wage rate is $10 per hour. The wage is the cost of Rogers's time. For every hour he spends at leisure, Rogers gives up $10 in wages—time is literally money. However, a "rational" individual generally does not work every possible hour, even though leisure is costly. People spend time on leisure to the extent that leisure's benefits exceed its costs.

This model may seem unrealistic. It ignores the possibility that an individual's labor supply behavior can depend on the work decisions of other family members. Neither does the model consider whether the individual can work as many hours as desired. Indeed, the entire notion that people make their decisions by rationally considering costs and benefits may appear unrealistic.

However, the whole point of model building is to simplify as much as possible, so one can reduce a problem to its essentials. A model should not be judged on the basis of whether or not it is 100 percent true, but on whether it is plausible, informative, and offers testable implications. Most work in modern economics is based on the assumption that utility maximization is a good working hypothesis. This point of view is taken throughout the book.

Imagine that Mr. Rogers has found his utility-maximizing combination of income and leisure based on his wage rate of $10. Now the government imposes a tax on earnings of 20 percent. Then Rogers's after-tax or *net* wage is $8. How does a rational individual react—work more, work less, or not change? In public debate, arguments for all three possibilities are made with great assurance. In fact, however, the impact of an earnings tax on hours of work *cannot* be predicted on theoretical grounds.

To see why, first observe that the wage tax lowers the effective price of leisure. Before the tax, consumption of an hour of leisure cost Rogers $10. Under the earnings tax, Rogers's net wage is lower, and an hour of leisure costs him only $8. Since

[1] A graphical exposition of the theory of labor supply appears in Chapter 18 under "Labor Supply."

leisure has become cheaper, he will tend to consume more of it—to work less. This is called the **substitution effect.**

substitution effect
The tendency of an individual to consume more of one good and less of another because of a decrease in the price of the former relative to the latter.

Another effect occurs simultaneously when the tax is imposed. If Rogers works the same number of hours after the tax, he receives only $8 for each of these hours, while before it was $10. In a real sense, Rogers has suffered a loss of income. To the extent that leisure is a **normal good**—consumption increases when income increases, and consumption decreases when income decreases—this income loss leads to less consumption of leisure. But less leisure means more work. Because the earnings tax makes Rogers poorer, it induces him to work more. This is called the **income effect.**

normal good
A good for which demand increases as income increases and demand decreases as income decreases, other things being the same.

income effect
The effect of a price change on the quantity demanded due exclusively to the fact that the consumer's income has changed.

Thus, the tax simultaneously produces two effects: It induces substitution toward the cheaper activity (leisure), and it reduces real income. Since the substitution and income effects work in opposite directions, the impact of an earnings tax cannot be determined by theorizing alone.

The importance of the ambiguity caused by the conflict of income and substitution effects cannot be overemphasized. The theoretical model helps understand the relationship between income taxes and labor supply, but only empirical work—analysis based on observation and experience as opposed to theory—can tell us how labor force behavior is affected by changes in the tax system. Even intense armchair speculation on this matter must be regarded with considerable skepticism. Here, then, we see one major role for economic theory: to make us aware of the areas of our ignorance.

In other contexts, economic theory can be the reason for thinking that a research question is important in the first place. Consider a government policy of mandating safety-design features (such as seat belts, air bags, and antilock brakes) in automobiles. The goal of such measures is to improve public safety. Yet, as pointed out by Peltzman [1975], economic theory suggests that this measure might actually backfire and increase fatalities. The basic logic is simple—economic theory says that, in general, when the cost of some activity goes down, people are more likely to engage in that activity. In this case, the safety-design features reduce the "cost" of driving fast and recklessly, because in the event of an accident, the injuries may be less severe. By this logic, then, mandating safety features could lead to more reckless driving and more associated accidents.

Empirical work is required to determine whether the reduction in fatalities from the additional safety-design features more than offsets the increase in fatalities due to more reckless driving. An additional testable proposition stemming from theory is that we would expect a disproportionate increase in pedestrian fatalities stemming from mandated safety-design features because pedestrians are exposed to the increase in reckless driving but do not experience the countervailing protection of the safety devices. Here we see another important function of economic theory: to generate hypotheses whose validity can be assessed through empirical work.

▸ CAUSATION VERSUS CORRELATION

The examples we have cited so far point to the importance of establishing a causal relationship between a certain government policy and an outcome of interest. In order for us to infer that government action X causes societal effect Y, three conditions must hold:

1. The cause (X) must precede the effect (Y). This makes sense, because a causal relationship is only possible if the cause leads to (that is, precedes) the effect.
2. The cause and effect must be **correlated.** Two events are correlated if they move together. The correlation may be positive (X and Y move in the same direction) or negative (X and Y move in opposite directions). If Y does not change when X does, then X cannot be causing Y.
3. Other explanations for any observed correlation must be eliminated.

correlation
A measure of the extent to which two events move together.

The last condition is tricky. It requires that other influences of Y (which we call factor Z) get ruled out before attributing X as the cause. Suppose, for example, that we want to know whether participating in a government job-training program increases an individual's wages. Suppose we collect earnings data from a group of individuals, some of whom enrolled in a job-training program and some of whom did not. We refer to the workers who went through the program as the **treatment group,** because they received the "treatment" that we are evaluating. The workers who did not receive the treatment are referred to as the **control group.**

treatment group
The group of individuals who are subject to the intervention (e.g., government program) being studied.

control group
The comparison group of individuals who are not subject to the intervention (e.g., government program) being studied.

Suppose we find that the treatment group of workers had higher wages than the control group. This suggests that the first two criteria for causation are met, but in order to infer that the job-training program caused the higher wages, we must consider whether other explanations exist for the observed relationship between the two events. One possible explanation is that the workers in the treatment group were more highly motivated than the workers in the control group. Higher motivation might induce workers both to enroll in the job-training program *and* to work harder once they get a job. The point is that these treatment workers might have obtained higher wages even in the absence of the job-training program. This suggests that factor Z (higher motivation) leads both to enrollment in the program and to higher wages, which means that one cannot conclude that the program caused the higher wages. Put another way, the fact that there is a correlation does not prove causation.

The importance of distinguishing between correlation and causation comes up in a variety of contexts. For example, there is a positive correlation between whether a man is married and his wages. On this basis, some pundits and policymakers have suggested that the government should institute financial incentives for people to marry. The problem is that there are other possible explanations for the correlation between men's marital status and their wages. It could be that men with better personalities do better in the job market and are more likely to find a spouse. One must rule out other explanations before promoting a policy that encourages marriage as a means of boosting wages.

▸ EXPERIMENTAL STUDIES

In our hypothetical example we saw that the observed relationship between enrolling in the job-training program and wages was due to a third influence—motivation level. The problem is that the characteristics of the control group workers differed from the characteristics of the treatment group workers. As a result, the estimated higher wages for the treatment group relative to the control represented a **biased estimate** of the true causal impact of the job-training program. A biased estimate is one that conflates the true causal impact with the impact of outside factors. In order to be compelling, empirical economics should eliminate bias when estimating the causal relationship between two events.

biased estimate
An estimate that conflates the true causal impact with the impact of outside factors.

In order to rule out other factors, we would like to know the **counterfactual,** which is what would have happened to the treatment group had they not received the treatment. Of course, in our job-training example it is impossible to know the true counterfactual because the treatment workers did indeed enroll in the job-training program. In order to make things interesting, let's momentarily leave the real world for the world of science fiction in which time travel is possible. First, we form a treatment group of workers who we enroll in the job-training program and we measure their subsequent wages. Then we go back in time and put the same people in a control group that does not enroll in the program, and we measure their subsequent wages. In this scenario, our treatment group consists of the exact same people as our control group. The only difference is that the former enrolled in the job-training program and the latter (in an alternative timeline) did not. In other words, our control group *is* the counterfactual. Any increase in the treatment group's wages relative to the control group's wages can therefore reliably be attributed to the causal effect of enrolling in the job-training program.

counterfactual
The outcome of people in the treatment group had they not been treated.

In a world without time travel it is impossible to use the same people for both the control group and the treatment group. Luckily, there is a good alternative, which is to use an **experimental (or randomized) study,** in which subjects are *randomly* assigned to either the treatment group or the control group. With random assignment, the people in the control group are not literally the same people as those in the treatment group, but they have similar characteristics on average. Importantly, because selection into the treatment group is outside the individual's control, it is less likely that other factors can lead the investigator to confuse correlation for causation.

experimental study
An empirical study in which individuals are randomly assigned to the treatment and control groups.

Experimental studies are considered the gold standard of empirical studies because of this potential to eliminate bias. They are frequently used in the natural sciences such as medicine. For example, in order to test the effectiveness of a drug, researchers can randomly assign people to either a treatment group (in which case they receive the drug treatment) or to a control group (in which case they receive a placebo instead of the drug). Any observed differences in their medical outcomes can therefore be attributed to the drug rather than differences in other characteristics. It was on this basis that, several years ago, scientists determined that the antibiotic drug streptomycin was an effective treatment for tuberculosis.

Conducting an Experimental Study

In an experimental study of the effect of the job-training program on wages, the first step is to randomly assign a sample of workers to either enroll or not enroll in the program. If we start with a small sample of workers, then it is still possible that there will be large differences in the average characteristics of those in the control and treatment groups. But as our sample size increases, we can expect the characteristics of both groups to be the same on average.

We can test this assumption by collecting data on the characteristics of the workers before they are assigned to the treatment and control groups and then comparing the average values after assignment to make sure there are no large differences across the two groups. With random assignment, not only do we expect the two groups' observed characteristics (such as education) to be the same on average, we also expect their unobserved characteristics (such as motivation level) to be the same on average as well. The final step is to compare the average wages across the two groups after the members of the treatment group have gone through the program. Because the two groups have the same characteristics at the start of the study, any difference in wages between the two groups can be attributed to enrollment in the job-training program.

Pitfalls of Experimental Studies

It is hard for economists to conduct controlled experimental studies. Sometimes the difficulty is due to ethical issues. Suppose, for example, that policymakers want to know how many fewer illnesses would result from a given reduction in pollution. An experimental study would randomly assign people to different groups, some of which would be exposed to low levels of pollution and others to very high levels. While this would yield unbiased estimates of the effect of pollution reduction on health, most people would agree that experimenting on people in this way is unethical.

Technical problems arise as well. Consider our hypothetical experiment in which workers are randomly selected to enroll or not enroll in the job-training program. Randomization ensures that the treatment and control groups have similar characteristics on average. But what if some of the treatment group workers who were enrolled in the job-training program do not actually attend the program? When we later compare wages between the treatment and control groups, we might draw misleading inferences if we don't know which workers in the treatment group failed to get treated. In the same way, members of the control group may find ways to get into the treatment group or obtain an experience similar to the treatment group, such as enrolling in substitute programs. In short, people in an experiment are not passive objects, and their behavior may undo the effects of randomization.

Another problem can arise when some workers involved in the experiment fail to respond to follow-up surveys requesting their wage information. For example, suppose that the job-training program actually does increase wages. However, suppose also that low-wage workers are less likely to report their future wages to the researcher. In this case, the average post-treatment wages of the control group are artificially high, because the low-wage people are not included in the computation of the average. We might then erroneously conclude that the treatment and control groups have the same wages. The basic problem is that even though the experiment started with random samples, when the final data are collected, the control group has been contaminated by the nonrandom disappearance of certain of its members.

A final problem is that people in an experiment may not behave the same way as they would if the entire society were subjected to the policy, especially if the experiment has limited duration. For example, suppose we conducted an experiment to estimate how much more frequently people go to the doctor when they have generous health insurance. We can randomly select some people to receive generous health insurance for a year and others to receive less generous health insurance for the same year. The problem is that the treatment group subjects might go to the doctor very frequently because they know the experiment will only last one year, after which health care will become much more expensive for them. The measured effect of the treatment will be a biased estimate of the impact of a government policy that provides generous health insurance indefinitely.

This leads to a more general concern with experiments. They are adept at achieving unbiased estimates of a causal relationship in a particular context. However, it is not clear whether the causal inferences from one context can be generalized to other populations, settings, and even to related treatments. For example, in the mid-1980s the state of Illinois conducted a controlled experiment in which a random sample of unemployed people was told they would receive a $500 bonus if they found a job within 11 weeks. (See Woodbury and Spiegelman [1987].) The findings suggested that the bonus reduced the time a person remained unemployed. Given the careful design of the study, this finding is likely unbiased. But what could a public official in California in 2007 learn from this experiment, which relied on a different population and took place in a different time period? In short, to what extent do the experimental results generalize? In the same way, suppose that one were back in Illinois in the 1980s, and the government could afford a bonus of only $250 rather than $500. How would that policy affect time spent unemployed? Presumably a bonus of $250 would have less of an impact than $500, but how much less is entirely unclear. The experimental results by themselves do not provide much guidance.

This example illustrates the "black box" aspect of experiments. Experiments provide reliable estimates of the impact of a very specific policy on behavior, but they do not provide in-depth understanding (i.e, what goes on inside the black box) of why any changes have occurred. Consequently, we don't learn much about the impacts we can expect if the policy is applied in other contexts or if it is configured in a somewhat different fashion. This returns us to our discussion of the role of theory. By making assumptions on how people behave, in particular that they rationally maximize utility, theory can help us generalize particular experimental results to other populations or policies.

Thus, although experimental studies offer a credible way to evaluate the impact of a policy, they are not foolproof. In particular, researchers must carefully track the subjects in the control and treatment groups to make sure the original random assignment remains over time, and they must be careful about generalizing the results to other settings or policies.

▸ OBSERVATIONAL STUDIES

Experimental studies are simply out of the question for many important issues. For example, as mentioned earlier, knowing the impact of tax reductions on labor supply is of major interest. An experimental study of this issue would require randomly giving some people tax cuts and others not. Even if this were legally and politically possible, we would still face the problem that the people in the tax cut group would know that they were part of an experiment, and this could affect their behavior. Under these circumstances, instead of experiments, economists rely on **observational studies,** which use data obtained by observing and measuring actual behavior outside of an experimental setting.

observational study
An empirical study that relies on observed data that are not obtained from an experimental setting.

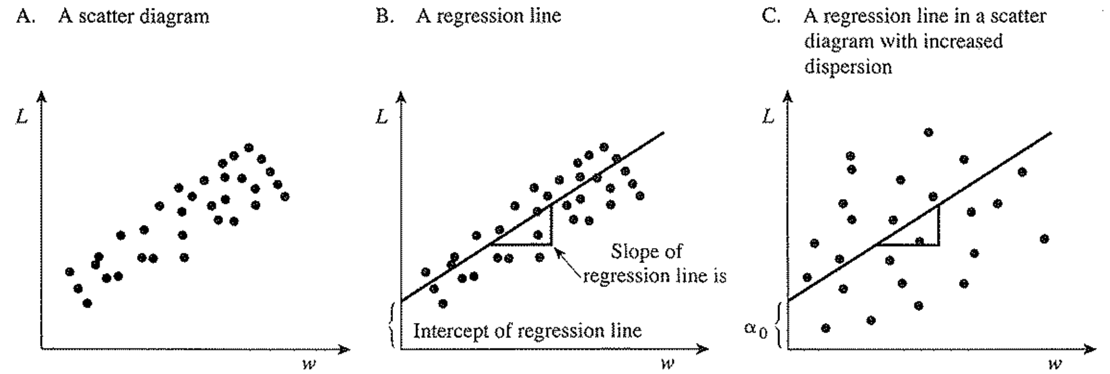
Observational data come from a variety of sources. Some are collected by surveying people, such as telephone surveys of consumers or written surveys submitted by households every 10 years for the census. Other observational data come from administrative records, including historical records of births and deaths, or government data on national economic performance.

Without randomization, observational studies must rely on other techniques to rule out factors that might contaminate causal inferences. **Econometrics** is the use of statistical analysis of economic data in order to estimate causal relationships. Specifically, econometrics uses *regression analysis* to estimate the relationship between two variables while holding other factors constant. We next explain how this technique works.

econometrics
The statistical tools for analyzing economic data.

Figure 2.1 Multiple regression analysis

Panel A shows there is a positive correlation between hours of work and after-tax wages. Panel B shows the regression line that fits through these data points, which yields an estimate of the magnitude of the relationship between the two variables. The estimated relationship between the two variables is more reliable in Panel B than in Panel C, because the data points in Panel C are more scattered.



Conducting an Observational Study

Suppose we are interested in estimating the effect of a reduction of the income tax on annual hours of work (which we denote as L, for labor supply). A change in the income tax changes the net wage rate (w) that a person receives. So we can state the problem as: Is a change in the tax rate followed by a change in hours worked, is there an observed correlation between these two events, and can we rule out other Z-factors that can explain this observed correlation? In observational studies, variables that are thought to be causal (such as the net wage here) are referred to as *independent variables*. A variable that is thought to be an outcome (such as labor supply here) is referred to as the *dependent variable*.

To see how observational studies work, suppose we have information on the hours of work and on the after-tax wages for a sample of people for a given year. We can plot these data points, as shown in Figure 2.1A. This figure indicates a positive correlation between after-tax wages and labor supply—labor supply increases as after-tax wages increase. We are interested in estimating the magnitude of this relationship. This is the task of regression analysis, which fits a **regression line** through the observed data points. Obviously, no single straight line can cross through all these points. The purpose of regression analysis is to find the line that best fits this relationship, as shown in Figure 2.1B.[2] The slope of this line, known as the regression coefficient, is an estimate of the relationship between after-tax wages and labor supply. Suppose, for example, that the regression coefficient is 1.5. This suggests that if the net wage goes up by \$10, then labor supply will increase by 15 hours per year.

regression line

The line that provides the best fit through a scatter of data points.

[2] In this example, we are assuming a linear relationship between the two variables. The "best" line minimizes the sum of the squared vertical distances between the points on the line and the points in the scatter. (See Wooldridge [2006].) Econometrics also allows for a nonlinear relationship between variables, which is frequently a preferable approach.

The reliability of the estimated regression coefficient depends on the distribution of the data in the scatter plot. To see why, suppose our scatter of points looked like those in Figure 2.1C. The regression line is identical to that in Figure 2.1B, but the scatter of points is more diffuse. Even though the estimated regression coefficient is the same, one has less faith in its reliability. Econometricians calculate a measure called the **standard error,** which indicates the reliability of the estimated coefficient. When the standard error is small in relation to the size of the estimated parameter, the coefficient is said to be statistically significant.

standard error

A statistical measure of how much an estimated regression coefficient might vary from its true value.

This example assumed there is only one explanatory variable, the net wage. Suppose that instead there were two independent variables: the net wage and nonlabor income (such as dividends and interest). Multiple regression analysis estimates the relationship between each independent variable and the dependent variable, holding all other independent variables constant. Multiple regression analysis is a very valuable tool, because in virtually all interesting problems, more than one independent variable can causally affect the dependent variable. If we can control for all factors that explain the dependent variable, this technique allows us to find the independent causal effect of whatever variable is under consideration.

Types of Data Regression analysis can be conducted using different types of data. Figure 2.1 relied on data on after-tax wages and labor supply for a sample of people in a given year. Data that contain information on individual entities (for example, workers, consumers, firms, states, or countries) at a given point in time are known as **cross-sectional data.** A cross-sectional regression analysis relies on variation across different individual entities in order to estimate the regression line.

cross-sectional data

Data that contain information on individual entities at a given point in time.

While cross-sectional data contain information on a group of individual entities at one point in time, **time-series data** include information on a single individual entity at different points in time. For example, we might have information on after-tax wages and labor supply for each year of one person's adult life. A time-series regression analysis relies on variation across time for one individual entity in order to estimate the regression line.

time-series data

Data that contain information on an individual entity at different points in time.

Finally, **panel data** (also called *longitudinal data*) combine the features of cross-sectional data and time-series data. That is, a panel data set contains information on individual entities at different points in time. You can think of panel data as a time series of cross-sectional data. For example, a panel data set might have information on thousands of different people from a variety of different years. We'll argue below that panel data have some unique advantages when it comes to doing empirical work in public finance.

panel data

Data that contain information on individual entities at different points of time.

Pitfalls of Observational Studies

Because observational studies rely on data collected in a nonexperimental setting, it is difficult to ensure that the control group forms a valid counterfactual. While the estimated regression coefficient provides a measure of the correlation between the independent and dependent variables, one cannot assume a causal relationship because outside factors could affect both of these variables.

Consider the hypothetical labor supply example above. Our regression analysis used cross-sectional data, in which some people had high after-tax wage rates and some had low after-tax wage rates. The analysis suggested that there is a positive correlation between after-tax wages and work hours. But remember, correlation does not necessarily imply causation. It could be that other factors influence both after-tax wages and hours worked, in which case the observed relationship is biased. Perhaps, for example, highly ambitious people have higher wages and also work longer hours. If so, then our observed positive correlation between after-tax wage rates and hours of work is at least partly due to differences in ambition.

As already noted, one way to address the bias in observational studies is to include other independent variables, which are referred to as control variables. Regression analysis allows us to estimate the independent effect of the variable we care about while taking into account the control variables. In our labor supply example, one might include variables such as age, nonlabor income, and gender, all of which might affect labor supply but could also be related to after-tax wage rates. But there are two problems. First, we might not think of all the control variables that should be included or all the relevant control variables may not be available in the data set. Second, some variables are very hard to measure, even in principle. Ambition is a good example. If either reason leads us to omit a control variable that is correlated with after-tax wages and influences labor supply, we will obtain biased estimates.

Despite the limitations of observational studies, they can provide useful information about the possible impacts of different government programs on societal outcomes. The key point is that these studies must be interpreted with care, recognizing the possibility that outside factors might bias any causal inferences.

▸ QUASI-EXPERIMENTAL STUDIES

quasi-experimental study

An observational study that relies on circumstances outside of the researcher's control to mimic random assignment.

Experimental studies have excellent properties when it comes to eliminating bias, but they may be difficult or impossible to perform. Observational studies have knotty problems with bias, but the data are relatively easy to obtain. Can one obtain some of the advantages of each? A class of observational studies known as **quasi-experimental studies** (also known as *natural experiments*) are used by empirical economists to estimate a causal relationship. These studies identify situations in which outside circumstances in effect randomly assign people to control and treatment groups. The difference between an experiment and a quasi-experiment is that an experiment explicitly randomizes people into a treatment or control group, whereas a quasi-experiment uses observational data but relies on circumstances outside of the researcher's control that naturally lead to random assignment.

A clever early example of a quasi-experiment comes from the work of John Snow, a 19th-century physician. At the time it wasn't known that germs cause diseases, and there were many competing theories to explain outbreaks of cholera. Snow wanted to find out whether cholera was caused by exposure to contaminated water.[3] He discovered that two water companies served much of London. One company had its intake point upstream from the sewage discharges along the Thames and so had fairly pure water, while the other company had its intake point downstream from the sewage discharges and so provided contaminated water. A natural strategy would be to compare the households who received water from one company to the households who received water from the other company. However, a potential problem arises. What if the people who received the polluted water were systematically different from the others? If they lived in poorer neighborhoods, for example, a different incidence of cholera could be attributed to factors other than dirty water. Snow indeed

[3] This example comes from Freedman [1991].

considered this important issue, and demonstrated that the assignment of water companies to different houses was essentially random:

> The pipes of each Company go down all the streets, and into nearly all the courts and alleys. A few houses are supplied by one Company and a few by the other, according to the decision of the owner or occupier at that time when the Water Companies were in active competition. In many cases a single house has a supply different from that on either side. Each company supplies both rich and poor, both large houses and small; there is no difference either in the condition or occupation of the persons receiving the water of the different Companies [Snow 1855].

In effect, Snow convincingly showed that his observational study virtually replicated a randomized study, because the treatment and control groups had similar characteristics. This randomization enabled him to rule out other factors, so he could safely conclude that the substantially higher number of cholera victims in houses receiving the contaminated water was due to the sewage.

Conducting a Quasi-Experimental Study

A successful quasi-experiment hinges critically on whether the researcher has identified a situation in which assignment to the treatment group is random. We now discuss a few approaches to establishing a valid quasi-experimental research design.

Difference-in-Difference Quasi-Experiments From time to time, policymakers suggest raising the state tax on beer in order to reduce teen traffic fatalities. Does it work? An ideal experiment would randomly assign different beer taxes to different states and then measure whether teen traffic fatalities decline in the high-tax states relative to the low-tax states. Obviously, such a study is not possible.

Now, suppose we learn that between 1989 and 1992 a group of states substantially increased their tax rates on beer, and that following the tax increases, teen traffic fatalities in these states declined by 5.2 per 100,000 teens. Could we infer that the tax increase for beer caused the reduction in teen traffic fatalities? No, because it could be that teen fatality reductions would have occurred even without the tax increase.

We would therefore want to examine what happened to a control group of states. A sensible control group would consist of those states that did not increase their beer taxes between 1989 and 1992. If the control group of states serves as a reasonable counterfactual, then we can assume a similar reduction would have occurred for the treatment states had they not increased their beer tax.

difference-in-difference analysis

An analysis that compares changes over time in an outcome of the treatment group to changes over the same time period in the outcome of the control group.

Therefore, in order to estimate the effect of the beer tax, it would make sense to compute the *change* in teen traffic fatalities in the treatment states and compute the difference between it and the *change* in the control group states. As it happens, in the control group states, teen traffic fatalities declined by 8.1 per 100,000 teens. That is, there was actually a relative increase in teen traffic fatalities in the states that increased their beer tax. Hence, contrary to the view one might obtain looking simply at the data from the treatment states, it appears that the tax increases did not accomplish the goal of reducing teen traffic deaths.

This example, based on actual estimates obtained by Dee [1999],[4] is typical of a technique known as **difference-in-difference analysis.** The reason for the name is that it compares the difference in a treatment group's outcome after receiving the treatment

[4] The treatment states were California, Delaware, New Jersey, New York, and Rhode Island.

to the difference in the outcome of the control group over the same period. This technique achieves unbiased results if one can safely assume that the changes that occurred to the control group form a valid counterfactual; that is, that they reflect what would have happened to the treatment group had it not been treated. Note that a difference-in-difference analysis is only possible if one has panel data, because the computation requires knowing how the behavior of a given group of individuals changes over time (which is the time-series part) and then comparing it to the change over the same time period for another group of individuals (which is the cross-sectional part).

instrumental variables analysis

An analysis that isolates random variation in the intervention in order to approximate an experimental study.

Instrumental Variables Quasi-Experiments Sometimes an investigator suspects that assignment into a treatment group may not be random, thus violating a requirement for obtaining an unbiased estimate. An approach to dealing with the problem is called **instrumental variables analysis.** The idea behind instrumental variables analysis is to find some third variable that may have affected entry into the treatment group but in itself is not correlated with the outcome variable.

An important issue that many local governments face is whether to reduce kindergarten class sizes. Proponents argue that such a policy leads to higher student test scores. An experiment to investigate this issue would randomly assign kindergarten students to different class sizes and then measure differences in test scores between those in large versus small classes. In fact, such experiments have been conducted. (See Krueger [1999].) As discussed earlier, one possible drawback of such an experiment is that the temporary nature of the experiment might influence the outcome.

An observational study might rely on regression analysis to estimate whether students in smaller classes score higher than students in larger classes. However, such a study would likely yield biased results because the treatment and control groups differ in many ways that can influence both class size and test score. For example, parents who are relatively more concerned about educating their children might choose schools or school districts with smaller class sizes. Such parents might also engage in other activities (such as reading with their children) that lead to higher test scores for their children. Therefore, an observed negative correlation between class size and test scores is misleading because both are caused by the outside factor of parental concern.

Hoxby [2000] developed a quasi-experiment in order to address this potential bias. Hoxby observed that there are random fluctuations in the timings of births in any given school area. These fluctuations will lead to some years where the kindergarten classes are larger than in other years. While there are many factors that determine whether a child attends a large or small kindergarten class, the variation in births from year to year represents a random component of this outcome. Hoxby therefore relied on the instrumental variables method, which takes advantage of the random determinant of class size to identify the effect on test scores. She used random fluctuations in enrollment year-to-year as an instrumental variable. This measure is correlated with class size, but does not directly influence test scores. On the basis of this exercise, Hoxby found that class size does not have a discernible effect on test scores.

Regression-Discontinuity Quasi-Experiments Eligibility for some policy programs is determined by whether a measurable characteristic of a person is above or below a specific cut-off point. For example, the government might make public health insurance available only to households whose annual incomes are below $20,000. An observational study that compared health outcomes of those who received the public health insurance to those who did not would likely be biased because the

two groups differ in many ways. Suppose, though, that instead of comparing the health of everyone above $20,000 with the health of everyone below $20,000, we compare the outcomes for those who were *just barely* eligible to those who *just barely* missed being eligible for the program. This is an attractive strategy, because while households that make substantially above and below $20,000 differ in many ways from each other, households that make $20,001 should be fairly similar to those that make $19,999. This approach is called **regression-discontinuity analysis.** The fundamental assumption that must be met for this approach to replicate an experiment is that the characteristics of those who just barely missed eligibility are the same on average to those who just barely made it.

regression-discontinuity analysis

An analysis that relies on a strict cut-off criterion for eligibility of the intervention under study in order to approximate an experimental design.

Suppose that a city is trying to decide whether to make summer school mandatory for its poorly performing students. It first wants to determine whether this step would improve academic performance. An ideal experiment would randomly assign some poorly performing students to summer school and then measure differences in future test scores between them and a control group of poorly performing students who did not attend summer school. However, political constraints probably would not permit such randomization. Instead, the city might rely on regression analysis to estimate whether students who attended summer school have higher future scores than those that did not. Unfortunately, such a study would likely yield biased results because the students who attend summer school tend to be poorer academic performers in the first place, so we would expect their future scores to be lower than those of other students even if summer school actually helped them.

Jacob and Lefgren [2004] developed a regression-discontinuity quasi-experiment to address this potential bias. In 1996, the Chicago Public Schools instituted a policy that tied summer school attendance to performance on standardized tests. If a student scored below a certain cut-off on the test, he or she was required to attend summer school; if the student scored above the cut-off, then summer school was not required. Jacob and Lefgren focused on the subsequent test scores of students who just barely qualified for summer school relative to those who just barely missed qualifying. They found a jump in follow-up reading and math scores for third graders (but not sixth graders), suggesting the existence of a positive causal effect, at least for some grade levels.

Pitfalls of Quasi-Experimental Studies

Quasi-experimental studies attempt to estimate causal relationships using observational data. The biggest pitfall is that the natural experiment may not truly mimic random assignment to the treatment group. If the underlying trends in teen fatalities were fundamentally different in states that increased beer taxes from those in states that did not, then the differences-in-differences estimate of the impact on teen traffic fatalities would be biased. If the fluctuations in births in a school area were not random or did not play a significant role in determining whether a child was in a small or large kindergarten class, then the estimated test impacts would be biased. And if those who just barely qualified for summer school eligibility were different from those who just barely missed qualifying, then the estimates of impacts on future test scores would be biased. Studies based on quasi-experiments look for situations that replicate randomization, but these attempts are not as straightforward as a pure randomized experiment.

Another concern is that quasi-experiments can only be applied to a limited number of research questions. Many interesting and important economic questions simply do not lend themselves to natural experiments. For example, as we'll discuss in

Chapter 11, the government provides guaranteed retirement income to people through the Social Security program. A critical question is whether people save less on their own when they know that they will receive Social Security payments when they retire. The problem is that Social Security was introduced to the entire nation at the same time—*everyone* received the same "treatment." Hence, there are only very limited opportunities to identify natural experiments. One economist expressed this concern by stating that "if applied to other areas of empirical work [natural experiments] would effectively stop estimation" [Hurd, 1990].

Quasi-experiments also share with experiments the concern about how to generalize the results to other settings and treatments. As discussed earlier, there is a "black box" element to these studies in that they provide reliable evidence of what happens given a very specific change in policy, but they are limited in explaining why the changes have occurred. Thus, it is difficult to use the results to predict the impact of other policies.

CONCLUSIONS

Economic theory plays a crucial role for empirical researchers by framing the research question and helping isolate a set of variables that may influence the particular behavior of interest. Empirical work then tests whether the causal relationship between a policy and an outcome suggested by the theory is consistent with real-world phenomena.

A randomized experiment is the cleanest way to establish a causal relationship between a policy and some type of behavior. However, it is not always clear whether the results of such experiments can be generalized to other contexts. In any case, economic researchers frequently must rely on observational data, which does not have the randomized feature of a controlled experiment. In these cases, the most reliable empirical analyses exploit natural experiments that mimic random assignment to the treatment or control groups.

It is not an easy task to conduct reliable empirical research. Each study will focus on a narrow research question, will rely on a certain set of model assumptions, and will test the impacts on a statistical sample of people rather than the full population. Therefore, honest researchers will frequently come to very different conclusions about the implications of a policy. Do we therefore have to abandon all hope of learning about the factors that influence economic behavior? Definitely not. In many cases one can reconcile the different empirical findings and construct a coherent picture of the phenomenon under discussion. Feldstein [1982, p. 830] has likened the economist who undertakes such a task to the maharajah in the children's fable about the five blind men who examined an elephant:

> The important lesson in that story is not the fact that each blind man came away with a partial and "incorrect" piece of evidence. The lesson is rather that an intelligent maharajah who studied the findings of these five men could probably piece together a good judgmental picture of an elephant, especially if he had previously seen some other four-footed animal.

We will refer to empirical results throughout this book, and explain the pros and cons of the research designs that generated them. In cases where the profession has failed to achieve consensus, we will draw upon this chapter to explain why. More generally, it is hoped that this introduction to empirical methodology induces a healthy skepticism concerning claims about economic behavior that occur in public debate. Beware any argument that begins with the magic words "studies have proved."

Summary

- One goal of the field of public finance is to estimate how various government policies affect individuals' behavior.
- Economic theory provides a framework for thinking about the factors that might influence the behavior of interest, and helps generate hypotheses that can be tested through empirical research. However, theory alone cannot say how important any particular factor is.
- An important purpose of empirical work in public finance is to estimate the causal relationship between a government policy and some kind of behavior. Three conditions must hold in order to infer a causal relationship between a government program and an outcome: 1) the program precedes the outcome, 2) the program and outcome are correlated, and 3) other explanations of the observed correlation are eliminated.
- It is critically important not to confuse correlation with causation. The fact that two variables are correlated does not prove that one causes the other.
- Experimental studies randomly assign subjects to either a treatment group or control group. Random assignment reduces the likelihood that outside factors will lead the researcher to confuse correlation with causation.
- Experimental studies offer a credible way to evaluate the impact of government programs, but they are not foolproof. In particular, researchers must make sure the random assignment remains over time and be careful about generalizing the results.
- Because experimental studies are often impossible to conduct, public finance economists rely on observational studies that use data obtained on real-world economic behavior.
- Econometrics is the use of statistical analysis of economic data in order to estimate causal relationships. An important econometric tool is regression analysis, which estimates the relationship between two variables while holding other factors constant.
- Observational data can be cross-sectional, time-series, or panel. Observational data are collected in nonexperimental settings. Therefore, the possible influence of outside factors can make it difficult to estimate causal relationships.
- A quasi-experiment uses observational data but relies on outside circumstances to replicate a randomized experiment.
- Quasi-experiments can be structured in several ways, such as a difference-in-difference analysis, instrumental variables analysis, and regression-discontinuity analysis.

Discussion Questions

1. In 2006, President George W. Bush proposed a cut in marginal income tax rates. Explain why it is difficult to predict the impact of such a tax cut upon labor supply on the basis of theory alone. If there were no political or legal impediments to doing so, how could you design an experimental study to estimate the impact of lower marginal tax rates on labor supply?
2. In an article on how exercise improves health, the *New York Times* reported on an observational study that found that each hour spent running added two hours to a person's life expectancy [Brody, 2006]. A week later, a letter to the editor questioned whether the results really proved anything about the impact of exercise on health, and suggested that the study could just as well be showing that "those with a strong heart and good health are otherwise more likely to enjoy running and do it more regularly." How does this challenge to the exercise study relate to the problems faced by economists trying to assess the causal effects of economic