Econometric tools are increasingly being used to study issues in education. The primary focus of econometrics in education research is to establish causal inference—to identify whether some factor, be it a student's race, teacher quality, or a policy, for example, affects education outcomes. This entry provides background on various econometric methods that are commonly used in education research and provides a basic understanding of the benefits and drawbacks of each strategy. It also provides examples of particular pieces of research that effectively use each strategy. The entry begins with background on distinguishing between causality and correlation, then it discusses the "gold standard" strategy of the randomized controlled trial (RCT). Since conducting an RCT is usually not feasible for a researcher, the entry then discusses methods that try to pull causal relationships from existing data, starting with basic regression and moving on to natural experiments, difference-in-differences (DD), instrumental variables (IV), and a few other strategies.

**Causality and Correlation**

Before diving into specific econometric strategies, it is important to review the distinctions between correlation and causality. Consider a set of students in a school. One looks at their test scores and finds that high-income students have higher test scores. This is a correlation—a relationship between two variables. It is tempting to say from this that there is an "effect" of income on test scores, that is, that higher income causes test scores to increase. This is the same as saying that income has a causal relationship with test scores. However, as will be described in detail later in this entry, just because two variables have a correlation, their relationship is not necessarily causal. In fact, most of the time, these relationships are not causal.

The example given earlier also highlights the importance of distinguishing between correlation and causality for policy. A reasonable policy response to the finding of the income-achievement relationship may be to increase income support and welfare programs. However, such a policy presumes that the lower income causes achievement to fall. Suppose instead it is not income directly but the fact that low-income students attend low-quality schools that leads to low achievement. If this is the case, then the income support policy may help a bit, but it would likely not be as effective as trying to directly improve the schools themselves. Thus, policy decisions based on correlative results are more likely to be ineffective, or potentially even harmful, than policies based on causal evidence.

The differences between correlation and causality can be illustrated through the analysis of how class sizes affect student achievement. Economists and education researchers have long theorized that smaller classes lead to improvements in student performance. However, establishing a causal relationship between these variables has proved elusive due to the fact that high-achieving students tend to enroll in schools with smaller classes. Thus, this positive relationship between students and class size may not be causal. This is a general problem in education research called *endogenous selection*.

A closely related problem is *omitted variables*, where some characteristic of a student that affects both the variable the researcher is interested in and the outcome is unobserved. For example, smaller classes may be an attractive quality of the teaching experience, and thus, high-quality teachers may gravitate toward smaller classes. In this case, the positive relationship between class size and achievement may be due to teacher quality rather than class size. Thus, these omitted variables can also explain the relationship.
There are very few questions in education research in which a causal relationship can be established by simple comparisons. Fortunately, economists and education researchers have developed a variety of tools that are able, in some cases, to extract causal relationships from data.

Regression

Almost all econometric techniques used in education research are based on linear regression, though a handful of other techniques that are sometimes used will be discussed briefly at the end of this entry. A basic regression model uses ordinary least squares (OLS) techniques where one takes a set of data and tries to draw the “best fit” line through those data by minimizing the sum of the squares of the distances from the line to the observed data points (Figure 1). Consider the problem of evaluating the impact of class size on student achievement.

Figure 1 Ordinary Least Squares Estimate of Class Size on Test Scores

![Figure 1](image)

**Figure 1** shows some hypothetical data on class size and achievement. Each dot reflects the class size and the average test score for that class. The line shows the results of an OLS regression of test scores on class size as defined by the following equation:

\[
\text{Test score} = 3.0 - 0.06 \times \text{Class size.}
\]

(1)

This says that at class size equals 0, test scores start at 3.0 and drop by 0.06 points for each one student increase in class size. More generally, this type of OLS model is defined by
where $Y$ is some outcome measure, $X$ is an educational input we care about, and $\epsilon$ takes into account everything that is not included in the model. For example, if $Y$ is a test score and $X$ is class size, then $\epsilon$ includes things like teacher quality, student socioeconomic status, and school quality. The “$i$” subscript denotes that we are talking about an individual person or thing such as a student or a school. OLS uses data on $Y$ and $X$ to estimate $\beta$, which says how much $Y$ changes for a 1-unit change in $X$, for example, how much does the test score change when increasing class size by 1. The model also estimates $\alpha$ (the “constant”), which tells us how much $Y$ equals when $X$ equals 0. OLS further estimates a standard error term for $\beta$. The standard error is a measure of the reliability of the estimate of $\beta$, or how much would we expect $\beta$ to change on average if we took a different random group of people from the population. The estimated standard error of $\beta$ is $\sigma$. These two pieces of information allow one to say whether a result is statistically significantly different from 0. Generally, given a large enough sample, if the following equation holds, then the estimate of $\beta$ is significant:

$$\frac{\beta}{\sigma} \geq 1.96 \text{ or } \frac{\beta}{\sigma} \leq -1.96.$$  

Typically, a regression model has more than the two variables in Equation 2. Thus, a typical OLS regression model is as follows:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_N X_{Ni} + \epsilon,$$  

where there are multiple “$X$” variables. Typically, a researcher cares most about one variable—called the variable of interest—and then adds others as control variables, which are variables that are included in the model because they may affect both the variable of interest and $Y$. Indeed, adding controls allows us to remove some variables from $\epsilon$ and include them in the model. So, for example, in our class-size model, one may want to estimate a model like this:

Test score = $\alpha + \beta_1 \times$ Class size + $\beta_2 \times$ Student income + $\beta_3 \times$ Teacher experience + $\epsilon.$  

The reason adding income and teacher experience to the model is important is because both of these factors can affect both test scores and class size. One might also include higher order terms for the variable of interest (e.g., $X^2$) or interactions between two variables of
interest (e.g., \(X_1 \times X_2\)). Doing this is useful when a researcher suspects that the relationship between the variable(s) of interest and the outcome variable is not linear. Finally, in the case of test scores, if the researcher has multiple years of data on a student, he or she will often include the prior year’s test score as a control as well. This allows one to interpret an estimate as an impact on test score gains rather than on levels.

While OLS is commonly used, it suffers from severe problems. Most important of these is statistical bias. That is, the estimate \(\beta\) is systematically different from the true value because it only captures correlation rather than causality. In Equation 1, it is estimated that larger class sizes reduce achievement. But what if low-socioeconomic status leads to larger class size? Then, the estimate on class size may reflect the effect of income. This is called an omitted variable bias—something is left out of the regression that affects both the \(X\) and \(Y\) variables. In some cases, one can include the omitted variable as a control. This is what is done in Equation 5. Student income and teacher experience are important omitted variables, thus, rather than leave them out of the model, it is better to include them. However, often one cannot include an omitted variable because it is unobserved. For example, teacher and school quality are also likely to affect both test scores and class size, but one cannot observe these factors in most data. Thus, omitted variable bias is only rarely eliminated using OLS. Luckily, econometrics provides a number of tools that remove this type of bias and allow researchers to find the causal impacts of education interventions. The rest of this entry will go through the strategies most commonly used in economics of education research.

Randomized Controlled Trials

An RCT, or experiment, is an effective strategy for estimating the causal effects of education interventions since a well-done RCT is free of bias. Hence, recently, experiments in education research have exploded in popularity.

An RCT is defined by a few key characteristics. First, within a given set of participants, some people are randomly assigned to receive an intervention—the treated group—while others receive no intervention—the control group. Second, the intervention is implemented with the intention of studying its effects. Thus, an RCT can be fully designed by a research team or by the agency administering the intervention (e.g., school district, state) as a pilot program prior to full implementation.

Since treatment in an RCT is based on random assignment, analysis is generally simple and relies on OLS, though complications such as attrition, noncompliance, and problems with the randomization process can arise. Typically, to assess the outcome from an RCT, a researcher will estimate the regression

\[
Y_i = \alpha + \beta \text{Treated}_i + \gamma_1 X_{1i} + \cdots + \gamma_N X_{Ni} + \varepsilon_i
\]

Thus, the researcher regresses an education outcome on whether or not a person, school, or district receives the treatment and estimates the impact of treatment on this outcome (\(\beta\)). As with any OLS model, the researcher can add control variables to improve precision or account for minor problems with randomization.

An example of an effective RCT was the Project STAR (Student/Teacher Achievement Ratio)
experiment in Tennessee. This experiment, conducted in the mid-1980s, randomly assigned a set of kindergarten students in multiple schools in Tennessee to small or large classes. Researchers have used this experiment to show that there are significant positive causal effects of class size on test scores and the likelihood of taking college entrance exams, as well as reductions in crime.

Natural Experiments

While growing in prominence, RCTs are still rarely used in education research as they involve high costs, and not every question is suitable to being answered via an experiment. Thus, as a second-best option, education economists often seek out natural experiments. These are situations where the education system is modified in a way that generates random differences in exposure to a treatment across students, schools, or staff. For these random changes to occur, typically the change in the system—often a policy change—needs to be sudden and unexpected. Furthermore, the policy generally must affect similar groups in different ways. Using our class-size example, a good natural experiment would be a situation where a school district decides to implement class-size reductions for some schools but not for others. Ideally, the choice of these schools would be random, but one can also extract causal estimates in some situations where the choice of schools was nonrandom; that is, only schools with high free-lunch eligibility rates are chosen. Economists have a number of econometric tools for exploiting these natural experiments to answer important questions in education.

Difference-in-Differences

The most common natural experiment technique used in education research is DD. This method can generate causal estimates by relating changes before and after treatment in a treated group to changes in a group that does not receive a treatment. To see this, consider the example in Figure 2. The figure shows two groups that are eligible to receive financial aid to attend college. At year zero, the treated group becomes eligible to receive additional aid, while the comparison group’s aid does not change. The DD approach takes advantage of the fact that, while different in levels, attendance prior to the increase follows similar trends in both groups. Thus, as in work by Susan Dynarski that looks at how the elimination of a Social Security financial aid program affected enrollment, one can calculate the impact of financial aid on college attendance by calculating

$$DD = \Delta_{\text{Treated}} - \Delta_{\text{Comparison}} = \frac{\text{CollegeRateTreatedPost} - \text{CollegeRateTreatedPre}}{\text{CollegeRateComparisonPost} - \text{CollegeRateComparisonPre.}}$$

(7)

Figure 2 Difference-in-Differences Estimates of Impact of Financial Aid on College Attendance
Typically, one uses a regression-based equivalent to models like Equation 8 as this allows for the addition of controls. Thus, the general DD model is of the following form:

\[ Y_{it} = \alpha + \beta_1 Treated_i + \beta_2 Post_t + \beta_3 Post_t \times Treated_i + \gamma_1 X_{1it} + \ldots + \gamma_N X_{Nit} + \varepsilon. \]  

(8)

In this model, \( \beta_3 \) provides the causal estimate. Note that there is now a “\( t \)” subscript to allow variables to change over time.

For DD to provide an unbiased causal effect, the groups must have similar trends over time in the absence of treatment. If the trends differ, then one may mistakenly attribute differences due to diverging or converging trends to the treatment. Unfortunately, this needs to be assumed as one cannot observe the trend for the treated group in the absence of treatment. Given this, researchers need to use the data available to try to rule out potential differential trends.

**Lotteries**

When an open-enrollment school has more applicants than slots available, admission is often determined by a random lottery. The student receives a randomly assigned number, and if that student’s number is chosen before space fills up, he or she is offered admission to the school. From a research perspective, lotteries are very attractive as they closely replicate an RCT at a much lower cost to the researcher. The procedure for analyzing a lottery is very similar to that of an RCT. Julie Berry Cullen, Brian Jacob, and Steven Levitt use lotteries to look at the impact on achievement of attending a higher quality high school. Their basic...
model assesses the impact of being admitted to a higher quality school using a model like the following:

\[ Y_i = \alpha + \beta \text{Admitted}_i + \gamma_1 X_{1i} + \cdots + \gamma_N X_{Ni} + \epsilon_i. \]  

(9)

Thus, instead of a “treated” variable, one is interested in the effect of being admitted to the school through the lottery. This analysis can be conducted by OLS as in the RCT framework or, if compliance with admission is not perfect, an IV framework, described in the next section.

Nonetheless, there are some drawbacks to lotteries relative to RCTs. First, since researchers have no control over who enters the lottery, there may be concerns that lottery participants are different from other students and thus do not tell us about the impacts of these programs beyond this particular subpopulation. Second, often schools are oversubscribed because they are perceived by parents to be better than alternative options. If this parent perception is accurate, then lotteries may only tell us about the effectiveness of the best schools.

Instrumental Variables

The IV method isolates changes in a variable of interest solely due to factors unrelated to the outcome variable. Thus, if one is worried about omitted variable bias, IV ensures that the estimated impact is based only on changes not induced by the omitted variable and thus establishes a causal effect. This is done through the use of an additional variable (the “excluded instrument” or “instrument” for short) that is correlated with the variable of interest but uncorrelated with the outcome variable except through the variable of interest. For example, a number of papers have used changes in minimum-dropout-age laws as instruments for the impact of dropping out of school on life outcomes. The idea behind this instrument is that people born in different states and years face different restrictions on dropping out, forcing some to stay in school longer in a way that is unrelated to other factors that could affect life outcomes.

In another example, Scott Imberman uses shopping centers as an instrument for charter school penetration to test the impact of charters on public school performance. The worry is that charter schools may choose locations with low-quality public schools. Imberman argues that shopping centers are popular locations for charter schools—thus correlated with charter penetration—but are uncorrelated with student outcomes.

To estimate an IV model, the most common strategy is two-stage least squares. Using the charter competition example, one first estimates how the number of nearby shopping centers affects charter penetration:

\[ \text{CP}_{it} = \gamma_0 + \gamma_1 \text{ShoppingCenters}_{it} + \gamma_2 X_{2,it} + \cdots \]
\[ + \gamma_N X_{N,it} + \mu_{it}. \]  

(10)
Then, one uses the estimates from Equation 10 to create $CP_{it}$, the predicted values of $CP_{it}$. The last step uses these predicted values in place of CP to estimate the causal impact of CP on achievement:

\[
\text{Achievement}_{it} = \beta_0 + \beta_1 CP_{it} + \beta_2 X_{2, it} + \ldots + \beta_N X_{N, it} + \epsilon_{it}.
\]

(11)

Despite IV's usefulness, it is often difficult for researchers to find instruments. First, the instrument needs to be completely uncorrelated with factors that affect the outcome variable regardless of whether those factors can be observed. This is not testable and thus must be assumed. The requirement for a complete lack of correlation is very strict, and few variables can satisfy it. Second, the correlation between the instrument and the variable of interest needs to be large; otherwise, the standard errors become too big. Nonetheless, instruments have been used in many other contexts in education research. For example, Tom Dee uses IV to estimate the impact of schooling on civic engagement, and Caroline Hoxby uses it to estimate the education benefits from school districts competing with each other for students.

**Regression Discontinuity**

The basic idea behind regression discontinuity (RD) is that if an intervention is implemented due to a student or school reaching some threshold, then one can compare those who just barely exceed the threshold with those who just barely fall below, as usually only random factors determine who exceeds the threshold within this small group. While the RD method can be very powerful, there are key limitations. First, one cannot have any manipulation around the cutoff. For example, if students who fail have the opportunity to retake the exam, this can cause problems, as being above the cutoff is no longer random. Second, the RD only provides the causal effect for a very specific group of people—those near the cutoff. Thus, one cannot say, for example, what the impact of attending an elite school is for students at the top of the class. Hence, researchers must be careful about extrapolating results to a more general environment. Finally, the variable that determines the cutoff needs to be quantitative and continuous. That is, one cannot do RD on a categorical variable. For example, suppose an education intervention is provided to special education students and students who have limited English proficiency but not other groups. While one may be able to evaluate such an intervention with another method listed earlier such as DD, one cannot conduct an RD in this situation.

Atila Abdulkadiroğlu, Joshua Angrist, and Parag Pathak used RD to estimate the impact of attending elite public high schools in Boston and New York on achievement. These schools admit students based solely on a threshold score on an entrance exam. Thus, the authors are able to use an RD design that compares those who barely exceed with those who barely miss the cutoff.

RD designs can be “strict” or “fuzzy.” A strict RD occurs when compliance with the cutoff is perfect—everyone who exceeds it gets treated while everyone below does not. In a fuzzy RD, compliance is imperfect. Figure 3 graphs a hypothetical likelihood of admission to the elite school versus points on the entry test. The panel on the left shows the strict case, while the
one on the right shows the fuzzy case when the cutoff is 50. Figure 4 shows that, in either case, what a researcher is looking for is a break in the trend for the outcome variable at the cutoff.

**Figure 3 Discontinuities in Elite School Enrollment**

![Figure 3 Graph](image)

**Figure 4 Relationship Between Admission Test Score and Future Achievement**

![Figure 4 Graph](image)

To estimate a strict RD, the most common method is to choose a range of observations close to the cutoff and estimate a regression model of the form

$$\text{Achievement}_{it} = \beta_0 + \beta_1 \text{Above}_{it} + f(\text{EntryScore}_{it}) + \varepsilon_{it}. \quad (12)$$
\( \beta_1 \) provides the causal impact of being above the admissions threshold on achievement. Note that in the strict model, \( \beta_1 \) is also the causal impact of enrolling in the elite school on achievement. A very important part of this model is \( f(\text{EntryScore}_{it}) \), which is some polynomial function of the entry score that is allowed to vary above and below the cutoff. Researchers will often also add control variables or use nonparametric methods such as local-linear regression.

For a fuzzy RD, one can use a two-stage least squares approach, where first one estimates the impact of the cutoff on enrollment and then the impact of enrollment on achievement.

\[
\text{Enroll}_{it} = \gamma_0 + \gamma_1 \text{Above}_{it} + f(\text{EntryScore}_{it}) + \mu_{it} \\
\text{Achievement}_{it} = \beta_0 + \beta_1 \text{Enroll}_{it} + f(\text{EntryScore}_{it}) + \epsilon_{it}
\]

(13)

Other Methods

Fixed Effects

If one has data that follow individuals over time, fixed effects may be used. These models estimate impacts using changes within individuals over time. Their advantage is that they remove bias resulting from omitted variables that do not change over time, such as innate ability. However, they do not account for bias from omitted variables that do change over time; hence, generally, economists prefer natural experiments or randomized trials. Random-effects models that provide smaller standard errors but require stronger assumptions are also used.

Propensity Score Matching

This method follows a two-step process, where one first estimates the likelihood of an individual being treated with an education intervention and then generates predicted probabilities of treatment—the “propensity score.” Then, the researcher calculates the differences in outcomes between treated observations and control observations with similar propensity scores. Although this method does not account for unobservable omitted variables, it has advantages over OLS in that it does not make restrictions on the form of the estimating equation and ensures that treated individuals are only compared with similar untreated individuals. However, it typically requires large amounts of data to get sufficiently reliable estimates.

Conclusion

This entry provides a broad overview of the econometric tools now used in education research. The use of these tools has been increasing dramatically over time and has had a marked effect on education research. The entry also highlights the importance of figuring out the causal impacts of educational interventions. Because econometric models are well suited to establishing causal effects of education policy, it is likely that their use will continue to increase in the future, which makes understanding these methods essential for education policy researchers, policymakers, and anyone else with an interest in education policy.
See also Instrumental Variables; Omitted Variable Bias; Ordinary Least Squares; Quasi-Experimental Methods; Regression-Discontinuity Design

Scott A. Imberman
http://dx.doi.org/10.4135/9781483346595.n89
10.4135/9781483346595.n89

Further Readings